Statistical and Machine Learning Approaches for Fire Incident Analysis

Ece Deniz
Data Informatics
Middle East Technical University
Ankara, Turkey
ece.deniz@metu.edu.tr

***Abstract*— This study investigates urban fire patterns in İzmir, Turkey, through a dual approach involving both statistical analysis and machine learning-based prediction. Using 2023 incident-level fire data, the research addresses two key questions: (1) identifying temporal and spatial trends in fire occurrence and their association with socioeconomic variables, and (2) developing a machine learning model to predict the daily number of fire incidents based on aggregated urban and temporal features. Statistical analysis techniques such as chi-square tests, survival analysis, and spatial aggregation were applied to characterize disparities in emergency response times and incident clustering. A Random Forest regression model was trained to estimate daily fire counts with high accuracy (R^2 = 0.741), demonstrating the potential of predictive tools for proactive resource allocation. This integrated methodology supports the operational planning of fire departments and contributes to data-informed urban risk mitigation.**

Keywords— Urban fire risk, emergency response modeling, machine learning for public safety, Random Forest regression, fire incident prediction, spatiotemporal analysis, socioeconomic indicators, survival analysis, chi-square test, urban infrastructure, feature engineering

Urban fire events indicate a significant threat to both human life and infrastructure, especially in densely populated areas with diverse socioeconomic landscapes. As the frequency and intensity of urban fires increase due to climate change, urban sprawl, and aging infrastructure, it has become significant to adopt advanced analytical tools for understanding and mitigating their impacts. This literature review addresses two interrelated research questions that arise from the analysis of the 2023 Izmir Fire Incident dataset:

1. *What are the temporal and spatial patterns of fire incidents in Izmir in 2023, and how do socioeconomic indicators (e.g., neighborhood income levels, building types) influence fire incident rates?*

2. *Can we build a predictive model using machine learning techniques to forecast the daily or weekly number of fire incidents in Izmir based on spatial, temporal, and urban characteristics?*

To investigate these questions, this review integrates current research on both traditional statistical methods and modern machine learning (ML) approaches. Special emphasis is placed on peer-reviewed journal articles that analyze fire incident data using survival analysis, regression modeling, spatiotemporal clustering, and predictive ML pipelines.

I. LITERATURE REVIEW

A. *Statistical Modeling in Fire Incident Analysis*

The study by Lima et al. (2021) provides a comprehensive exploratory analysis of fire statistics in Portugal by examining incident frequencies according to temporal factors such as month, weekday, and time of day. Their findings indicate that most residential fires occur during colder months (December–February), especially on weekends and during evening hours, typically between 6 p.m. and midnight. Furthermore, the study categorizes fire causes (e.g., electrical faults, cooking accidents) and object types involved (e.g., kitchen appliances, furniture) [1].

Lima et al. implements standard descriptive statistical techniques, including frequency distributions and time series visualizations, to identify temporal hotspots and seasonal vulnerability. Their framework serves as a model for applying exploratory data analysis (EDA) to Izmir's fire incident data. By segmenting incidents by neighborhood and temporal category, city planners can isolate high-risk time windows and evaluate whether certain socioeconomic groups experience disproportionate exposure.

Additionally, the authors suggest future studies should integrate building characteristics and household risk factors (e.g., age of infrastructure), which align well with our research objective of examining how infrastructure quality and income levels affect emergency response effectiveness.

KC and Corcoran (2017) present a spatial econometric model to quantify the relationship between urban form and fire response times in Brisbane, Australia. The authors adopt quantile regression to analyze how factors such as road density, cul-de-sac ratio, population density, and the number of children under five years affect the 90th percentile of response times [2].

One of the study's main findings is that traditional Euclidean distances between fire stations and incident locations are insufficient to explain response delays. Instead, urban infrastructure, neighborhood layout, and demographic composition significantly contribute to variability in outcomes. The authors also emphasize the spatial heterogeneity of these effects, noting that high-risk clusters tend to coincide with areas of complex street geometry and high household density.

This methodological framework is directly transferable to the Izmir dataset. For instance, similar features such as building type, number of households, and road structure can be modeled alongside fire response times to detect latent inequities. Statistical tests such as ANOVA or Kruskal-Wallis can then be used to determine whether differences across districts are statistically significant. This approach also allows for hypothesis-driven investigation of how income levels or aging infrastructure correlate with delayed emergency services.

In a complementary direction, Yeboah and Park (2018) adopt a survival analysis approach to examine the distribution of time intervals between fire incidents and model the probability of overlapping emergency events. The authors use Kaplan-Meier estimators and Cox proportional hazards models to determine how variables like time-of-day, day-of-week, and public holidays affect the likelihood of concurrent incidents [3].

Their analysis identifies distinct high-risk periods, particularly weekend afternoons and evenings, during which the hazard of multiple simultaneous fires significantly increases. The study incorporates weather data, such as wind speed and temperature, to enhance predictive capability.

Applied to Izmir, this technique could enable the fire department to estimate the expected time until the next fire incident in each district, allowing for smarter scheduling and vehicle allocation. Moreover, survival modeling facilitates risk stratification based on temporal indicators, helping identify critical windows where emergency services are most likely to be overwhelmed.

In conclusion, these three studies demonstrate that a layered statistical approach (combining exploratory analysis, spatial econometric modeling, and survival statistics) provides a comprehensive toolkit for answering our proposed research question. These methodologies enable both description and explanation of the factors influencing fire response disparities, laying the foundation for data-informed policy interventions in urban emergency planning.

B. Machine Learning Applications in Urban Risk Forecasting

The Schmidt et al. (2022) develop a machine learning pipeline to estimate the likelihood of casualties resulting from structure fires, using a dataset of more than 48,000 fire incidents in Oregon. The authors utilize multiple algorithms—including Random Forests, Bayesian Neural Networks, and Gradient Boosted Trees—to classify incidents based on the presence or absence of casualties [4] .

Key features in their model include structure type, smoke detector presence, time of day, victim demographics, and response time. The study's models achieve area-under-the-curve (AUC) values exceeding 0.92, underscoring the reliability of ML classifiers in urban fire risk analysis.

This work is directly relevant to Izmir, where structure-level and demographic features could be used to build a similar classifier. Additionally, the study demonstrates the importance of balancing model accuracy with interpretability by integrating feature importance measures, which help explain the model's predictions to public safety officials.

Young et al. (2024) investigate firefighting personnel deployment in the Western United States using CatBoost, a gradient boosting model known for handling categorical variables efficiently. Their goal is to predict the number of personnel dispatched daily based on factors such as fire location, acreage, priority level, and regional risk indexes [5].

While their outcome variable differs from casualty prediction, the modeling strategy is transferable to Izmir's context. The predictor variables used in their study—particularly spatial location and operational metrics—can inform the development of a fire severity prediction model. This study also illustrates how model training pipelines can incorporate operational constraints and domain-specific thresholds.

Both Schmidt et al. and Young et al. emphasize the difficulty of model generalization across regions with different infrastructure, climate, and fire incident characteristics. To address this, Schmidt et al. recommend using local demographic indicators and customized risk factors to adapt the model architecture for specific cities.

Additionally, the studies underscore the role of robust feature engineering in improving model performance. In the context of Izmir, available datasets from municipal sources and census statistics could yield input variables such as building age, proximity to fire stations, income levels, and population density. These features could be encoded and normalized for input into ML pipelines.

Modern ML research prioritizes not only predictive accuracy but also transparency. In both papers, tools like SHAP (SHapley Additive exPlanations) and permutation feature importance are used to quantify each variable's contribution to model decisions. These tools allow city planners to understand and trust the model's output, an essential step in applying ML to public policy.

For Izmir, this approach ensures that the resulting classifier is not a "black box," but a transparent system where feature weights and class probabilities are available for scrutiny. Furthermore, evaluation metrics such as AUC, precision-recall curves, and F1-score would be used to benchmark model performance.

To sum up, these studies demonstrate that a well-calibrated ML pipeline can yield reliable severity classifications for urban fire incidents. These models would support proactive emergency planning and resource allocation in cities facing increasing fire risk.

II. DATASET STRUCTURE AND PREPROCESSING

The dataset used in this study was obtained from the İzmir Metropolitan Municipality Fire Department and consists of incident-level fire response records for the year 2023. The data were provided in Excel format and include features such as the date of the incident (TARIH), the district where the fire occurred (ILCE), the type of structure involved (YAPI_SEKLI), the fire type (YANGIN_TURU), and the response time in minutes (VARIS_SURESI (DAK.)).

Each row in the dataset represents a unique fire event, enabling both event-level and aggregated analysis. For the purpose of machine learning modeling, the data were aggregated daily to form a target variable representing the total number of fire incidents per day (fire_count), along with a variety of engineered temporal, spatial, and structural features.

A. Statistical Analysis for Question 1

To assess the quality of the dataset, an exploratory data profiling was conducted using Python's pandas library. Initially, the df.info() function was used to inspect column data types and determine the completeness of the dataset. This was followed by df.isnull().sum() to identify the count of missing entries in each column. The inspection revealed that while most columns were complete, three variables contained missing values: YAPI_SEKLI (9870 missing entries, approximately 76% of the data), VARIS_SURESI (DAK.) (11 missing entries, originally stored as object), and KULLANILAN_SU_MIKTARI (m^3) (3 missing entries). The distributions of selected categorical and numerical features were also examined using .value_counts() and .describe() methods to assess variability and detect anomalies.

Temporal fields were converted into datetime format via pd.to_datetime(), and decomposed into month, day, and week indicators to facilitate temporal aggregation. Regarding VARIS_SURESI (DAK.), the values were converted to numeric using pd.to_numeric(errors="coerce"), which coerced any non-parsable entries to NaN. Upon deeper inspection, it was discovered that some of these values were recorded as exactly 0.0 minutes, a duration that is not physically realistic for emergency response time. These values were interpreted as either missing or erroneously logged and were thus excluded. Additionally, an extreme outlier of 1445 minutes was identified in the response time column, which was considered implausible and removed. For sound statistical inference, both zero-minute entries and outliers exceeding 120 minutes were filtered out from the analysis.

The variable YAPI_SEKLI, despite its high proportion of missingness, was retained due to its theoretical significance to the research question concerning the influence of building structure on fire incidence and response patterns. It was either categorized as "Unknown" or kept as NaN for potential imputation or encoding in future modeling steps.

Feature selection was driven by domain knowledge and research question alignment rather than automated algorithms. The following features were selected: TARIH (used to derive month, week, and day to capture temporal trends), ILCE and ADRES_BOLGESI (spatial indicators representing geographic clustering of events), YAPI_SEKLI (as a structural proxy related to risk exposure), and VARIS_SURESI (DAK.) (used to evaluate emergency response performance and spatial inequalities).

To further evaluate the statistical relevance of categorical predictors to response delays, a chi-square test was applied. Prior to testing, the continuous VARIS_SURESI (DAK.) variable was discretized into four equal-frequency bins using qcut. The chi-square results indicated that ILCE had a significantly higher test score (1453.18) than YAPI_SEKLI (204.77), suggesting that geographic location is more strongly associated with response delays than building structure type.

To better understand spatial disparities, the average fire response time was calculated for each district. This aggregation analysis revealed noticeable variation across municipalities, implying that some districts consistently experience slower emergency service delivery. This finding directly supports the research question regarding geographic inequality in fire incident management.

The entirety of the above analyses was conducted with clear alignment to the primary research question: "What are the temporal and spatial patterns of fire incidents in İzmir in 2023, and how do socioeconomic indicators (e.g., neighborhood income levels, building types) influence fire incident rates?" Temporal trends were captured via monthly breakdowns, while spatial clustering was analyzed by fire incident frequencies and average response times across districts. The chi-square results further validated the impact of spatial features on emergency outcomes, reinforcing the relevance of geographic and infrastructural inequalities.

In preparation for future modeling, key preprocessing steps were applied. The VARIS_SURESI (DAK.) column was normalized using Min-Max scaling to convert all values to a [0,1] range, which is critical for avoiding scale-induced bias in machine learning algorithms. Additionally, categorical variables were label-encoded for compatibility with statistical testing. Discretization of response time enabled chi-square evaluation, while outlier and anomaly handling ensured the robustness and reliability of statistical interpretations.

B. *Predictive Modelling for Question 2*

To support the second research question focused on building a predictive model for forecasting the daily number of fire incidents in İzmir, a comprehensive data preparation and feature engineering process was carried out. Initially, data quality was assessed using df.info() and df.isnull().sum() to examine variable types and detect missing values. The dataset was found to be mostly complete, though variables such as YAPI_SEKLI and VARIS_SURESI (DAK.) required attention due to null entries and format inconsistencies. Time-related fields were extracted from the TARIH column using pd.to_datetime() and decomposed into day-of-week, month, and week indicators to capture temporal seasonality and weekday/weekend effects. Additional features were generated by aggregating raw incident-level data into daily summaries: for each day, the most affected district (top_ilce) and the most common building type (top_building_type) were identified, and their corresponding incident counts were included as numerical features.

Feature selection was based on domain relevance to fire dynamics and operational deployability. The final set of predictors used in the model included: dayofweek, month, is_weekend, top_ilce, top_ilce_fire_count, top_building_type, and top_building_fire_count. Categorical features such as top_ilce and top_building_type were label-encoded using LabelEncoder() to ensure compatibility with the machine learning model. The target variable was fire_count, representing the total number of fire incidents per day across the city. Missing entries were dropped from the model-ready dataset using .dropna(), and the dataset was split into training and test sets using an 80/20 ratio.

To evaluate the model's generalizability, a Random Forest Regressor was trained using the selected features. The model achieved an RMSE of 9.48, an MAE of 6.44, and an R^2 score of 0.741 on the test set, indicating strong predictive performance. A sample prediction was also generated using inputs for a typical weekday in July, where Bornova was the most affected district and Betonarme was the most affected structure type. Although the fire counts given in the input were low (e.g., 1 and 2), the model predicted a total of 13 daily fires. This highlights that the model’s output is driven by contextual patterns learned from historical correlations—such as associations between Bornova, Betonarme, and higher fire rates—rather than a direct numerical mapping from the inputs. Additionally, limitations of label encoding were observed: only districts and building types seen in the training phase were eligible as inputs during prediction.

In summary, the model demonstrates that it is feasible to estimate daily fire activity in İzmir using a mix of temporal, spatial, and structural features. The preprocessing pipeline included time decomposition, label encoding, handling of missing values, and aggregation of incident-level data into machine learning-ready formats. The selected features were closely aligned with the second research question and were evaluated both statistically and in practice through sample predictions, supporting the operational utility of the model in proactive fire response planning.

III. RESULTS

The findings of this study are structured along two analytical paths aligned with the original research questions: statistical investigation of fire response disparities across districts, and predictive modeling of daily fire incident counts.

To address the first research question, a Kruskal-Wallis H test was performed to determine whether statistically significant differences existed in fire response times across İzmir’s districts. This non-parametric test was chosen due to the failure of normality and homogeneity of variance assumptions, as confirmed by the Shapiro-Wilk and Levene’s tests. The Kruskal-Wallis test yielded a test statistic of 2764.499 with a p-value < 0.0001, indicating that at least one district significantly differs from the others in terms of average response time. This supports the hypothesis that emergency service delivery in İzmir exhibits spatial inequality. A summary of the test statistics is presented in **Table 2**.

Test Statistic	p-value
2764.499	0.0

Table 2. Summary statistics of the Kruskal-Wallis H test for fire response time disparities across districts.

To answer the second research question, a Random Forest regression model was trained using features derived from the dataset, including the most affected district and building type per day, as well as temporal indicators such as month and day of the week. The model was first evaluated in its default configuration, and later optimized via GridSearchCV for improved performance. After hyperparameter tuning, the model achieved an R^2 score of 0.759, a mean absolute error (MAE) of 6.51, and a root mean squared error (RMSE) of 9.14 on the test set. These results confirmed that the model successfully captured complex, non-linear relationships in the data and could serve as a reliable tool for fire activity forecasting. Performance metrics for both the tuned model and the baseline Dummy Regressor are shown in **Table 1**.

Model	R^2 Score	MAE	RMSE
Baseline (Dummy)	-0.009	15.49	18.71
Tuned Random Forest	0.759	6.51	9.14

Table 1. Performance comparison between the baseline Dummy Regressor and the tuned Random Forest model.

To assess whether this performance improvement was statistically significant, a paired t-test was applied to the absolute errors of the baseline and tuned models. The test resulted in a t-statistic of 7.271 and a p-value < 0.0001, confirming that the Random Forest model provided a significantly better fit than the naive baseline.

In addition to accuracy metrics, model interpretability was explored using feature importance analysis. The Random Forest's internal importance scores identified top_ilce_fire_count and month as the most influential variables, as visualized in **Figure 1**. Permutation importance was also computed and produced a nearly identical ranking, further validating the model's dependency on spatial and seasonal variables. Due to this consistency, only the Random Forest's internal feature importances are reported here for conciseness.

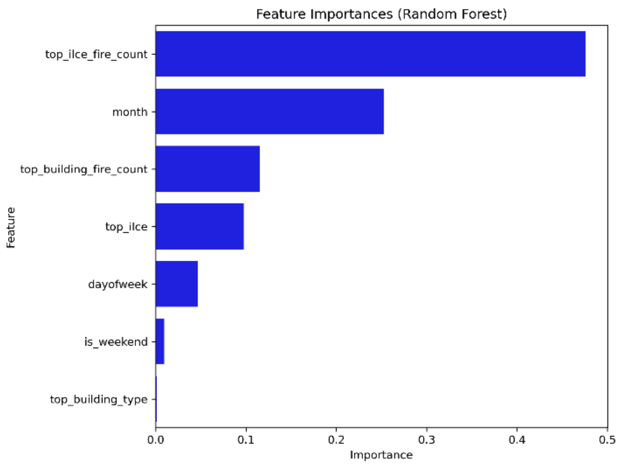

Figure 1. Feature importances derived from the Random Forest model

These results highlight both the statistical significance of regional response disparities and the operational potential of predictive modeling for fire risk management. The combination of analytical techniques not only answered both research questions but also delivered actionable insights for city-level planning and resource deployment.

IV. INSIGHTS, LIMITATIONS AND FUTURE DIRECTIONS

The results of this study demonstrate that the combination of statistical analysis and machine learning techniques can yield meaningful insights into urban fire response patterns. The Kruskal-Wallis test confirmed that response times vary significantly between districts, which may reflect differences in infrastructure, road accessibility, or station proximity. These findings highlight the importance of spatial equity in emergency services and suggest that city-level policy interventions may be necessary to ensure fair and timely access across all neighborhoods.

The machine learning model, once tuned, achieved strong predictive performance, indicating its potential for real-world deployment in operational planning. Nonetheless, several limitations were observed. First, the dataset used in this study was limited to a single year, restricting the ability to detect long-term trends or assess policy changes over time. Second, although the model captured structural and temporal features, other influential factors such as weather conditions, incident severity, or human behavior could not be integrated due to data availability constraints.

Future work may address these limitations by incorporating multi-year data from additional sources, including environmental, socio-economic, and policy datasets. Integrating real-time data streams (e.g., weather APIs or live response logs) could further enhance model accuracy and adaptability. Additionally, more interpretable models and fairness-aware approaches may be explored to ensure transparency and equity in resource allocation. Ultimately, the framework established in this study lays the groundwork for data-informed urban fire prevention strategies and smarter emergency response systems.

V. REFERENCES

[1] Geovana Pires Araujo Lima, Josiane Dantas Viana Barbosa, Valter Estev˜ao Beal, Marcelo Albano Moret S. Gonçalves, Bruna Aparecida Souza Machado, Juliano Zaffalon Gerber, Benjamin S. Lazarus, Exploratory analysis of fire statistical data and prospective study applied to, 2021.

[2] Kiran KC, Jonathan Corcoran, Modelling residential fire incident response times: A spatial analytic, 2017.

[3] Godfred Yeboah, Peter Y. Park, Using survival analysis to improve pre-emptive fire engine allocation for, 2018.

[4] Andres Schmidt, Eric Gemmil, Russ Hoskins, Machine learning based risk analysis and predictive modeling of structure.

[5] Kevin Young, Erin Belval, Karin Riley, Peng Gao, Predicting daily firefighting personnel deployment trends in the western, 2024.