A Wildfire Vulnerability Index for Businesses Using Machine Learning Approaches

Andrew Byrnes
Dun & Bradstreet[1]
byrnesa@dnb.com

Lisa Stites
Dun & Bradstreet[1]
stitesl@dnb.com

Abstract

Climate change and housing growth in ecologically vulnerable areas are increasing the frequency and intensity of wildfire events. These events can have profound impact on communities and the economy, but the financial and operational impacts of wildfire on businesses have not been evaluated extensively yet. This paper presents a Wildfire Vulnerability Index (WVI) that measures the risk of a business failing in the 12 months following a wildfire event. An XGBoost algorithm champion model is compared to two challenger models: 1) a model that extends the champion model by incorporating building and property characteristics and 2) a model that utilizes a neural network approach. Results show that while all models perform well in predicting business failure risk post-disaster event, the model that uses building and property characteristics performs best across all performance metrics. As the natural environment shifts under climate change and more businesses are exposed to the consequences of wildfire, the WVI can help emergency managers allocate disaster aid to businesses at the highest risk of failing and can also provide valuable risk insights for portfolio managers and loan processors.

1 Introduction

Wildfires are a growing problem as factors such as climate change and housing growth exacerbate the spread and impact of these events [1]. Drought and persistent heat are creating warmer, drier conditions that lead to longer wildfire seasons and climate change has been identified as a key driver in increasing wildfire conditions [2, 3]. Additionally, housing growth in the wildland-urban interface, which is areas where housing meets or intermingles with wildland vegetation, has increased the risk of human-caused wildfires and the risk of damage to structures due to housing locations in areas that are at greater risk for wildfire exposure [4]. The negative consequences associated with wildfire events impact both communities and the economy. Direct impacts include damage or loss of businesses and homes, while indirect impacts can include reduced economic activity from decreases in physical economic activity (e.g., foot traffic and tourism) because of smoke compromising air quality [1, 5, 6].

Although there is an abundance of wildfire impact research with regards to residential housing, less work has been dedicated to understanding the impact of wildfire on commercial businesses. To

[1] Any opinions provided are the author's own personal opinions and are not necessarily the views of Dun & Bradstreet. The authors reserve the right to update the work.

understand the effect of wildfire on a business, the vulnerability of a business, in terms of its ability to survive in the months following a wildfire event, should be evaluated. In this context, business survival can be defined as maintaining business-as-usual operations and remaining active in the months after a wildfire event, as opposed to going out of business or becoming commercially inactive. In essence this contributes to resilience, which has been studied in the literature relating to community, economic, and business resilience [7, 8, 9]. Prior studies have also developed wildfire vulnerability indexes and modeling that are limited to building characteristics, landscape, and social vulnerability [10, 11, 12]. While some aspects of these previously developed indices are useful in predicting business vulnerability, there exists a gap in extending wildfire vulnerability to businesses' operational status, influenced by financial health, and their ability to recover from a natural disaster.

This paper addresses the gap by presenting a Wildfire Vulnerability Index (WVI) that measures the risk of business failure, which is defined as going out of business or becoming inactive, in the months that follow a wildfire event. The WVI is based on the concept that a natural disaster event, such as wildfire, can negatively impact operating infrastructure or the ability to operate the company in a business-as-usual manner, which can impact a company's financial situation and cause persistent change in its operational status. These impacts on a subset of commerce can more broadly affect the business community and macroeconomic landscape because business relationships are complex, spanning geographies and crossing industries. As extreme weather events become more prevalent under climate change, identifying and quantifying the linkage between operational status, business activity, and macroeconomic distress from natural disaster events is becoming increasingly important. Building upon a previously developed champion model, this research tests both a new dataset in the model and a new modeling approach. The baseline champion model was developed using the extreme gradient (XGBoost) algorithm. Two challenger models are introduced that incorporate data on building characteristics and utilize a Neural Network framework. The models are compared across multiple metrics (e.g., AUC, Gini, KS, Capture Rate 10%) to determine the best model for predicting business failure in the months following a wildfire event.

2 Data and methods

2.1 Data

Data related to businesses and their financial performance was sourced from the Dun & Bradstreet (D&B) Data Cloud. D&B collects, curates, and validates business information from thousands of data sources on over 500 million global businesses. Constant updates allow businesses to be tracked over time from incorporation to bankruptcy or ceasing operations. Data sourced from D&B includes information such as business firmographics, business activity signals, trade history, financial statements, and D&B risk scores. These risk scores predict future risk behaviors ranging from financial stress to business growth. Historical wildfire perimeter shapefiles were collected from the National Interagency Fire Center (NIFC) and historical temperature, wind, and precipitation data were sourced from the National Oceanic and Atmospheric Administration (NOAA). Land cover data collected from the United States Geological Survey (USGS) was also incorporated. Finally, Federal Emergency Management Agency (FEMA) disaster assistance data and several factors from the FEMA National Risk Index, such as expected annual losses from wildfire, social vulnerability, and community resilience were included (see appendix for more details on data sources).

For development of the challenger model with new data, property and building characteristic data were selected that ultimately are sourced from county property register offices. This data was sourced from the time period April 2023. Within this dataset are features including the number of building stories, construction materials, roofing material, construction year, and property type.

2.2 Data pre-processing and feature selection

To identify businesses located near wildfire perimeters, geospatial joins were applied to match the latitude and longitude of the business address to wildfire perimeter polygon shapes represented in the NIFC shapefiles. The sample was limited to businesses located within 50 miles of wildfire perimeters between the years 2015 to 2019. Approximately 1 million businesses that were active as of the wildfire event and affected by historical wildfires were sampled. The observation period was one month prior to the start of a wildfire, with the performance period running from the month of impact to 12 months later. The target variable was defined as if the business entity went out of business or became inactive in the 12 months following the wildfire event. The event rate at which businesses went inactive in the dataset was 10.44%. This sample was enriched with the dataset features discussed in section 2.1 to create the analytical dataset.

From the initial list of over two hundred variables, the features were reduced using a feature evaluation framework to calculate the overall information value (IV) of all attributes and identify features that provided at least some value in predicting the target event in a univariate manner. Variable clustering was also used, which selected only the top variables from each cluster – this further reduced the number of features. The top features identified were then included in the final model. Additional preprocessing to the dataset included one hot encoding to convert categorical variables to a series of binary variables which is required by the XGBoost algorithm. Finally, for the XGBoost models, model segmentation was performed to create three different models based on whether the businesses had data on financials and trade payments or not. This splits the dataset into roughly thirds. Segment 1 contained businesses with both financials and trade data, Segment 2 contained business with trade data, but not financials data, and Segment 3 contained businesses with neither financials nor trade data.

For development of the neural network challenger model, additional preprocessing was required to convert the variables to a common scale. For unbounded variables with large outliers, the log of the value was used. Next, all variables and the log-scaled variables were scaled using the min max scaler technique to convert all variables to a 0 to 1 scale. Finally, for any missing or null values, the values were imputed using either zero or the mean of the variable, depending on the meaning of the variable.

2.3 Proposed framework

The resulting subset of potential predictor variables was input into an XGBoost algorithm, a highly efficient, flexible, and optimized gradient boosting library that implements machine learning decision tree algorithms under a gradient boosting framework [13].[2] XGBoost was selected for its efficiency, ability to handle missing data, and ability to uncover complex nonlinear interactions between the features. A random search hyperparameter tuning technique was used to find the best specification of hyperparameters for the algorithm.[3]

For the neural network challenger model, a sequential model was leveraged.[4] This alternative methodology was selected for its ability to also uncover complex relationships between the variables and its ability to handle many predictors in the model. The model utilized two Dense layers and two Dropout layers to mitigate any overfitting. A random search hypermeter tuning technique was used to select the learning rate and number of neurons in each of the dense layers.

2.4 Model evaluation

To evaluate model performance, the analytical dataset was split into a train, validation, and test dataset using a 60%-20%-20% split, respectively. The Gini, KS, AUC, and capture rate within the top 10% were employed to determine the best model.

3 Results and discussion

The champion model demonstrates the ability to predict the risk to a business following a wildfire event. Within the test data, this baseline model performs well with a KS of 60.28% and a capture rate in the top 10% riskiest segment of 59.53%. This shows that this model performs well in risk ranking all businesses, as well as segmenting out the top riskiest businesses from the others –

---

[2] Models were written in Python 3.7 with the following packages: scikit-learn (0.24.2), xgboost (1.7.0-dev), numpy (1.26.0), pandas (2.1.1), geopandas (0.14.0), tensorflow.keras (2.9.1) and mlflow (1.29.0).

[3] The following parameters were tuned to optimize performance of the model and minimize overfitting on the training dataset: column sample by level, column sample by tree, learning rate, max depth, minimum child weight, number of estimators, and subsample.

[4] The Neural Network approach was developed using the tensorflow keras package.

important for disaster planning and mitigation.

The first challenger model, an XGBoost model with building data, was tuned on 10 iterations and the best model was selected based on the AUC of the validation dataset (see appendix for final parameters). This model demonstrates that it can also accurately predict the risk to a business following a wildfire event. The KS is higher than that of the champion model at 61.97% and a 62.37% capture rate in the top 10% riskiest segment, showing that this model performs well within the out-of-time test data. Additionally, this model performs better than the champion model across all other performance metrics.

After tuning the neural network sequential model, the best model was selected based on the AUC of the validation dataset in accordance with reasonable results in the remaining validation metrics (see appendix for final model summary). After training for 50 epochs with early stopping, this challenger model also shows promise in predicting business activity following a wildfire natural disaster (with a 58.09% KS and a capture rate in the top 10% riskiest segment of 55.96%).

Table 1: Summary of model performance on testing data set

Metric	Champion Model	XGBoost + Building Data	Neural Network Approach
AUC	87.49%	88.61%	86.21%
Gini	74.97%	77.21%	72.43%
KS	60.28%	61.97%	58.09%
Capture Rate 10%	59.53%	62.37%	55.96%

While the champion model and two challenger models developed here all demonstrate predictive capabilities, the XGBoost model with building data features performs best on the testing dataset. Since a building's structural resilience to wildfire can have impact on a business's capacity to continue business-as-usual operations post-disaster, the superior performance of this model is consistent in context.

From the XGBoost models, predictors that are important to predicting the operational health of businesses following a wildfire natural disaster can be analyzed by looking at the feature importances output by the model. Across all segments of the XGBoost model with building data, a firm's pre-disaster financial health is a top predictor of the health of a business post-disaster. This makes sense as businesses would expect to be financially stressed from a wildfire and those already in dire financial straits would be expected to be pushed over the edge if a wildfire were to hit, while a business in financial strength and with excess reserves might be able to absorb the impact. Additionally, building usage, which determines if a building is for residential or commercial purposes, is a top predictor and in the top 6 predictors for 2 of the 3 model segments.

4 Conclusion and future work

As climate change is set to increase the frequency of wildfire events around the globe, economic resiliency is key to sustaining global commerce and trade. Now more than ever, the ability to quantify and predict the risk that the business landscape faces from wildfires can serve as a proactive capability in disaster risk management. Understanding potential business risk and geographic or industry concentration can then help devise and evaluate effective mitigation and adaptation efforts. By combining predictive modeling with natural disaster simulation, the business community can understand the trade-offs of possible future states and associated actions.

One large gap in the modeling approaches presented in this paper is the lack of historical property and building data to match the timeframe of the analytical dataset. Based on the data available at the time of model development, the property and building data needed to be back populated based on recent data. However, buildings are frequently renovated, torn down, and rebuilt. With historical data, the models may be able to better capture the impact of building characteristics on business resilience to wildfire. Another future opportunity would be to develop a solution for a short-term impact model that would predict the operational status 30 days after wildfire impact and how long it would take for a business to turn the lights back on based on real-time signals. With a short-term model, supply chain disruptions in other regions and targeting of businesses that need immediate relief assistance can be more accurately quantified to get the local economy back on the right track. Finally, future work can connect this model with real-time wildfire information to enable score updates as new wildfires ignite and spread. This would allow stakeholders to see the risk to the business landscape change over time and make real-time decisions.

References

[1] Kramer, H. A., Mockrin, M. H., Alexandre, and P.M., Radeloff, V. C. "High Wildfire Damage in Interface Communities in California." *International Journal of Wildland Fire*, vol. 28, no. 9, 2019, pp. 641-50.

[2] National Oceanic and Atmospheric Administration (NOAA). "Wildfire climate connection", 07/24/2023.

[3] Westerling, A. L., et al. "Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity." *Science*, vol. 313, no. 5789, 2006, pp. 940-43.

[4] Radeloff, V.C., Helmers, D.P., Kramer, H.A., Mockrin, M. H., Alexandre, P.M., Bar-Massada, A., Butsic, V., Hawbaker, T.J., Martinuzzi, S., Syphard, A.D. *et al*. "Rapid Growth of the US Wildland-Urban Interface Raises Wildfire Risk." *PNAS*, vol. 115, no. 13, 2018, pp. 3314-19.

[5] Molina, J.R., González-Cabán, A., and Rodríguez y Silva, F. "Wildfires Impact on the Economic Susceptibility of Recreation Activities: Application in a Mediterranean Protected Area." *Journal of Environmental Management*, vol. 245, 2019, pp. 454-63.

[6] Kochi, I., Donovan, G. H., Champ, P.A., and Loomis, J.B. "The Economic Cost of Adverse Health Effects from Wildfire-Smoke Exposure: A Review." *International Journal of Wildland Fire*, vol. 19, no. 7, 2010, pp. 803-17.

[7] Cartier, E.A. and Taylor, L.L. "Living in a wildfire: The relationship between crisis management and community resilience in a tourism-based destination." *Tourism Management Perspectives*, vol. 34, 2020, pp. 1-12.

[8] Rose, A. "Economic resilience to natural and man-made disasters: Multidisciplinary origins and contextual dimensions." *Environmental Hazards*, vol. 7, no. 4, 2007, pp. 383-98.

[9] Hussen saad, M., Hagelaar, G., van der Velde, G., and Omta, S. W. F. "Conceptualization of SMEs' business resilience: A systematic literature review." *Cogent Business & Management*, vol. 8, no. 1, 2021, pp. 1-33.

[10] Papathoma-Köhle, M., Schlögl, M., Garlichs, C., Diakakis, M., Mavroulis, S., and Fuchs, S. "A Wildfire Vulnerability Index for Buildings." *Scientific Reports*, vol. 12, no. 6378, 2022, https://doi.org/10.1038/s41598-022-10479-3.

[11] Vaillant, N. M., Kolden, C. A., and Smith, A. M. S. "Assessing Landscape Vulnerability to Wildfire in the USA." *Current Forestry Reports*, vol. 2, 2016, pp. 201-13.

[12] Wigtil, G., Hammer, R. B., Kline, J. D., Mockrin, M. H., Stewart, S. I., Roper, D., and Radeloff, V. C. "Places Where Wildfire Potential and Social Vulnerability Coincide in the Coterminous United States." *International Journal of Wildland Fire*, vol. 25, 2016, pp. 896-908.

[13] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

A Appendix

A.1 Data sources

Table 2: Data sources summary

Data Source	Description
D&B Data Cloud	Business information on 500 million + businesses globally that is collected, cleaned, and validated on an ongoing basis. Data tracks businesses from creation to bankruptcy or ceasing operations. Data includes:

	Business firmographics - number of employees, age of the business or industry. Business activity signals - how many times the business is searched for by other businesses. Trade history - how timely the business pays its bills. Financial statements - revenue, profit, loss, and other data typically found on 10-k filings. Risk scores - predictive scores that measure the risk to a business or opportunity for growth.
National Interagency Fire Center	Data on the historical wildfires within the United States. This includes information on the date of the wildfire, the name, number of acres burned, and a geospatial polygon that represents where the wildfire occurred.
National Oceanic and Atmospheric Administration	Historical data on temperature, wind, and precipitation at a specific location.
United State Geological Survey	Data that classifies a location as rural, urban, marshlands, forested, or many other land cover types.
Federal Emergency Management Agency	Historical data on relief assistance by natural disaster type. This includes whether certain programs for relief were approved and also measures of how much relief assistance dollars or loans were approved once a natural disaster declaration had been approved. National Risk Index data that classifies disaster risk in the United States at the county and census tract level. Factors include expected annual losses, a measure of social vulnerability, and a measure of community resilience.
Building and Property Data	Data on buildings and properties including their location, usage type, detailed property type, number of stories, construction materials, roof materials, value, and acres.

A.2 Model details

Model segments are defined as follows:

- Segment 1 contains businesses with both financials and trade data.
- Segment 2 contains businesses with no financials data, but some trade data.
- Segment 3 contains businesses with neither financials nor trade data.

Table 3: XGBoost final parameters

Parameters	Segment 1	Segment 2	Segment 3
Column Sample by Sample	0.9	0.8	0.7

Column Sample by Tree	0.8	0.5	0.8
Learning Rate	0.05	0.075	0.1
Max Depth	4	5	5
Min Child Weight	9	8	5
Number of Estimators	400	600	400
Sub Sample	0.6	0.8	0.9

Table 4: Neural network model summary

Layer (type)	Output Shape	Param #
dense_3 (Dense)	(None, 79)	20,145
dropout_2 (Dropout)	(None, 79)	0
dense_4 (Dense)	(None, 15)	1,200
dropout_3 (Dropout)	(None, 15)	0
dense_5 (Dense)	(None, 1)	16
Total params: 21,361		
Trainable params: 21,361		
Non-trainable params: 0		