

HM 5 Writeup

Xiang Li

Introduction

The city of San Francisco is currently grappling with rising processing times for building permits, causing significant bottlenecks in the construction sector. In response to this challenge, the Department of Building Inspection has commissioned an exploration to identify the main determinants of longer processing times. The objective is to enable more efficient resource allocation across divisions to address the backlog effectively.

To achieve this objective, we conducted analyses using three distinct machine learning models: Random Forest, Support Vector Machine (SVM), and Gradient Boosting Regressor. Each model was trained and tested on a dataset spanning from Jan 1st, 2013 to Feb 25th, 2018, comprising approximately 200,000 records and 43 columns.

Data Preprocessing

The analysis begins with loading the building permit dataset and selecting relevant columns such as permit type, current status, and dates. Data cleaning steps include handling missing values and encoding categorical variables into numerical format. Additionally, extreme values are removed to ensure data integrity and reliability.

Exploratory Data Analysis (EDA)

Visualizations such as histograms and density plots are employed to explore the distribution of processing times. The analysis reveals a right-skewed distribution, with a mean processing time of approximately 49.54 days. The majority of permits are issued within a relatively short duration, highlighting potential factors influencing processing times.

Predictive Modeling

Machine learning algorithms including Random Forest, Support Vector Machine (SVM), and Gradient Boosting Regressor are utilized for predictive modeling. The Random Forest Classifier achieves an accuracy score of approximately 13.97%, indicating moderate predictive performance. Similarly, SVM with a linear kernel demonstrates an accuracy score of around 14.15%. Additionally, Gradient Boosting Regressor yields a Root Mean Squared Error (RMSE) of approximately 59.73 days, indicating the average deviation of predicted processing times from actual values.

Conclusion

Despite employing various machine learning algorithms, the predictive models demonstrate moderate accuracy in estimating building permit processing times. Further research into additional features and advanced modeling techniques may enhance predictive performance. Nevertheless, understanding the underlying factors influencing processing times is crucial for stakeholders to optimize permit approval processes and improve overall efficiency in construction permit management.

Appendix

