Weather Conditions and Police Calls in San Jose

Predictive Modeling for Emergency Response

Kang Gun Ham
Halıcıoğlu Data Science Institute
University of California, San Diego
kham@ucsd.edu

Jayla Cho
Jacobs School of Engineering
University of California, San Diego
jac090@ucsd.edu

ABSTRACT

Emergency response plays an important role in ensuring safety in a community. Understanding the factors that affect the Priority Codes of police calls can significantly improve the efficiency of these systems. This study investigates the relationship between the features of the weather conditions and the priority code assigned to police calls in San Jose, California, from 2022 to early 2024. We used a dataset comprising weather data and call records to develop a predictive model to forecast the priority code of emergency calls based on weather patterns.

Our study reveals a small link between weather and police calls. Despite the small correlation, our study suggests the potential for our model to estimate the percentage of Priority 1 calls based on weather features. This motivates us and shows the necessity of further exploring the weather's impact on calls, which is crucial for effective emergency response in day-to-day operations. This project was done in the hopes of advancing the emergency response systems, specifically police calls, by providing potentially actionable insights through our prediction. By using machine learning techniques and real-world data, we hope our approach and implementation allow emergency responders to better anticipate calls so that it promotes public safety.

CCS CONCEPTS

• **Information systems → Data mining;** Supervised learning; Regression analysis; Feature Selection;
• **Applied computing → Law, social and behavioral sciences;** Emergency response system; Public safety; Crisis management;

KEYWORDS

Predictive Model, Weather, Priority, Police Calls

1 Literature Review

Police call system plays key element in ensuring public safety and minimizing potential harm during emergencies. Understanding the factors that influence the prioritization of emergency calls is essential for improving the efficiency and effectiveness of these systems.

While previous research on this topic has primarily focused on socio-demographic factors such as race, age, sex, etc. the scope of the studies has expanded into weather conditions in an attempt to understand criminal behavior. Ellen G. Cohn's (1990) study delves into this relationship, discussing theoretical frameworks such as the situational approach, rational choice theory, and routine activities theory, which suggest that weather could significantly impact crime rates and behavior.

Building upon those ideas, more studies have provided valuable evidence supporting the influence of weather conditions on emergency response systems. Ramgopal et al. (2019) contribute empirical evidence by examining the impact of weather on emergency medical service (EMS) responses. Their study, conducted in Western Pennsylvania, analyzes weather characteristics—temperature, precipitation, and wind speed—in relation to EMS call rates. They find notable associations between weather conditions and EMS responses, highlighting that increasing temperature and certain weather events, such as rain and snow, correlate with higher call volumes.

These studies provide insights into the relationship between weather conditions and emergency response systems. While Cohn (1990) provides theoretical insights into this relationship, Ramgopal et al. (2019) and Horrocks and Kutinova Menclova (2011) offer empirical evidence supporting the influence of weather on emergency service responses and crime rates, respectively. Putting these theoretical concepts with empirical ideas together, these studies contribute to a deeper understanding of environmental factors shaping criminal behavior and inform strategies for crime prevention and emergency response planning.

Despite these advancements, further investigation is needed to explore the specific relationship between weather conditions and police call prioritization. Our research aims to address this by analyzing weather data and call records from San Jose, California, to enhance the efficiency of emergency response systems and contribute to public safety.

By leveraging machine learning techniques with real-world data, we seek to find actionable insights to better the workflow of law enforcers and improve the efficiency of emergency response systems. Through a detailed analysis of the literature and our research objectives, we aim to contribute to advancing knowledge in this field and enhancing public safety in San Jose and beyond.

2 Dataset

2.1 Identify Dataset

The police call dataset was sourced from the 'San Jose CA Open Data Portal' as a CSV file that provides comprehensive information regarding calls received by law enforcement agencies in San Jose for various incidents or emergencies.

The Weather Dataset, obtained from the 'VisualCrossing Weather Data & API' as a CSV file contains a comprehensive record of meteorological conditions observed over a specific period, typically organized daily. This dataset offers detailed insights into various weather parameters recorded at San Jose.

We limited the data range from 2022 to 2024 to exclude data from the pandemic period, ensuring a consistent analysis. Data from the pandemic era can significantly deviate from normal patterns, potentially distorting our analysis.

2.2 Exploratory Data Analysis (EDA)

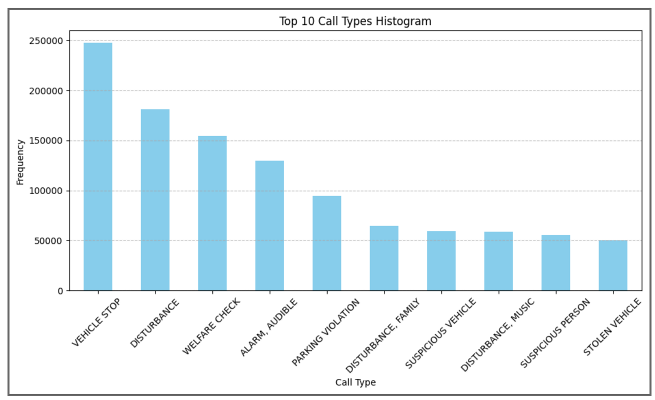

Figure 1: Top 10 Call Types

The Police Call dataset between 2022 to 2024 consists of 3,998,463 rows and 14 columns. Each row represents an individual call, while each column provides specific information regarding the nature of the call such as Time Stamp, Priority, and Call type.

In the initial preprocessing steps for the Police Call data, we standardized data types to enhance consistency and readability. Specifically, we converted 'eid' to a string format and 'CDTS' (Call Date Time Stamp) to a datetime format for efficient time-based analysis.

To streamline the dataset and focus on essential information, we conducted an in-depth exploration of coded columns such as 'calltype_code' and 'final_dispo_code'. By generating maps linking each code to its corresponding meaning, represented in the 'call_type' and 'final_dispo' columns respectively, we gained clarity and context regarding the nature of each call and its resolution. For instance, we transformed code '1066' to 'SUSPICIOUS PERSON' and code 'A' to 'Arrest Made'.

Subsequently, we pruned redundant or extraneous columns, including 'EID', 'CALL_NUMBER', 'CITY', 'STATE', 'call_type', and 'final_dispo'. By eliminating these superfluous attributes, we streamlined the dataset to focus solely on pertinent information crucial for our analysis, thus enhancing efficiency and clarity.

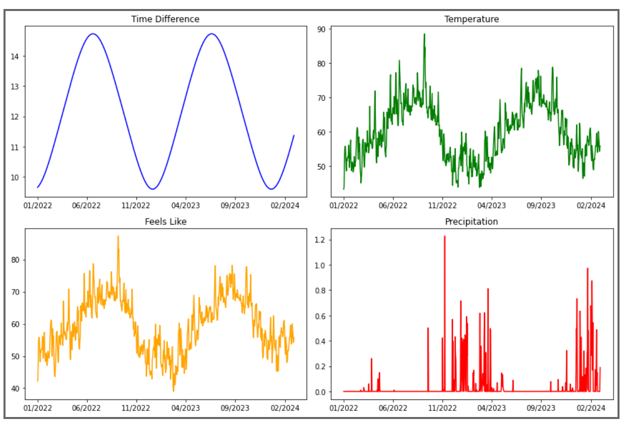

Figure 2: Weather Data Analysis

Now the weather dataset between 2022 to 2024 comprises 790 rows and 33 columns. Each row corresponds to a single day, with columns containing various weather-related information for that particular day. A wide range of meteorological variables is included in the dataset, such as temperature, precipitation, humidity, wind speed, atmospheric pressure, and cloud cover.

In the preprocessing phase for the Weather data, we commenced by refining data types to ensure consistency and facilitate analysis. Notably, we transformed the datetime column into a datetime type, ensuring accurate temporal representation for subsequent time-based analyses.

To enrich our dataset with insightful features, we engineered a novel attribute named 'daytime' by computing the duration between sunset and sunrise times. This additional column provides valuable information regarding the duration of daylight, which may influence various weather-related phenomena and human activities.

Furthermore, we pruned redundant columns such as 'description' and 'icon', which provide redundant or non-essential information for our analysis. By discarding these extraneous attributes, we streamlined the dataset to focus solely on key weather variables essential for our investigation, enhancing clarity and relevance.

Overall, these preprocessing steps for both the Police Call and Weather datasets lay a solid foundation for subsequent analyses, ensuring data consistency, relevance, and usability in our investigative endeavors.

3 Predictive Task

In this section we will go over the analysis and models that we have built. We developed our predictive model using Python and machine learning libraries. The GitHub[1] repository includes all implementation and sources of the dataset.

3.1 Linear Regression

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables, assuming a linear relationship between them.

In this analysis, we employed linear regression to predict the number of police calls received per day using weather data as independent variables. We

[1] link to Github: https://github.com/kanggunh/DSC148-Project.git

utilized the LinearRegression model from the scikit-learn library to train a linear regression model. This model aims to capture the linear relationship between weather variables and the number of police calls.

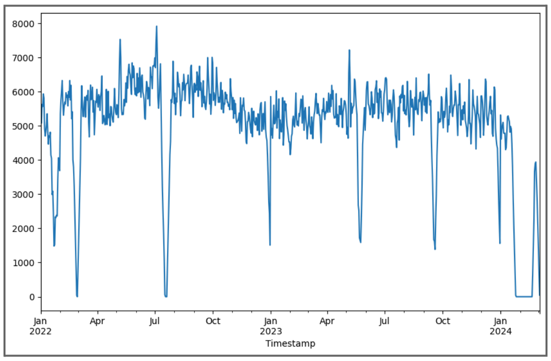

Figure 3: Number of Calls for 2022-24

3.2 Gradient Boosting for Regression

For this second model, we employed Gradient Boosting for Regression to better understand the impact of weather on emergency calls in San Jose. In this approach, we utilized Priority Codes as our dependent variable. We approached this with two distinct strategies: one based on the mean of the priority code and the other on the percentage of Priority 1 codes. However, we observed that the first strategy yielded limited meaningful insights, as the mean Priority values ranged between 3 to 4 for most of the days.

Given the intricate nature of Gradient Boosting models, careful parameter turning is essential to ensure optimal performance and prevent overfitting. To address this, we used GridSearchCV to fine-tune parameters, which is a technique that searches through a predefined parameter grid.

Here is a brief overview of the parameters we tuned and their significance. The **n_estimators** determine the number of sequential trees to be modeled, playing a crucial role in preventing overfitting. The **learning_rate** controls the magnitude of adjustment made with each new tree. Lower values are generally preferred to prevent overfitting but they may make the model computationally expensive. The **subsample**, **max_depth**, **min_samples_split**, has also been tuned in the hopes of making the model generalize well to unseen data. By adjusting all these parameters, we aimed to optimize our Gradient Boosting model's performance to accurately predict our desired measurement of urgency.

4 Result

4.1 Evaluating Linear Regression

To assess the performance of the linear regression model, we utilized the r2_score and mean_squared_error metrics from the scikit-learn library. The r2_score represents the proportion of the variance in the dependent variable that can be predicted from the independent variables, while the mean_squared_error calculates the average squared difference between the predicted and actual values.

In addition to the linear regression model, we explored alternative approaches, including Ridge, Lasso, and ElasticNet regression, to evaluate potential enhancements in predictive performance and analyze the impact of regularization techniques.

The linear regression model produced an r-squared value of 0.09 and a mean squared error of 2,198,253, indicating that approximately 9% of the variance in the number of police calls can be explained by the weather variables. The Ridge regression model achieved a similar r-squared value of 0.09 and a mean squared error of 2,196,531. Similarly, the Lasso regression model and ElasticNet regression both resulted in an r-squared value of 0.09 and a mean squared error of 2,203,054 and 2,191,561 respectively.

4.2 Evaluating Gradient Boosting Regression

To assess the predictive performance of the model in estimating the percentage of Priority calls based on weather conditions, we utilized two key evaluation metrics: Root Mean Squared Error (RMSE) and Mean Percentage Error (MPE). RMSE measures the deviation between predicted and observed values, while MPE calculates the average percentage of the difference between those two values. Lower RMSE values closer to zero indicate better predictive accuracy.

Before parameter tuning, we obtained RMSE and MPE values of 0.0181 and 0.0031, respectively. On the other hand, we got 0.0152 and 0.0028 after tuning the hyperparameters. This tuning resulted in significant improvements, with RMSE decreasing by 16.14% and MPE decreasing by 10.79%, indicating enhanced accuracy and precision in predicting the percentage of Priority calls. However, these evaluations do not provide insights into how well the model predicts since the percentage of Priority 1 every day is low in the first place.

5 Discussion

5.1 Interpretation of Results

In an effort to provide meaningful context, we conducted a correlation analysis of the features employed in the gradient-boosting model.

Figure 4 presents a heatmap depicting the Pearson correlation coefficient among all the features utilized in this model. Analyzing the correlation between the percentage of priority calls and the weather features reveals coefficients ranging from approximately -0.1 to 0.2. This indicates a weak relationship between weather conditions and the percentage of Priority calls.

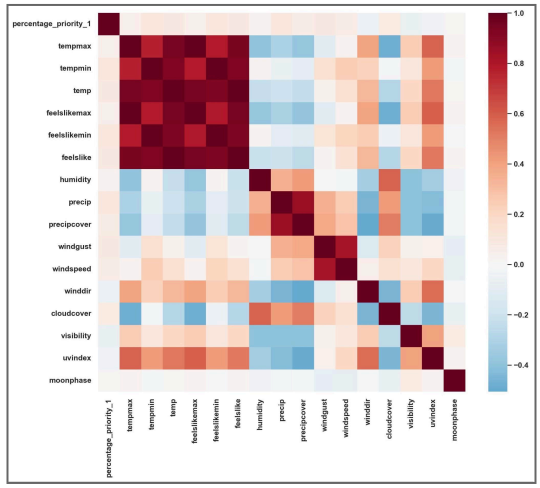

Figure 4: Correlation Heatmap

The obtained results from Linear Regression, Ridge Regression, Lasso Regression, ElasticNet Regression, and Gradient Boosting Regression indicates that the weather variables have very weak predictive power for the number of police calls. Specifically, the weather variables used in the model are insufficient to explain the variability in police call volume.

5.2 Limitations of the study

The weak predictive power of weather variables on police call volume could stem from several limitations inherent in the study. Factors such as population density, socioeconomic status, cultural norms, total police personnel, etc. could potentially have a greater impact on police call volume.

Daily averages of weather data may fail to capture short-term weather fluctuations or extreme weather events that could significantly impact human behavior and, consequently, police call volume. For example, a sudden rain or a heatwave during the day may have triggered certain types of police calls, but this effect might be diluted or masked when averaged over an entire day.

In addition, human behavior is influenced by a multitude of factors beyond weather conditions alone.

While the weather may play a role in certain types of incidents, such as traffic accidents during adverse weather conditions, it is unlikely to be the sole determinant of police call volume.

6 Conclusion

After conducting analysis utilizing both linear regression and gradient boosting techniques on the weather data and police call data, it was observed that no substantial correlation or predictive relationship between the variables could be established.

Despite our findings, it's important to note that previous research has suggested a potential association between weather conditions and crime rates. While our analysis did not find evidence to support this relationship within the context of police call volumes, it's possible that the impact of weather on specific types of crime or crime reporting behaviors may vary.

Factors such as seasonality, temperature, and precipitation have been implicated in influencing criminal activity in some studies. However, our study's focus on police call volumes may not fully capture the nuanced dynamics of weather-related effects on crime.

With that being said, future research should focus on developing a more comprehensive predictive model by incorporating all the things mentioned above and exploring different modeling techniques. For example, examining the impact of time, integrating geospatial data to capture localized variations, and investigating the ethical and social implications of predictive modeling.

By addressing these aspects, we can advance our understanding of emergency response dynamics and contribute to the development of more robust predictive models for public safety.

ACKNOWLEDGMENTS

We would like to thank Professor Jingbo Shang for his lectures and guidance in the course DSC 148 - Introduction to Data Mining. His expertise and support enriched our learning experience and contributed to the completion of this project.

Additionally, we would like to thank our class teaching assistants, Bill Hogan, Weitang Lui, and Yufan Zhuang, for their help throughout the course this quarter.

REFERENCES

[1] Ellen G. Cohn. 1990. Weather and Crime. The British Journal of Criminology 30, 1 (Winter 1990), 51–64. https://doi.org/10.1093/oxfordjournals.bjc.a047980

[2] Sriram Ramgopal, Jennifer Dunnick, Sylvia Owusu-Ansah, Nalyn Siripong, David D. Salcido, and Christian Martin-Gill. 2019. Weather and Temporal Factors Associated with Use of Emergency Medical Services. Prehospital Emergency Care 23, 6 (2019), 802–810. https://doi.org/10.1080/10903127.2019.1593563

[3] James Horrocks and Andrea Kutinova Menclova. 2011. The effect of weather on crime. New Zealand Economic Papers 45, 3 (2011), 231–254. https://doi.org/10.1080/00779954.2011.572544

[4] San Jose CA Open Data Portal. (2022-2024). Police Call Dataset Police Calls. Retrieved from https://data.sanjoseca.gov/dataset/police-calls-for-service

[5] VisualCrossing Weather Data & API. (2022-2024). Weather Dataset Weather_Sanjose. Retrieved from https://www.visualcrossing.com/weather-data