Understanding and predicting blight ticket compliance (Wei Chua)

Introduction

According to a 2020 study conducted by the University of Michigan on the Detroit community, nearly three quarters of respondents reported blighted properties in their area. [1] Blight is a situation where owners fail to maintain their property resulting in either danger of the building collapsing, the building potentially being a hub for immoral acts or deemed unsanitary for human inhabitation. [2] As a result, residents in neighborhoods of blighted properties report **lower neighborhood satisfaction** and **lower perceived safety**. [1] This is a serious issue affecting the city of Detroit, so each year the city issues millions of dollars in fines to owners of blighted properties, but many of these remain unpaid. Since enforcing blight is **tedious and costly**, it would be useful to understand and predict blight ticket compliance so that the city council can be more effective at enforcing blight fines.

This project is the fourth assignment in the Applied Machine Learning in Python course [3] by the University of Michigan offered through Coursera.

Aims

1. Predict the probability of blight ticket compliance given a number of input features such as judgement amount.
2. Understand the relationship between the input features and blight ticket compliance to understand why people might fail to comply.
3. With the understanding of what causes low blight compliance, the city could take action on these factors to ensure payments are made on time.

As the rate of compliance increases, the city could fund more blight remediation programs.

Data description

train.csv - tickets issued from 2004 to 2011

test.csv - ticket issued from 2012 to 2016

address.csv - file containing the ticket IDs and corresponding addresses

latlons.csv - file containing addresses and corresponding coordinates (latitude and longitude)

Data processing

Reduced the number of input features

- Removed input features with a large of number of unique categories, which are unlikely to be useful for prediction such as violator name (approx. 120000 unique categories).
- Removed features with a large proportion of missing data.
- Removed features which are identical for all instances and would not be useful for prediction such as city.
- Removed information such as payment date to prevent data leakage.

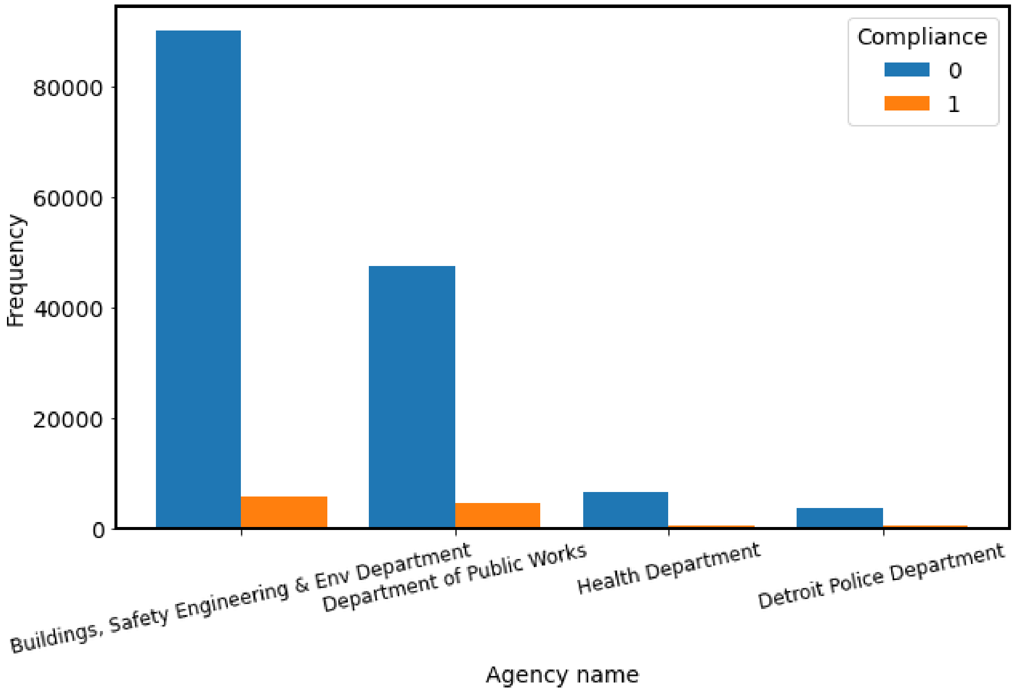
Introduced a new feature, time to hearing

- Calculated the duration between issuing the ticket and the hearing date as a new input feature.

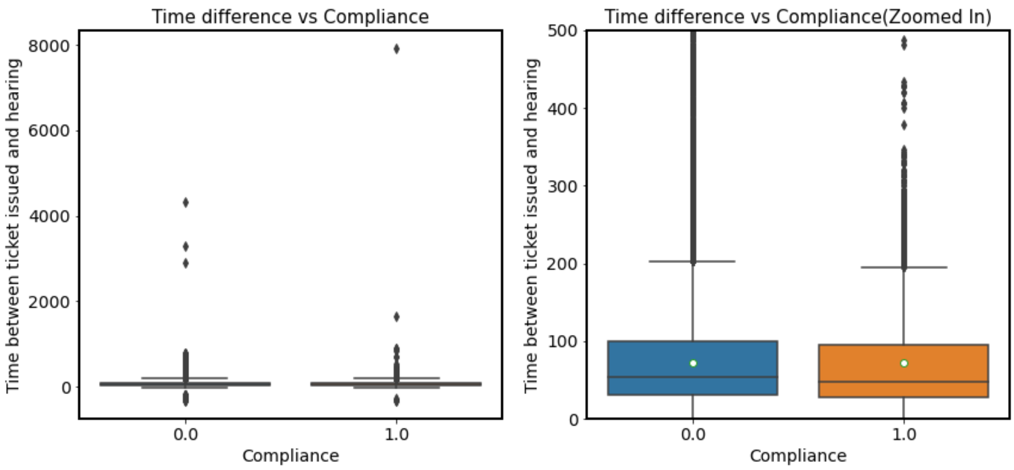
Used the label encoder to convert categorical data

- Converted agency name, violation code and disposition, which are categorical data into numerical data.

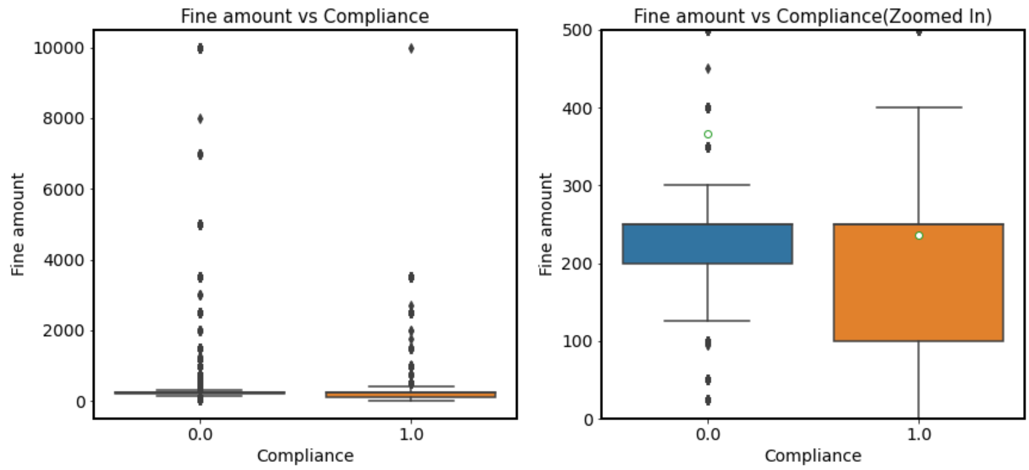
Exploratory Data Analysis



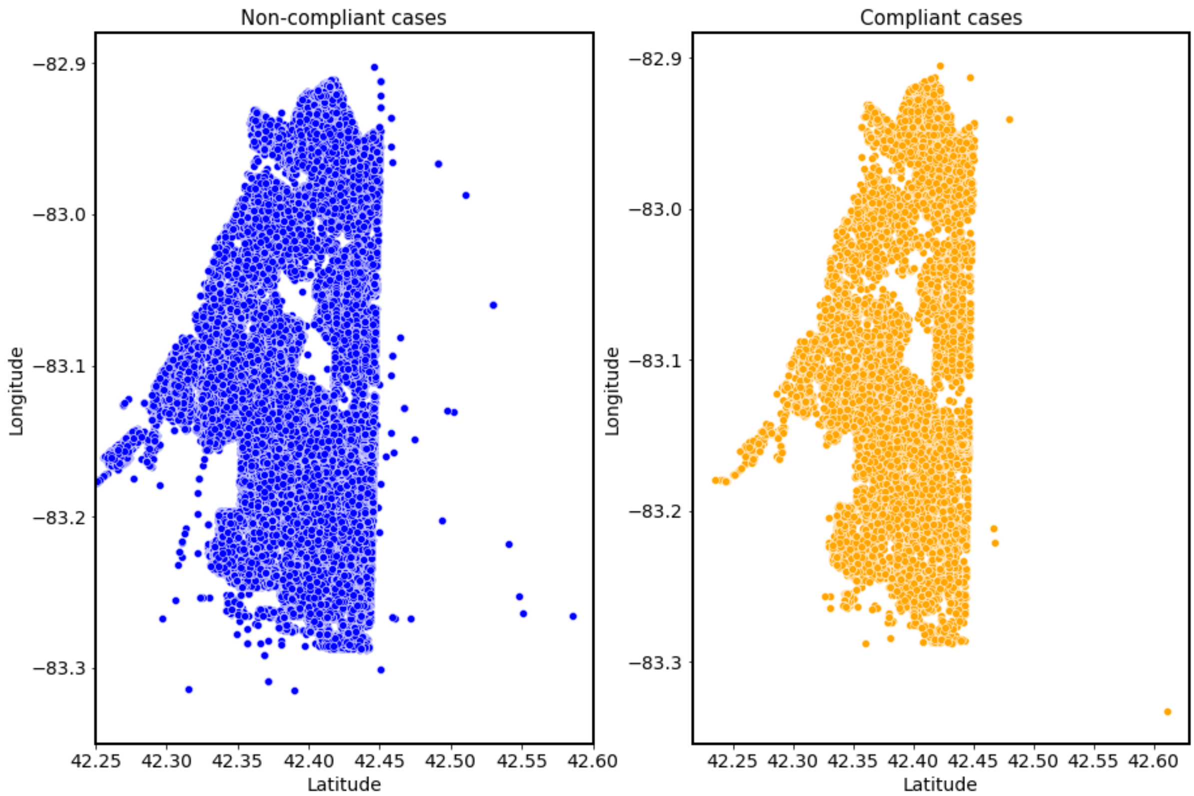
The bar chart of frequency against agency name for compliant and non-compliant cases show a larger percentage of non-compliant cases from the Buildings, Safety Engineering and Env Department and the Health Department compared to other agencies.



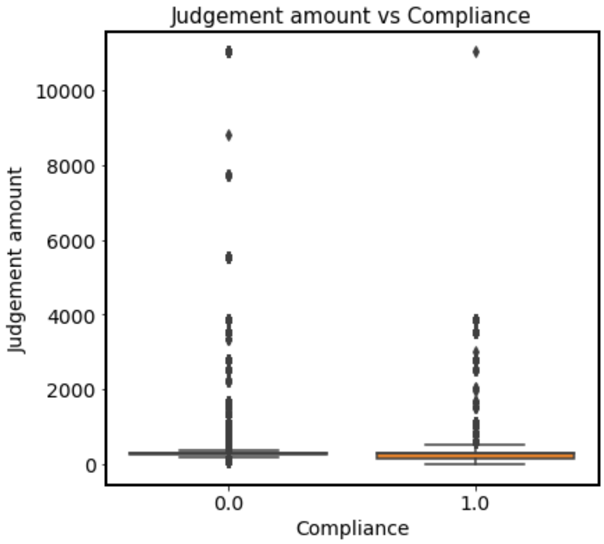
The mean duration between ticket issued and hearing is slightly higher for non-compliant cases. (The means are annotated with white circles.)

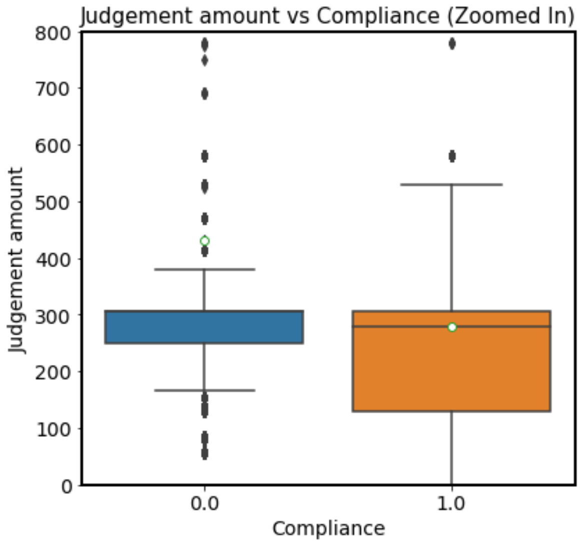


The mean fine amount is higher for non-compliant cases than for compliant cases.

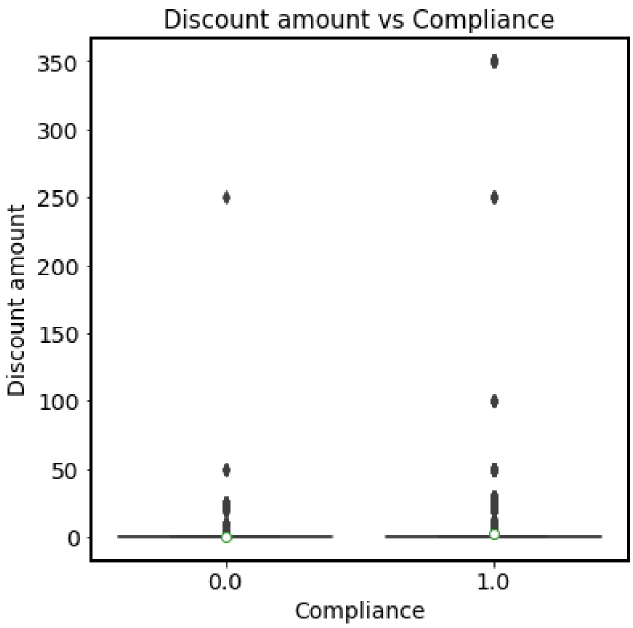


The non-compliant dataset shows more variation compared to the compliant dataset, which is much smaller. Interestingly, the majority of scatter points for non-compliant cases match the pattern of scatter points for compliant cases, which would suggest that both cases share the same subset of addresses.



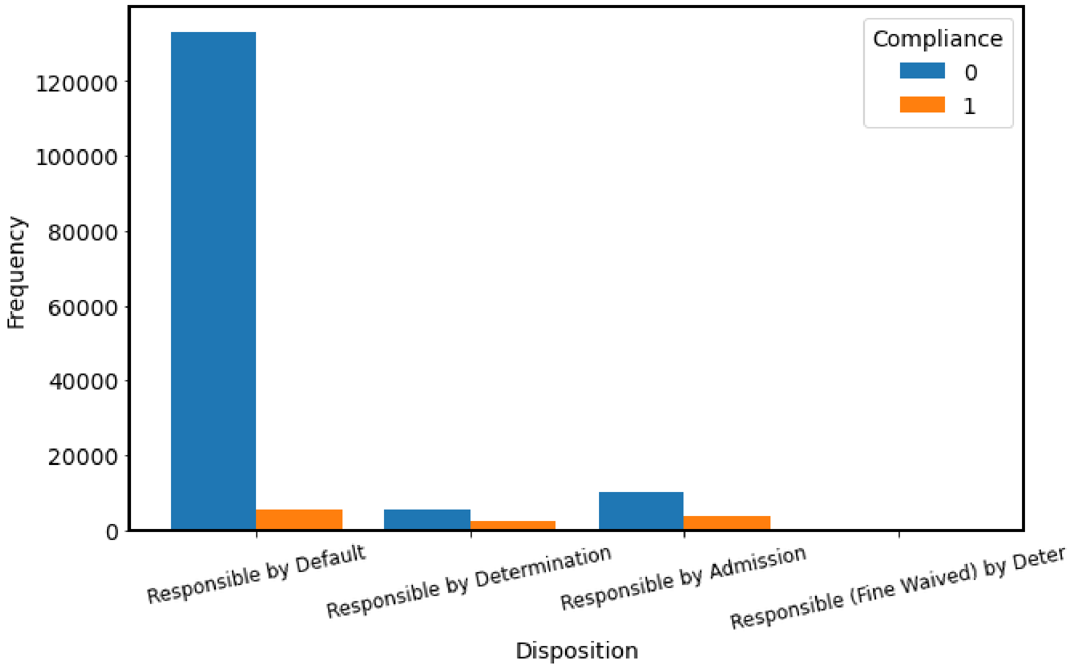


The mean judgement amount for non-compliant cases is much higher than for compliant cases.



The distribution of amounts of discount are very similar. The compliant cases have a larger variation in discount amount and a greater number of outliers. The mean for compliant cases is slightly higher than for non-compliant cases.

There are only two instances of discounts among non-compliant cases vs 998 instances of discounts among compliant cases. Ideally, a larger dataset would be needed to establish a relationship between discount amount and compliance as the current data for discount amount is very limited.



The percentage of non-compliant cases for the responsible by default category is much higher than for other categories.

Modelling and model evaluation

Model evaluation metrics

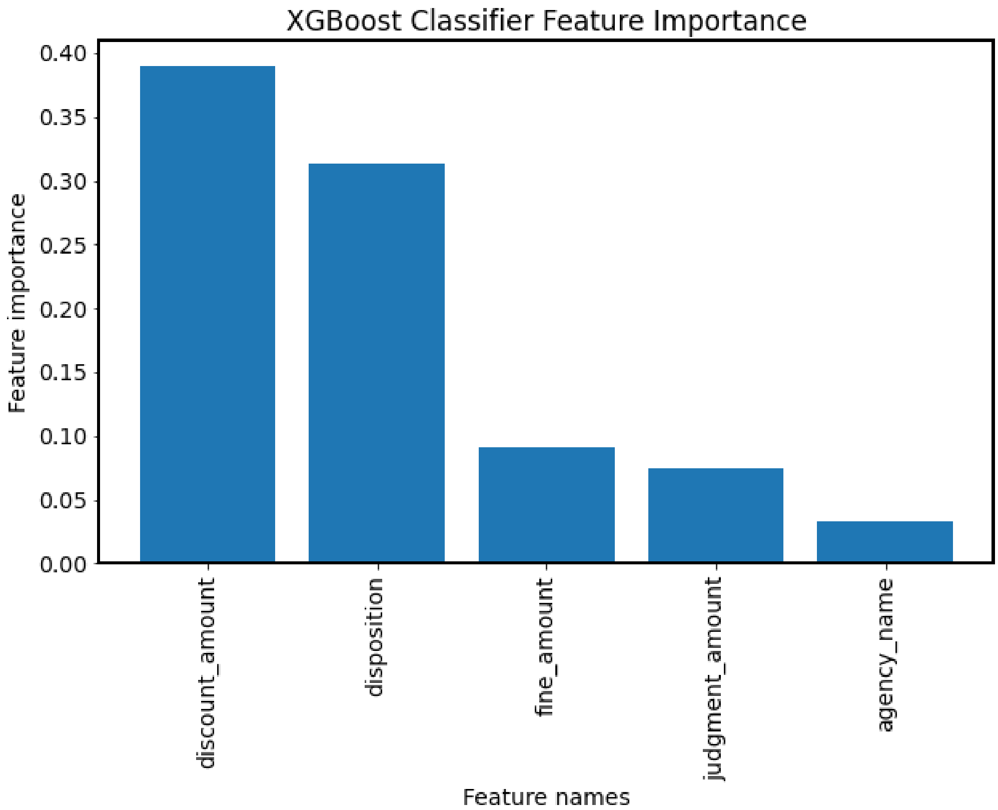
Model evaluation metrics include training and test set accuracy, area under the ROC curve and specificity which is defined as below.

$$Specificity = \frac{TrueNegative}{TrueNegative + FalsePositive}$$

Specificity was included as a metric as the aim is to identify as many individuals as possible who might not pay the fines on time, so that enforcers can channel more effort early on to ensure that these individuals pay the fines. The area under the ROC curve was included to assess the models' performance at distinguishing positive and negative classes.

Modelling

The baseline model chosen was a logistic regression model. This was compared against Random Forest and XGBoost classifiers using the metrics stated above. Optimal parameters for all three models were determined using GridSearchCV. XGBoost was chosen due to high AUC ROC score (0.704), high specificity (0.995) and high accuracy for both training (0.945) and test sets (0.94). The model also does not seem to overfit the data.



The most important features for prediction using the chosen model are discount amount, followed by disposition, fine amount, judgement amount and agency name.

Results and Actionable Takeaways

The optimized model predicted 1.2% compliance among all the instances in test.csv. This is reasonable considering the historical rate of blight compliance is under 10%.[5] However, this prediction may be improved by using unsupervised machine learning methods to balance the dataset. Currently, compliant cases only form 7.16% of the entire dataset.

Based on the feature importance bar chart, there are some key actions that enforcers can take to increase blight compliance.

- Discount amount, fine amount and judgement amount: The mean fine and judgement amounts are lower for compliant cases. The violators may be experiencing financial difficulty and are unable to pay off the tickets. This makes sense as the average individual would probably not leave their property to deteriorate if they have the financial means. One way to solve this issue may be offering the violators an opportunity to get involved in neighborhood blight remediation efforts in exchange for a fine reduction. Enforcers should pay more attention to the income of blight violators and offer alternatives such as paying a small portion of their income over a longer period of time. Another possibility is to focus more effort on ensuring middle to high income violators pay off their tickets as soon as possible. This can include not allowing violators to collect rent on blighted properties, which is currently implemented by the City of Detroit. [4]
- Disposition: The percentage of non-compliant cases was highest for violators who were determined to be responsible by default. Further analysis should be carried out to understand why these violators refuse to comply. One possibility is that some violators purchase several properties and fail to maintain them, resulting in

many deserted properties [4], hence they may be reluctant to pay for the fines. A method to increase compliance would be to encourage the violators to pay before the hearing date by possibly offering a higher discount rate the earlier they pay. Another method would be to prevent such violators from buying any more properties or renting and selling acquired properties.

- Agency name: The Buildings, Safety Engineering and Env Department as well as the Health Department face larger percentages of non-compliance. The city council can carry out an investigation to find out the reason for low compliance of fines enforced by these two agencies.

References

1. 2020. *Detroit Metro Area Communities Study*. Blight in Detroit. [online] University of Michigan. Available at: <https://detroitsurvey.umich.edu/wp-content/uploads/2021/02/Blight-Report-7-21-2020.pdf> [Accessed 7 September 2022].

2. City of Detroit. n.d. *Blight | City of Detroit*. [online] Available at: <https://detroitmi.gov/departments/law-department/blight> [Accessed 7 September 2022].

3. Applied Machine Learning in Python course. Available at <https://www.coursera.org/learn/python-machine-learning>[Accessed 7 September 2022].

4. 2017. Understanding Blight Ticket Compliance in Detroit. [online] Available at: https://midas.umich.edu/wp-content/uploads/sites/3/2017/09/understanding-blight-ticket.pdf [Accessed 7 September 2022].

5. Michigan Data Science Team. 2017. *Michigan Data Science Team - Detroit Blight Analysis*
. [online] Available at: <https://www.mdst.club/projects/detroit-blight-analysis> [Accessed 2 October 2022].