Predicting Fire Department Response Times

Jackson Gawarecki

Jgawarecki2@bellarmine.edu

1/16/2023

This predictive analysis project utilizes data from the publicly accessible Louisville Metro Open Data website to develop a model for accurately predicting the response times of fire departments across the city of Louisville. Focusing on critical factors such as fire department jurisdiction (agency), time of day, call type, and priority level, it aims to provide valuable insights for emergency management and planning. Recognizing the importance of rapid response in firefighting, the National Fire Protection Association's benchmark of 5 minutes and 20 seconds for 90% of dispatched incidents serves as a crucial metric. Previous models, like XGBoost in the Montréal Fire Department, inspire the use of Decision Tree Regression, Random Forest, and XGBoost in this project. Python, with libraries including Pandas, Scikit-learn, NumPy, Seaborn, XGBoost, and Matplotlib, facilitates data processing, training, and testing. The proposed models offer a balance between interpretability, flexibility in handling categorical data, and the ability to capture complex patterns in the dynamic dataset. By predicting and managing response times effectively, this project contributes to the overall efficiency of fire departments, potentially saving lives and resources.

Using data from the publicly accessible Louisville Metro Open Data website, this predictive analysis project seeks to develop a model that can accurately predict the response times of fire departments across the city. The primary factors considered in the analysis include the specific fire department (such as Louisville, Fern Creek, etc.), the time of day, the nature of the call (whether it's a medical emergency, structural fire, gas leak, etc.), and the assigned priority level ranging from 1 to 10. By analyzing these variables, the project aims to construct a robust model that offers insights into the anticipated arrival times of firefighters on the scene, providing valuable information for emergency management and planning.

Response time in firefighting is extremely important as it could be the literal difference between life and death. It is of utmost importance for the government and civilians to have a firm grasp on how long emergency response teams may take to arrive on the scene. The National Fire Protection Association set a benchmark goal of 5 minutes and 20 seconds for 90% of dispatched incidents. This benchmark is a goal time, and it is crucial for departments to understand factors influencing their actual response times. The model is made to point out crucial factors and in turn enhance the effectiveness of fire departments.

This problem of finding, predicting, and managing response times has been acknowledged and worked on in various environments and contexts. Historically, departments have based their estimations on factors such as geographic proximity to the incident and the availability of units and transportation. This is not always enough to have a ‘good’ prediction because dynamic factors, such as weather and traffic patterns, are not addressed. One notable example of a more advanced model is taken from the article, “*Using Data Science to Predict Response Times of Firefighters”* written by Cyril Pecoraro. The author uses data from the fire department of Montréal to outline a potential solution for predicting response time using XGBoost models, extreme gradient boosting. This use of XGBoost allows for consideration of a broader, more dynamic set of influencing variables than historically accounted for.

This project utilizes a dataset sourced from the Louisville Metro Open Data website, specifically from the section relating to fire departments. The dataset, originally named “Louisville Metro KY - Louisville/Jefferson County Fire Districts calls for service,” provides a record of fire-related incidents and events in the Louisville Metro and surrounding areas. In all 167,587 rows, each event is attributed to the agency that dealt with it. This includes the Louisville Fire Department, Fern Creek FD, and many others. Along with the date, there are 6 other columns dealing with the time of day and the respective times for the chain of events (when the initial call was made to when the scene was clear). Another group of important columns is the event type which deals with what threat firefighters are responding to and the priority level of that event from 1 to 10. Finally, there is more geographic data including the location address and the zip code.

In this project, the Decision Tree Regression model will serve as one of the main tools in predicting response times. Given the categorical nature of the dataset, dummy variables will be extremely prevalent throughout the process. The flexibility of Decision Tree Regression allows for categorical factors such as event type and agency to be properly considered and understood. The model’s interpretability is also crucial because it allows a clear path to how a decision was made. Its ability to capture and model non-linear relationships provides a link between the complex and dynamic nature of emergency situations. Decision Tree Regression also ranks features based on their importance in making predictions. With this information, we can glean some insight into how we may be able to increase efficiency. The Decision Tree Regression model is a fitting choice for the project as it offers a balance between ease of interpretation, flexibility in handling data types, and the capacity to understand and capture complex and dynamic patterns in the dataset.

Another model that would suit the dataset well would be the Random Forest model. This model uses many decision trees to create an ensemble approach to the regression model. It shares some positives with Decision Tree Regression, such as flexibility and ranking of factors, but offers a more robust model that can deal with intricate patterns and data. The ensemble method provides more accurate and dynamic predictions because it combines many different decision trees. This also reduces potential overfitting problems that simple Decision Tree Regression models might hold. Another important feature of Random Forest is the model's reliability in predicting new, unseen data. The Random Forest model, with its ability to handle complex relationships, provide accurate predictions, and scale easily with large datasets makes it a good fit for the project.

Extreme Gradient boosting, or XGBoost, is another model well suited for the dataset. Like Random Forest, XGBoost is another ensemble machine learning algorithm. It combines simple, weak learners like decision trees to form a more robust and accurate model. The ensemble approach combats problems such as overfitting and in turn, improves generalizations so that accurate predictions can be made using unseen data. Unlike Random Forest, however, XGBoost sequentially adds its weak learners (decision trees), assigning weights to the observations and optimizing the next iteration based on the errors of the previous one. Overall, XGBoost provides a combination of accuracy, interpretability, and scalability that would help interpret the data well. The iterative steps also provide the model with a more accurate output capable of overcoming pitfalls other models have.

Much of the project will be completed using the programming language python because of its many easy to access libraries and the familiarity I have with the language itself. Within python, some libraries that will be accessed are Pandas, Scikitlearn, NumPy, Seaborn, XGBoost and Matplotlib. NumPy and Matplotlib will be used for basic statistics and visualization of data. They are the backbone of the other imported libraries. Pandas and Scikitlearn are imported to aid in the training, testing, and predictive analysis of the dataset. More specifically, train-test-split, Decision Tree Regressor, Random Forest Regressor, and other statistical models (accuracy score, mean squared error, and confusion matrix). These will be used for the bulk of the cleaning, training, and testing of our models. XGBoost is also being imported as it is one of the three models intended to be used. Finally, Seaborn is imported to aid in more complex data visualizations.

Response time of fire departments is a nationally benchmarked statistic. Every department understands that keeping that number low means potentially saving money and more importantly, lives. Using data gathered from the Louisville Metro Open Data website, this analysis project aims to accurately predict the general response time of local fire departments based on the jurisdiction, time of day, call type, and priority level of the initial call. Decision Tree Regression, Random Forest Regression, and XGBoost, with the help of python and its libraries, will be employed to find a robust model that can accurately predict the response time output.

References-

Louisville Metro KY - Louisville/jefferson county fire districts calls for service. Louisville Kentucky Open Data. (n.d.). https://data.louisvilleky.gov/datasets/LOJIC::louisville-metro-ky-louisville-jefferson-county-fire-districts-calls-for-service/explore

Pecoraro, C. (2018, May 1). *Predicting the response times of firefighters using Data Science*. Medium. https://medium.com/crim/predicting-the-response-times-of-firefighters-using-data-science-da79f6965f93#:~:text=Prediction%20of%20Travel%20Time,estimations%20about%20the%20response%20time.

Katirai, Matin, "Sprawl, equity and fire department response times across the U.S." (2009). *Electronic Theses and Dissertations.* Paper 729.
https://doi.org/10.18297/etd/729

Fire response time. FIRE RESPONSE TIME. (n.d.). https://fems.dc.gov/page/fire-response-time#:~:text=NFPA%20Standard%201710%20establishes%20an,29%20for%20a%20detailed%20description).

Moore-Merrell, L. (2019, July 16). *Understanding and measuring fire department response times*. Lexipol. https://www.lexipol.com/resources/blog/understanding-and-measuring-fire-department-response-times/?fwp