Environment and Planning B:
Urban Analytics and City Science

Is the Noise still Going on?: Predicting Repeat Noise Complaints with Historical Time Course and Random Forest Classifiers

Journal:	*Environment and Planning B: Urban Analytics and City Science*
Manuscript ID	EPB-2022-0613.R1
Manuscript Type:	Manuscript
Keywords:	Noise Complaints, Random Forest Classifier, Time Series Analysis, Spatial and Temporal Clustering

Urban Analytics and City Science

Article

11

Is the Noise still Going on?

Predicting Repeat Noise Complaints with Historical Time Course and Random Forest Classifiers

Abstract

Noise can have serious adverse effects on residents' physical and mental health. Since the COVID-19 pandemic, the City of Westminster in London has seen a continuous increase in noise complaints, with a significant number of repeat complaints from the same address within a short time scale. The authorities' ability to respond to complaints is challenged. This study explores a method for predicting and identifying repeat complaints to improve the efficiency of the authorities in dealing with noise complaints. Taking the noise complaint records of the City of Westminster during 2018-2022 as research objects, the research explores the cumulative distribution characteristics and clustering pattern of noise complaints in different spatial and temporal dimensions. On this basis, for a noise complaint from a specific address, the study fits random forest classifiers to predict whether the same address is likely to have another noise complaint in future time scales. It is found that about 18.5% of all complaints had at least one previous complaint at the same address in the previous seven days; during the lock-down period caused by the COVID-19 pandemic, areas with active commercial activities and higher housing prices experienced a significant decrease in complaints, while areas adjacent to parks and green spaces can share a similar upward trend in noise complaints. Prediction of repeat noise complaints with random forest classifiers is proved feasible. F1 scores of models to predict repeat complaints within 0 to 2nd days, 0 to 7th days and 0 to 30th days in the future are 0.55, 0.66 and 0.75, respectively. Suggestions are provided for local authorities to improve resource allocation related to noise complaint management.

Keywords

Noise Complaints; Time Series Analysis; Spatial and Temporal Clustering; Random Forest Classifier

Introduction

Noise exposure is increasingly a common and severe problem in global urbanisation (Tong & Kang, 2021a) and poses challenges to public health and urban governance. Long-term exposure to noise has been found to be associated with the risk of physical and mental health problems (WHO, 2011; EEA, 2021). Medical studies confirmed that exposure to noise environment could cause negative mental states of stress, anxiety and depression, and can be the inducement of multiple cardiovascular and endocrine diseases (Dzhambov & Dimitrova, 2016; Münzel et al., 2018). Noise can also significantly deteriorate the living experience of residents (Ottoz, Rizzi & Nastasi, 2018) and affect the property price (Bravo-Moncayo et al., 2022; EU, 2022).

According to noise sources, noise can be roughly classified as environmental noise (European Commission, 2002) and noise nuisances. Environmental noise is commonly associated with essential infrastructure in built-up areas, such as roads, railways, airports and industrial equipment, and authorities tend to develop unified legislation and long-term strategies to manage and monitor it. The noise nuisances can be generated by daily human activities, and the common noise sources include neighbours, commercial areas, and workplaces. Depending on the severity, a noise nuisance can be regarded as a 'statutory nuisance' in the UK (Great Britain. Department for Environment, Food & Rural Affairs, 2015). Complaints regarding noise nuisances are often representative among complaints of various urban problems (Kang, 2006; Peng et al., 2022). Dealing with noise complaints relies on dynamic and flexible responses from local authorities.

However, with the COVID-19 pandemic, the rapid growth of noise complaints and the repeat complaint problems bring difficulties to effective complaint management and resource allocation. The City of Westminster in London, for example, receives more than 17,000 noise complaints annually (Westminster City Council, 2021). According to the internal statistics, the borough has seen a continuous increase of noise complaints of 13% during 2019-2021, and 84% of all complaints came from addresses that have historical complaints within a short timescale. There seem to be potential concentrated and repetitive patterns of noise complaints, but no clear prioritisation is applied for responding to repeat complaints, which adds difficulties in efficiently allocating resources. The negative impact of noise exposure and the increasing pressure of noise complaint response all call for a more effective noise complaint response strategy.

In general, taking the noise complaint records in the City of Westminster as the case study area, the project hopes to explore the potential spatial and temporal pattern of repeat noise complaints, and develop methods to predict the likelihood of a repeat complaint happening on certain addresses within a given timescale. The study is expected to provide policy suggestions on efficiently predicting and dealing with future noise complaints.

Literature review

Noise Nuisance, Complaint and Response

A series of legislations, such as the *Environmental Protection Act 1990*, *Environmental Protection Act 1990* and *Noise Act 1996*, state that local councils in the UK have a legal duty to investigate complaints related to potential statutory nuisances, and noise nuisances are included. According to the definition (Great Britain. Department for Environment, Food & Rural Affairs, 2015), a statutory noise nuisance should be noise prejudicial to health or unreasonably and substantially interfering with the use or enjoyment of a home or other premises. Local councils are given autonomy to manage the noise nuisances within the boundaries independently, and set hotlines and websites to receive noise complaints. In practice, the City of Westminster lists the typical noise cases applied to complaints. The cases include loud parties or music from residential and commercial premises, noisy house-held animals, audible alarms from premises and vehicles, construction sites operating noisily beyond permitted hours, etc. (Westminster City Council, 2021). Noise from railways and airports, political demonstrations and military occupation are usually exempt from noise nuisance. To respond to a noise complaint, the local council will contact the complainer within 45 mins through text messages to confirm whether the noise or nuisance is still happening (Westminster City Council, 2022). Depending on the repetitiveness of the noise, an officer from the nearest site could be assigned to visit the complaint address to witness the noise nuisance in person and investigate the noise source. Due to the complexity of noise sources, and uncertainty of the noise duration, noise complaints may repeatedly appear at the same address within short time scales, regardless of whether the address has been visited. In-person investigation towards these repeat complaints has been reported as a heavy occupation of manpower, and may cause difficulties in allocating resources reasonably. Higher efficiencies are expected to identify and respond to a noise complaint.

Increased Noise Complaints and the Related Socio-economic Factors

Concerns about noise were heightened due to the COVID-19 pandemic, and the spatial and temporal distribution characteristics of noise complaints and their relation with socio-economic dynamics have received extra attention in urban studies (Zambon et al., 2020; Fan, Teo and Wan, 2021; Tong et al., 2021; Yildirim & Arefi, 2021; Ramphal et al., 2022). The COVID-19 lock-down and changes in work-life patterns are found to be related to the increased noise complaints, especially in economically and socially disadvantaged neighbourhoods. For example, during the COVID-19 lock-down in 2020 Spring, a substantial increase in noise complaints was found in London, compared to the same period in 2019 Spring (Tong et al., 2021). The most significant rise in noise complaints was attributed to urban construction activities and neighbourhoods. With widespread work-from-home strategies, residents were spending more time in their rooms, making it easier to perceive noise that was previously less noticeable in office environments. Besides, areas with higher unemployment and lower housing prices reported more significant changes in the complaints (ibid.). Similarly, a New York-based study shows that since 2010, noise complaints have increased significantly in the most economically disadvantaged neighbourhoods, and this disparity has been exacerbated during the COVID-19 pandemic (Ramphal et al., 2022). In addition, social diversity and certain built environment factors may also contribute to the distribution and variability of noise complaints. Tong & Kang (2021b) conducted a study on a national scale in England and found that Regions with greater social diversity in ethnicities and religions tended to receive more noise complaints. Tong & Kang (2021a) also investigated the relationship between noise complaints and urban morphology factors, such as the road transport network, land use, and building morphology. Noise complaints tended to cluster around high-density built-up areas.

Learning from Complaints to Improve Urban Governance

Complaints bring administrative pressure but also opportunities to improve urban governance. From the perspective of social sensing, complaints submitted to government departments or informal platforms such as social media, as a kind of crowd-sourcing data, can be used to monitor urban dynamics (Liu et al., 2015; Chen et al., 2021; Osorio-Arjona et al., 2021), and investigate or predict potential disasters or events (Young et al., 2022; Sadiq et al., 2022). For example, Osorio-Arjona et al. (2021) analyse the emotional feelings and corresponding geographical location of citizens' complaints in the Twitter account of Madrid Metro, to detect the spatial distribution of problems in the public transport network and optimise the public transport services. Agonafir et al. (2022) investigate the spatial variability of New York City's flood vulnerability using street flood complaints from the New York 311 platform. A random forest regression model is built to evaluate the importance of factors such as location, topography and land use in predicting flood vulnerability. In addition, with the increasing complexity of urban management, there is an opportunity to learn from existing complaints to optimise the process of dealing with new complaints for local authorities. For example, Peng et al. (2022) develop a tree-based method to classify the complaint records received from the urban 12345 hotlines in China, assign complaints to the most appropriate department and help balance the workload between different government sections. Similarly, Chen et al. (2022) have proposed an intelligent government complaint prediction framework. By learning from the interaction between citizens and government sections, the framework can merge highly similar inquiry samples and help governments respond to citizens' concerns quickly.

Overall, the complaint information can serve as ideal materials to improve urban governance strategies, and developing intelligent methods for learning and dealing with government complaints has shown good prospects. The above literature provides a new possibility to deal with noise complaints. Based on the potential repetitive characteristics of noise complaints and their interaction with the spatial and socio-economic environment in local urban areas, it is possible to fit a prediction model based on each current complaint, to identify whether a new noise complaint is likely to occur in the same space in the future. This hypothesis is further explored and tested in this study.

Data and Objects

Case Study Area

This study takes the City of Westminster borough as the case study area. As shown in Figure S1, the City of Westminster is located in inner London, adjacent to the Thames River in the south and Camden and City in the East. By 2021, the City of Westminster has a population of 204,200 (ONS, 2023a), and the population density is 9509/km² (ONS, 2023b). The city of Westminster is famous for its highly diversified metropolitan functions. With Europe's largest night-time economy, more than 3,000 eating, drinking and nightlife establishments are concentrated in the central areas of Westminster, such as Oxford Street, Covent Garden, and Soho (Westminster City Council, 2009, p.5). Besides, the City of Westminster is the seat of well-known green spaces such as Hyde Park, Regent's Park and Green Park (Figure S2). The rich activities and resources make Westminster more troubled by noise problems than other areas in London. Westminster City Council (WCC) is the local authority of the City of Westminster. Responding to noise complaints is an important concern in its terms of reference.

Data

The noise complaint data used in the study is provided by WCC, including a total of 83741 noise complaint records from April 1st 2018 to May 24th 2022, within the City of Westminster boundary. As shown in Table S1, each complaint record contains a unique noise complaint index, the date and time the complaint was received, the noise source type of the complaint, an address index created by WCC, corresponding to complainant's address when feeling and report the noise, and the Lower Super Output Area (LSOA) where the complainant's address is located. In addition to the complaint data, this study extensively collected the socio-economic and built-up environmental data within Westminster, as an attempt to explain the potential spatial and temporal distribution characteristics of noise complaints in local contexts. More details about the data are listed in Table S2 in the supplementary material.

Methods

Research Framework

The present study aims to investigate the temporal and spatial distribution patterns of noise complaints in the City of Westminster, and explore the potential for predicting repeat complaints at the same address in future time scales based on these patterns. A sample of 81,507 noise complaints from April 1st, 2018 to March 31st, 2022, was selected for analysis, and a research framework is illustrated in Figure 1.

To predict the development of events with time series, a common scheme is to decompose the time series into typical time features like trend and seasonality, and to fit a potential prediction model with the features (Box & Jenkins, 1970). Some spatial and social inducements of events, as exogenous variables, may help further improve the accuracy of the model prediction (Cryer & Chan, 2008). The scheme above has been applied to some classical time series forecasting regression methods. Following the scheme, the study first attempts to explore the spatial and temporal distribution characteristics of noise complaints via the seasonal and trend characteristics of noise complaints, and exogenous variables that may affect noise complaints.

Three hypotheses are put forward to explore in this study. Firstly, due to the regularity of human activity at different time scales, the number of noise nuisance generated by human activity may also exhibit regular fluctuations, thereby constituting the seasonal characteristics of noise complaints. Secondly, for trend characteristics of complaints, temporary events in cities, such as festivals, sports matches, and accidents, may generate a series of interrelated complaint behaviours in a local area or in a time scale. The frequency and quantity of historical complaints may to some extent indicate the probability of future complaints. Thirdly, based on the research by Tong et al. (2021) and Ramphal et al. (2022), the distribution of noise complaints in cities can be associated with long-term effects of local socio-economic status, built environment density, land use type, etc. Major changes in the urban functional pattern, such as the lock-down caused by the COVID-19 epidemic, may also change the existing spatial and temporal distribution patterns of noise complaints.

Based on the independent variables extracted from the exploration of the above hypothesis, the study attempts to fit a random forest classifier to predict the possibility of noise complaint repetition. The reason for selecting the classification rather than the regression method is that the distribution of complaints on the address scale can be highly unbalanced, and most complaints are one-time complaints. It is difficult to form a continuous time series on address scale for regression analysis. The research methods are discussed in detail in the following section.

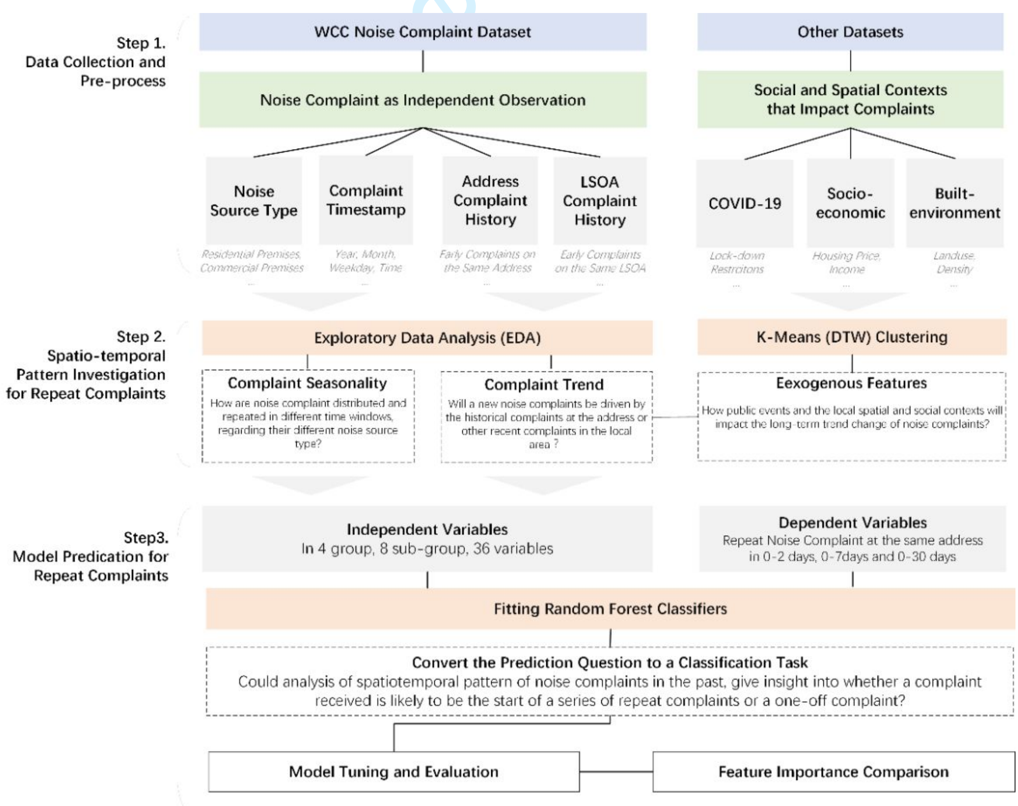


Figure 1 Research Framework

Seasonality, Trend and Exogenous Variables Related to Noise Complaints

Each noise complaint is initially considered an independent observation, labelled with the time of receipt and the type of noise source on the address scale. To identify the seasonality of complaints, the study investigates the frequency distribution characteristics in different time windows of the month, day, and hour. The potential distribution differences in terms of noise source types are also explored. Regarding the trend of complaints, the study hypothesises that activities causing noise nuisances in urban areas may trigger multiple interrelated noise complaints in both spatial and temporal dimensions, where the history of complaints at an address may indicate future complaint trends for that address or adjacent addresses. Based on this, for each complaint record, the study searched for other complaints that occurred earlier in the same day, in the previous 2nd-7th days, in the previous 7th-30th days and in the whole record history at the same address, to infer future complaint trends. For the LSOA where the address is located, the study investigated and summarised the quantity of complaints from different noise sources that occurred earlier in the same day. As for exogenous variables, the study first considered the possible impact of the UK's three nationwide COVID-19 lock-downs on noise complaints. The three nationwide lock-downs are distinguished as Mar 23th - August 14th, November 5th - December 2nd in 2020, and January 6th – July 19th in 2021, based on the time when relevant policies and restrictions were introduced. The study labels each noise complaint with whether it occurred during the three lock-down periods. In addition, the study investigated the changes in the number of noise complaints during the lock-down period for each LSOA, as well as their associations with local socio-economic characteristics and built environment features.

The K-Means Dynamic Time Warping (DTW) clustering method (Petitjean, Ketterlin & Gançarski, 2011), supported by the Python package 'Tslearn' (Tavenard et al., 2020), is used to identify similarities and dissimilarities between time series that characterise trends in monthly noise complaints for different LSOAs. Basically, K-Means clustering aims to partition data into given clusters based on the similarity between data points, and Euclidean distance is the similarity measure (Hartigan & Wong, 1979). K-Means (DTW) is a variation of this method that uses Dynamic Time Warping (DTW) to measure the similarity between two sequences that change over time. DTW works by warping the sequences non-linearly to make them align with each other. The algorithm calculates the minimum Euclidean distance between two data points x and y of the aligned sequences to find similar sequences. The algorithm can be formulated as (Sakoe & Chiba, 1978):

$$DTW(x,y)=\sqrt[\min_{\pi}]{\sum_{(i,j)\in\pi} d(x_i,y_j)^2} \qquad (1)$$

where $\pi=[\pi_0,\ldots,\pi_K]$ is a path that satisfies:

π is a list of index pairs $\pi_K=(i_k, j_k)$ with $0\le i_k<n$ and $0\le j_k<m$;

$\pi_0=(0,0)$ and $\pi_K=(n-1,m-1)$;

for all $k>0$, $\pi_k=(i_k, j_k)$ is related to $\pi_K -1=(i_k -1, j_k -1)$ as:

$i_k -1\le i_k \le i_k -1+1$; $j_k -1\le j_k \le j_k -1+1$

Based on clustering analysis, LSOAs belonging to the same cluster may share similar overall trends in the changes of noise complaint counts. For different clusters, the study compares the centroid value of the built environment and socio-economic characteristics of LSOAs corresponding to the clusters, to identify the possible impact of environmental factors on the spatiotemporal patterns of noise complaints.

Predicting Repeat Complaints Based on Classification Method

The study uses a random forest classifier to predict the probability of future noise complaints at the same address. As an ensemble learning method, random forest classifiers are generally regarded more robust than single classifiers (Breiman, 1996; Dietterich, 2000), and have the advantages of higher classification accuracy, less overfitting, and appliable to larger datasets, compared with other ensemble classifiers (Breiman, 2001; Rodriguez-Galiano et al., 2012). Nevertheless, random forests may have a poor performance curse on highly imbalanced training datasets (Chen & Breiman, 2004).

Taking each noise complaint record as an observation, the study searches the new noise complaints occurring at the same address in futural time scales (if they exist) to construct dependent variables. The time scales include: from the complaint time until 24:00 the next day (0-2 days), from the complaint time until 24:00 on the 7th day (0-7 days), and from the complaint time until 24:00 on the 30th day (0-30 days). The presence or absence of future noise complaints is input as the dependent variable in the classifier.

Four groups of independent variables, as shown in Table S3, including 8 subclasses and a total of 36 variables, are included in the classifier to examine their potential to improve the prediction of noise complaint repetition. These variables are related to the previously discussed seasonality, trend, and exogenous factors. 75% of complaints are randomly split into training dataset to fit the random forest classifier, while the rest serve as the test dataset to evaluate the model performance. As the unbalanced distribution of dependent variables may weaken the performance of random forest models, the oversampling method, Synthetic Minority Over-sampling Technique (SMOTE), has been applied to oversample the repeated complaint observations (Chawla et al., 2002) in training dataset before fitting the model. Precision, recall, and F1 score are used as the main scoring methods. Precision and recall reflect the model's performance in correctly classifying different samples and recalling positive samples (repeat complaints), respectively. F1 score is the harmonic mean of Precision and recall, and reflects the model's balance in prediction. Finally, the study analyses the importance of different independent variables in the prediction models, using the feature importance score method which is supported by the Python library of 'rfpimp' (Parr & Turgutlu, 2022).

$$Precision = \frac{tp}{tp + fp} \tag{2}$$

$$Recall = \frac{tp}{tp + fn} \tag{3}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

where tp refers to the true positives; fp refers to the false positive; fn refers to the false negative.

Considering that there is temporal order in the complaint records, the study attempts to improve the model's robustness in real-world scenarios through a series of model tuning methods. The workflow of model tuning is illustrated in Figure S3. First, the research applies the 'RandomizedSearchCV', a model tuning and cross-validation method supported by 'Scikit-learn' (Pedregosa et al., 2011), to select the best hyperparameters to reduce the model's overfitting. For each parameter combination randomly selected, the oversampled training set is re-split into a training set and a validation set, and the model's performance under the selected hyperparameter setting is cross-validated. Secondly, the 'TimeSeriesSpilt' method is applied to split the training and validation set in a temporal order during the cross-validation. In real-world scenarios, the model is trained on historical complaint data to predict future complaints. Splitting the training and validation set in a temporal order can enhance the model's robustness on new data.

Exploratory Analysis on Noise Complaint Dataset

Noise Source Types and Timestamps of Complaints

The study first investigates the general temporal distribution of noise complaints from April 1st 2018, to March 31st 2022, based on the time series constituted of the daily count of complaints. As shown in Figure S4, there were seasonal fluctuations in the count of noise complaints; since the outbreak of COVID-19 in the 2020 spring, the fluctuations in the number of noise complaints have become more significant; noise complaints showed an increasing trend during each lock-down period. In terms of the complaint type, there were 40670 noise complaints from residential premises during 2018-2022, which was more than twice the complaint count from streets (2nd, 17728) and commercial premises (3rd, 11621). The counts of complaints from building sites (4th, 7392) and property alarms (5th, 3876) follow (Figure S5).

A fine-grained observation within Figure S6 compares the distribution of different noise complaints by minute in day and by hour in week between 2019 and 2020. It is found that complaints from building sites happened more frequently during the day-time, while complaints from residential premises and streets were relatively more frequent during night-time; the peak period of noise complaints in a week could be the midnight hours between Friday-Saturday and between Saturday-Sunday. In 2020, there was a significant decrease in day-time complaints from street and commercial premises compared to 2019. However, there was a significant increase in day-time and weekend complaints from residential premises. The pandemic-induced home isolation and work-from-home strategies may have increased people's sensitivity to noise complaints.

The Complaint History and Repeat Complaints

The complaint history on address and LSOA scales are investigated to learn the trend characteristics of repeat complaints. Regarding address-scale distribution, as shown in Figure S7, about 40% (11143) of all addresses have recorded two or more complaints, while the complaints in these addresses account for 80% (68581) of the total number of complaints. This suggests that more noise complaints tend to occur at addresses that already have noise complaints. Besides, for each noise complaint, the study searches the preceding complaints within 3 different previous time scales at the same address. The count of current complaints with preceding complaints in the previous 7 days accounted for 18.5% of all 81507 complaint records (Figure S8), which indicates a potential dependency between complaints at the same address within short time scales.

On the LSOA scale, as shown in Figure S9, relatively more noise complaints and complaint addresses came from the east of the City of Westminster. These areas cover London's famous commercial activity blocks. Figure S10 displays the correlation matrix between the noise source type for each complaint, the number of same-day but earlier complaints of different types in the LSOA, and the density of various POIs in the LSOA. It can be observed that the type of current noise complaints is positively correlated with the number of earlier noise complaints of the same type in the LSOA. It can be further inferred that noise complaints may also be interdependent in space due to common noise sources, and the land use and street activities in the vicinity may be closely related to the type of noise complaints.

Trend Clustering of Noise Complaints

Clustering Results

The clustering analysis hopes to explore the similarity and dissimilarity of the trend feature of noise complaints in LSOA, and reveal the potential impact of COVID-19. To highlight the trend features, the time series of the monthly noise complaint count is first smoothed with the rolling mean method with a fixed subset size of 6. Then the series is decomposed as seasonality, trend and residual, and the seasonality feature is removed from the series. With the smoothed series, the Multidimensional Scaling (MDS) method (Cox & Cox, 2008), is applied to reduce the dimension of original time series into 2 components. And through Silhouette analysis (Figure S11), 3 is selected as the optimal clustering number for a higher silhouette score and more clusters to compare with. The clustering results for the long-term trend of noise complaints, Cluster 0, Cluster 1 and Cluster 2, and the corresponding LSOA locations are shown in Figure 2. A comparison of centroid values of different clusters, regarding the counts of different noise complaint types, socio-economic features, and built environment densities is shown S12.

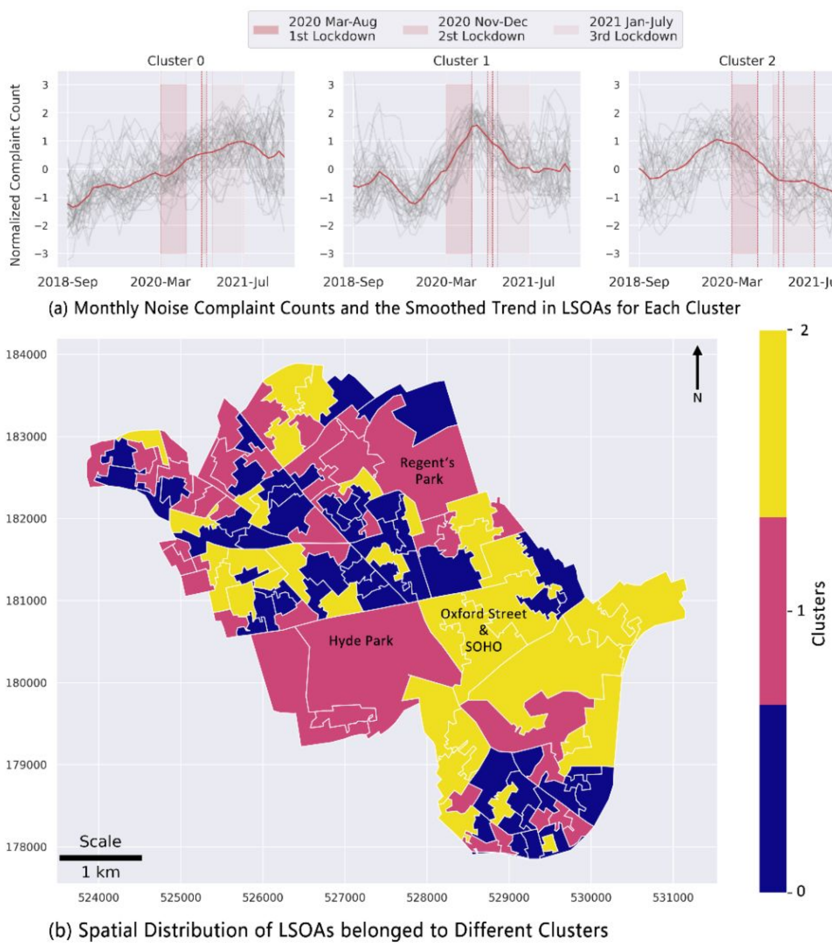


Figure 2 Results of K-Means (DTW) Clustering on Monthly Noise Complaint Count of LSOAs

It has been observed that noise complaints in different clusters show distinct responses to the outbreak of COVID-19 and related lock-down restrictions, and the responses can be further influenced by their specific social and spatial contexts. Specifically, noise complaints in Cluster 1 showed a rapid increase during the first COVID-19 lock-down in 2020, and there is a continuous decreasing trend after the ease of the first lock-down restrictions, until the end of all lock-down restrictions. LSOAs in Cluster 1 cover more open and leisure space in Westminster, and are featured with most residential premises complaints. It is assumed that the risk of COVID-19 may have driven urban residents to choose green spaces more for outdoor activities, leading to an increase in complaints from residents living near these green spaces. Similar to Cluster 2, noise complaints in Cluster 0 have a rapid increase during the first lock-down period, but they remained on an upward trend until the end of the third lock-down. LSOAs belonging to Cluster 0 are mainly located in the northern and southern parts of Westminster, where communities with higher residential mobility, higher built environment density, fewer educational and medical facilities, and lower incomes are situated. Based on previous findings by Tong et al. (2021) and Ramphal et al. (2022), it can be inferred that the relatively disadvantaged socio-economic conditions of these communities may have contributed to an increase in noise complaints during the lock-down restrictions.

In contrast, there has been a stable decrease in noise complaints in Cluster 2 since the COVID-19 outbreak, with a significant drop in complaints during the first lock-down period compared to the second and third ones. Cluster 2 corresponds to the main commercial functional areas of Westminster, with the highest housing and income levels, and is characterised by the most street complaints and points of interest (POI) density. The pandemic-related restrictions on service industries may have significantly reduced the level of activity in these areas, resulting in a reduction in various types of noise complaints related to street activities. Based on the above analysis, there can be significant spatial differentiation in the trend noise complaints in different LSOAs. The spatial differentiation can be driven by the impact of COVID-19 and is also related to the social and spatial characteristics of different LSOAs.

Predicting Repeat Noise Complaints

Model Tuning and Performance

This study aims to develop a random forest model based on the aforementioned multidimensional features associated with noise complaints. By using current complaints, the model predicts the likelihood of repeat complaints at the same address within different timeframes in the future. To enhance the model robustness, the 'RandomizedSearchCV' method is utilised to select the best parameter combination from a pre-set hyperparameter collection. The hyperparameter collection and the best hyperparameters are listed in Table S5.

Table S6 presents the confusion matrix of the model fitted with the best hyperparameters, and Figure S13 presents the precision-recall curves for classifying repeat complaints in different prediction tasks. It is found that the proposed prediction method can achieve high recall and classification accuracy for repeated complaints within 0-30 days, with scores of over 75%. However, for shorter-term prediction tasks such as predicting in the next 2 days or in the next week, the model performs better in identifying non-repeated complaints and weaker in identifying repeated complaints. This deficiency of the model may still originate from the imbalance of repeated and non-repeated samples in the training dataset. Nevertheless, this study has demonstrated that there is significant potential to predict repeat complaints based on complaint-related historical information, as well as the spatial and social contexts of the complaint location.

Feature Importance

Figure 3 illustrates the importance scores of different features in three tuned models that predict repeat complaints in the next 0-2nd days, 0-7th days, and 0-30th days. It is found that the noise source type of complaint and the address' complaint history can be the most important features in classifiers across all prediction tasks. The general importance ranking followed can be the complaint's timestamp, the complaint's relation with COVID-19 lock-down, the short-term complaint history in LSOA, the complaint trend clustering the LSOA belong to, and LSOA's socio-economic and built-environment contexts. Our analysis reveals that variables responding to trends and seasonality, and characterising the fine-granularity temporal and spatial characteristics of noise complaints, generally have higher importance in the classifier. In contrast, exogenous variables that characterise the long-term spatial and socio-economic characteristics around the complaint address have a relatively limited contribution to the model performance.

In addition, the importance of certain features can vary greatly depending on the time scale of the prediction task. For instance, as the prediction time scale expands from 0-2 days to 0-30 days, the significance of variables representing the complaint history of the location increases significantly. Specifically, variable that represent the total number of historical complaints at the address are of the utmost importance in the 0-7 day and 0-30 day models, and its importance is much higher than that of noise source type and short-term historical features of the address and LSOA. The long-term complaint history of the address may imply whether the address is in an environment susceptible to noise nuisances, or whether the complainant at the address is more sensitive to noise. Addresses with these characteristics can be affected by temporary nuisance events, and are also likely to repeat noise complaints over a longer time scale in the future. In contrast, distinguishing whether a complaint occurred during a lock-down period is more important for predicting repeated complaints within 0-2 and 0-7 days, and the importance of the variable in the model decreases as the prediction time scale increases. It can be speculated that the lock-down may have amplified residents' sensitivity to temporary noise disturbance events in the short term, but its impact may not sustain.

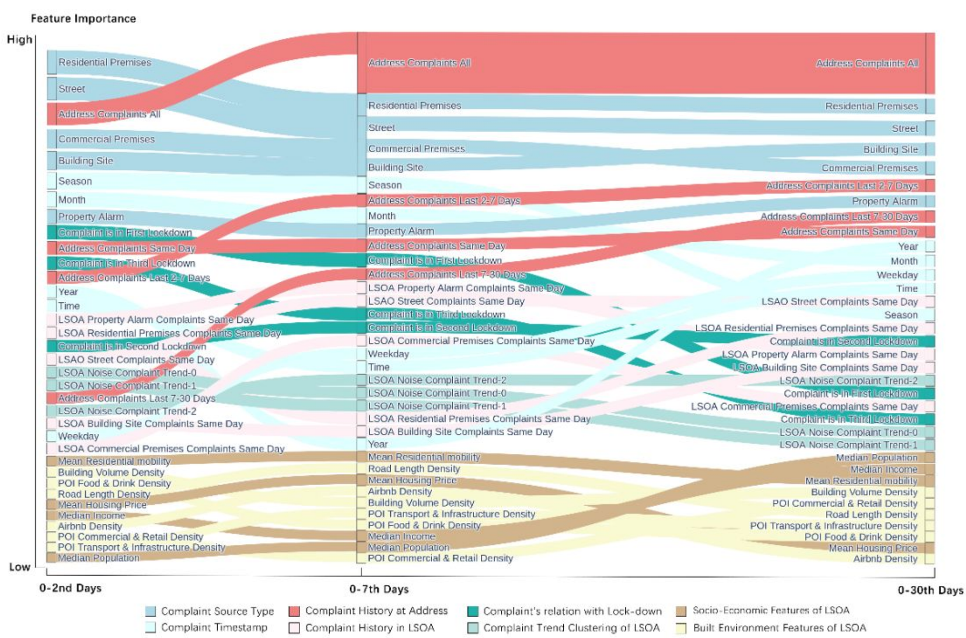


Figure 3 Feature Importance Ranking for Different Repeat Complaint Prediction Tasks

Conclusion

Findings

This study explores the spatial and temporal distribution characteristics and evolutionary trends of noise complaints in Westminster, UK, and verifies the feasibility of predicting repeat noise complaints with random forest classifiers. It is found that there are specific high-frequency periods and areas for noise complaints, and the temporal pattern of complaints of different types may vary significantly. There is a potential temporal dependency between current noise complaints and historical complaints happening at the same address. 18.5% of new complaints have at least one preceding complaint in the last 7 days. While on the LSOA scale, there appears to be a correlation between a current complaint's type and the number of earlier complaints of the same type in the same day. It is possible for the same noise source or nuisance event to trigger multiple noise complaints in nearby locations and times. According to the clustering analysis, the COVID-19 lock-down may have significantly affected the temporal-spatial pattern of noise complaints in different LSOAs. Specifically, areas with higher incomes and housing prices (Cluster 2) commonly experienced a decline in noise complaints during the lock-down periods. However, in areas which are relatively disadvantaged in social and economic features (Cluster 0 & 1), there are upward trends of noise complaints. The finding is consistent with the observation of Tong et al. (2021) and Ramphal et al. (2022) on the urban scale, indicating that the spatial distribution and long-term change of noise complaints can be related to the spatial and social contexts of local space. Besides, it is also found that the effect of lock-down can gradually subside after the ease of lock-down restrictions.

Multiple random forest classifiers are trained to predict the likelihood of repeat noise complaints for a given complaint and its address in 0-2nd days, 0-7th days or 0 -30th days in the future. With hyperparameter tuning and cross-validation, the F1 score of the tuned model for 0-30th days' repeat complaint prediction can be as high as 0.75. While it decreases as the prediction time scale becomes shorter and the prediction scenario becomes more precise. For feature importance, the history count of complaints in the address, and the type of the current complaint play the most important roles in all the models. The fine granularity variables associated with the trend and seasonality of noise complaints play a more important role in predicting repeated complaints compared to the local socio-economic and built environment features of the noise complaint area.

Suggestions for Noise Complaint Governance

Based on the findings of this study, policy recommendations are proposed for managing noise complaints by optimising resource allocation and precise governance. Firstly, based on the temporal and spatial distribution patterns of different types of noise complaints, the long-term trends of complaint numbers in the local area, and socio-economic and built-environmental characteristics of the area, the corresponding noise complaint pattern zones can be established for spatial management optimisation. For example, there are significant differences in the main types of noise complaints between commercial activity-intensive areas and residential-intensive areas in Westminster, which may lead to further differences in the distribution of complaint time; the changes in noise quantity between the two during the COVID-19 lock-down period show almost opposite trends, and the sensitivity of residents to noise nuisances may also differ. Based on these differences, appropriate resource allocation can be made for areas that are susceptible to noise nuisances at the policy level. In specific noise investigation scenarios, the spatial arrangement of complaint investigation personnel and vehicles can be dynamically adjusted based on the current time and dominant complaint type, reducing the additional time consumption when investigators travel to noise complaint locations.

Secondly, this study has demonstrated that the complaint history of noise source types and addresses plays the most significant role in predicting repeated noise complaints. In particular, understanding the short-term and long-term complaint history of an address can help identify whether the address is in an area vulnerable to noise disturbance or has prominent sensitivity to noise. Authorities can try to use the complaint history of addresses and other more accurate property attribute data to identify and label addresses with abnormal complaint behaviour and gradually and targetedly eliminate noise nuisance risks around the addresses, such as re-planning the business premises' operating hours near high-frequency complaint addresses or introduce regeneration to the noise-sensitive properties.

Innovations and Limitations

This study's innovation lies in that a prediction method for repeat noise complaints is proposed for the first time. By identifying the history time course of noise complaints at different spatial and temporal scales, the method tries to compensate for the impact of the lack of detailed address information in the complaint data due to privacy issues. However, it should be acknowledged that the model's short-term prediction capability for repeated complaints is still not satisfactory. By introducing some key attributes of complaint addresses and other dynamic urban activity information at the time of complaint, it is hoped that the model's performance can be further improved.

Reference

Agonafir, C. *et al.* (2022) 'A machine learning approach to evaluate the spatial variability of New York City's 311 street flooding complaints', *Computers, Environment and Urban Systems*, 97, p. 101854. Available at: https://doi.org/10.1016/j.compenvurbsys.2022.101854.

Box, G. & Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day. Available at: http://entsphere.com/pub/pdf/Papers/Signal%20Processing/Box,%20Time%20Series%20Analysis,%20Forecasting%20and%20control%20BOOK.pdf.

Bravo-Moncayo, L. *et al.* (2022) 'Traffic noise and property values: an instrumental variable strategy for hedonic valuation', *Journal of Environmental Planning and Management*, 0(0), pp. 1–20. Available at: https://doi.org/10.1080/09640568.2022.2079079.

Breiman, L. (1996) 'Bagging predictors', *Machine Learning*, 24(2), pp. 123–140. Available at: https://doi.org/10.1007/BF00058655.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, 45(1), pp. 5–32. Available at: https://doi.org/10.1023/A:1010933404324.

Chen, C. & Breiman, L. (2004) 'Using Random Forest to Learn Imbalanced Data', *University of California, Berkeley.* Available at: https://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf

Chen, N. *et al.* (2021) 'KE-CNN: A new social sensing method for extracting geographical attributes from text semantic features and its application in Wuhan, China', *Computers, Environment and Urban Systems*, 88, p. 101629. Available at: https://doi.org/10.1016/j.compenvurbsys.2021.101629.

Chen, S. *et al.* (2022) 'An Intelligent Government Complaint Prediction Approach', *Big Data Research*, 30, p. 100336. Available at: https://doi.org/10.1016/j.bdr.2022.100336.

Chawla, N.V. *et al.* (2002) 'SMOTE: Synthetic Minority Over-sampling Technique', *Journal of Artificial Intelligence Research*, 16, pp. 321–357. Available at: https://doi.org/10.1613/jair.953.

Clean Neighbourhoods and Environment Act 2005, c. 16. Available at: https://www.legislation.gov.uk/ukpga/2005/16/part/7

Control of Pollution Act 1974, c. 40. Available at: https://www.legislation.gov.uk/ukpga/1974/40

Cox, M.A.A. & Cox, T.F. (2008) 'Multidimensional Scaling', in C. Chen, W. Härdle, and A. Unwin (eds) *Handbook of Data Visualization*. Berlin, Heidelberg: Springer (Springer Handbooks Comp.Statistics), pp. 315–347. Available at: https://doi.org/10.1007/978-3-540-33037-0_14.

Cryer, J.D. & Chan, K.-S. (2008) *Time Series Analysis*. New York, NY: Springer (Springer Texts in Statistics). Available at: https://doi.org/10.1007/978-0-387-75959-3.

Dietterich, T.G. (2000) 'An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization', *Machine Learning*, 40(2), pp. 139–157. Available at: https://doi.org/10.1023/A:1007607513941.

Dzhambov, A.M. & Dimitrova, DD (2016) 'Exposures to road traffic, noise, and air pollution as risk factors for type 2 diabetes: A feasibility study in Bulgaria', *Noise & Health*, 18(82), pp. 133–142. Available at: https://doi.org/10.4103/1463-1741.181996.

European Environment Agency. (2021). Environmental noise in Europe - 2020. *European Environment Agency*. Available at: https://www.eea.europa.eu/publications/environmental-noise-in-europe

Environmental Protection Act 1990, c. 43. Available at: http://www.legislation.gov.uk/ukpga/1999/28/contents

Fan, Y., Teo, H.P. and Wan, W.X. (2021) 'Public transport, noise complaints, and housing: Evidence from sentiment analysis in Singapore', *Journal of Regional Science*, 61(3), pp. 570–596. Available at: https://doi.org/10.1111/jors.12524.

Great Britain. Department for Environment, Food & Rural Affairs. (2015). Noise nuisances: how councils deal with complaints. Available at: https://www.gov.uk/guidance/noise-nuisances-how-councils-deal-with-complaints.

Hartigan, J.A. & Wong, M.A. (1979). 'Algorithm AS 136: A K-Means Clustering Algorithm', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), pp. 100–108. Available at: https://doi.org/10.2307/2346830.

Kang, J. (2006). Urban Sound Environment. CRC Press, Boca Raton, FL.

Liu, Y. *et al.* (2015). 'Social Sensing: A New Approach to Understanding Our Socio-economic Environments', *Annals of the Association of American Geographers*, 105(3), pp. 512–530. Available at: https://doi.org/10.1080/00045608.2015.1018773.

Münzel, T. *et al.* (2018). 'Environmental Noise and the Cardiovascular System', *Journal of the American College of Cardiology*, 71(6), pp. 688–697. Available at: https://doi.org/10.1016/j.jacc.2017.12.015.

Noise Act 1996, c. 37. Available at: https://www.legislation.gov.uk/ukpga/1996/37/section/2

Office for National Statistics (ONS) (2023a). *How life has changed in Westminster: Census 2021*. Available at: https://www.ons.gov.uk/visualisations/censusareachanges/E09000033/ (Accessed: 6 May 2023).

Office for National Statistics (ONS) (2023b). *Population density - Census Maps.* Available at: https://www.ons.gov.uk/census/maps/choropleth/population/population-density/population-density/persons-per-square-kilometre (Accessed: May 2nd 2023).

Osorio-Arjona, J. *et al.* (2021). 'Social media semantic perceptions on Madrid Metro system: Using Twitter data to link complaints to space', *Sustainable Cities and Society*, 64, p. 102530. Available at: https://doi.org/10.1016/j.scs.2020.102530.

Ottoz, E., Rizzi, L. & Nastasi, F. (2018). 'Recreational noise: Impact and costs for annoyed residents in Milan and Turin', *Applied Acoustics*, 133, pp. 173–181. Available at: https://doi.org/10.1016/j.apacoust.2017.12.021.

Parr, T. & Turgutlu, K. (2022). 'Feature importances for scikit-learn machine learning models'. Available at: https://github.com/parrt/random-forest-importances (Accessed: August 21st 2022).

Pedregosa, F. *et al.* (2011). 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, 12(85), pp. 2825–2830.

Peng, X. *et al.* (2022). 'A social sensing approach for everyday urban problem-handling with the 12345-complaint hotline data', *Computers, Environment and Urban Systems*, 94, p. 101790. Available at: https://doi.org/10.1016/j.compenvurbsys.2022.101790.

Petitjean, F., Ketterlin, A. & Gançarski, P. (2011). 'A global averaging method for dynamic time warping, with applications to clustering', *Pattern Recognition*, 44(3), pp. 678–693. Available at: https://doi.org/10.1016/j.patcog.2010.09.013.

Ramphal, B. *et al.* (2022). 'Noise complaint patterns in New York City from January 2010 through February 2021: Socio-economic disparities and COVID-19 exacerbations', *Environmental Research*, 206, p. 112254. Available at: https://doi.org/10.1016/j.envres.2021.112254.

Rodriguez-Galiano, V.F. *et al.* (2012). 'An assessment of the effectiveness of a random forest classifier for land-cover classification', *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, pp. 93–104. Available at: https://doi.org/10.1016/j.isprsjprs.2011.11.002.

Sadiq, R. *et al.* (2022). 'Integrating remote sensing and social sensing for flood mapping', *Remote Sensing Applications: Society and Environment*, 25, p. 100697. Available at: https://doi.org/10.1016/j.rsase.2022.100697.

Sakoe, H. & Chiba, S. (1978). 'Dynamic programming algorithm optimisation for spoken word recognition', *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), pp. 43–49. Available at: https://doi.org/10.1109/TASSP.1978.1163055.

Tavenard, R. *et al.* (2020) 'Tslearn, A Machine Learning Toolkit for Time Series Data', *Journal of Machine Learning Research*, 21(118), pp. 1–6. Available at: https://jmlr.csail.mit.edu/papers/v21/20-091.html.

Tong, H. & Kang, J. (2021a). 'Characteristics of noise complaints and the associations with urban morphology: A comparison across densities', *Environmental Research*, 197, p. 111045. doi:10.1016/j.envres.2021.111045.

Tong, H. & Kang, J. (2021b). 'Relationships between noise complaints and socio-economic factors in England', *Sustainable Cities and Society*, 65, p. 102573. doi:10.1016/j.scs.2020.102573.

Tong, H. *et al.* (2021). 'Increases in noise complaints during the COVID-19 lock-down in Spring 2020: A case study in Greater London, UK', *Science of The Total Environment*, 785, p. 147213. doi:10.1016/j.scitotenv.2021.147213.

World Health Organization. (2011). *Burden of disease from environmental noise: quantification of healthy life years lost in Europe*. World Health Organization. Regional Office for Europe. https://apps.who.int/iris/handle/10665/326424.

Westminster City Council. (2009). Noise Issues and Options Report Consultation. Community Protection Department Westminster City Council. Available at: https://www.westminster.gov.uk/westminster-noise-strategy.

Westminster City Council. (2018). City of Westminster Profile 2018. Available at: https://www.westminster.gov.uk/media/document/city-of-westminster-profile.

Zambon, G. et al. (2020). 'Analysis of noise annoyance complaints in the city of Milan, Italy', *Journal of Physics: Conference Series*, 1603(1), p. 012029. Available at: https://doi.og/10.1088/1742-6596/1603/1/012029

Westminster City Council. (2021). The Westminster Noise Strategy. *Westminster City Council*. Available at: https://www.westminster.gov.uk/westminster-noise-strategy.

Westminster City Council. (2022). *Investigating your complaint*. Available at: https://www.westminster.gov.uk/planning-building-and-environmental-regulations/environmental-policies-regulations-and-guidance/noise-pollution/investigating-your-complaint.

Yildirim, Y. & Arefi, M. (2021). 'Noise complaints during a pandemic: A longitudinal analysis', *Noise Mapping*, 8, pp. 108–115. Available at: https://doi.org/10.1515/noise-2021-0008.

Young, J.C. *et al.* (2022). 'Social sensing of flood impacts in India: A case study of Kerala 2018', *International Journal of Disaster Risk Reduction*, 74, p. 102908. Available at: https://doi.org/10.1016/j.ijdrr.2022.102908.

Supplementary Materials

Data and Objects

Figure S1 The Location of the City of Westminster Borough in London

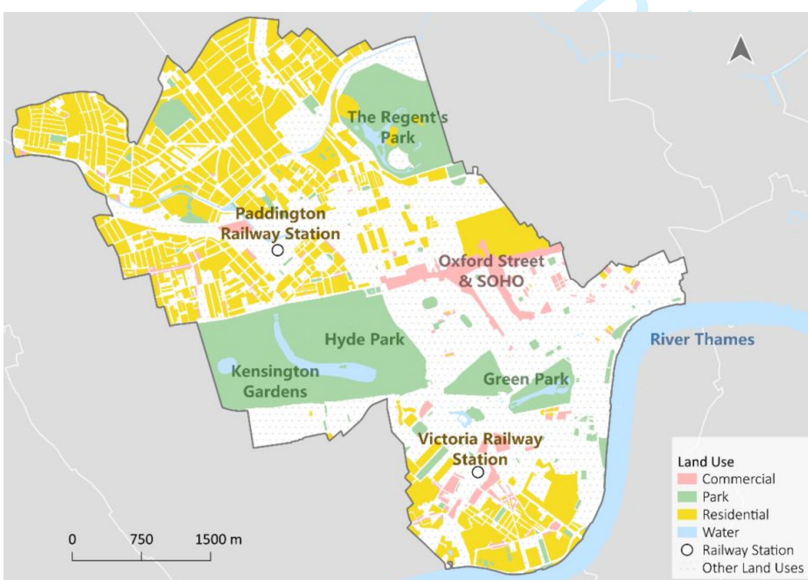


Figure S2 General Distribution of Commercial, Residential and Green Space

Column Name	Column Explanation
Noise Complaint Index	Unique index of each complaint record, e.g. NC1,NC2, ……
Received Date	Timestamp of each noise complaint (D/M/Y), e.g. 4/1/2018, ……
Received Time	Timestamp of each noise complaint (H:M:S) e.g. 0:11:00, ……
Noise Source Types	The noise source a complaint is report for e.g. Residential Premises, Commercial Premises, Street, Property Alarm, Building Site
Complainant Address Index	Anonymous address ID replacing the detailed address of Complainants.
LSOA 2011 Code	Lower Super Output Area the complaint is located in.

Table S1 Column Name and Explanation of the Noise Complaint Dataset

Data Source	Column Name in the Study
CACI - Households & Total Income 2022 (Provided by Westminster City Council)	Median Income
Consumer Data Research Centre - Residential Mobility 2018-2019 https://data.cdrc.ac.uk/dataset/cdrc-residential-mobility-index	Mean Residential Mobility
Office for National Statistics - Housing Price Median Quarterly 2018 https://www.ons.gov.uk/search?q=lsoa	Mean Housing Price
Office for National Statistics - Population Prediction 2020 https://www.ons.gov.uk/search?q=lsoa	Median Population
Ordnance Survey – Lodon Building & Road Centre Line https://digimap.edina.ac.uk/roam/download/os	Building Volume Density
	Road Length Density
Inside Airbnb - London Airbnb Listing 2022 http://insideairbnb.com/	Airbnb Density
Ordnance Survey – Point of Interest 2022 https://digimap.edina.ac.uk/roam/download/os	POI Transport & Infrastructure Density
	POI Food & Drink Density
	POI Education & Health Density
	POI Commercial & Retail Density

Table S2 Other Socio-economic and Built Environment Data in the Study

Methods

A. Noise Source and Timestamp	
Complaint Timestamp	Year
	Month
	Weekday
	Season
	Time
Noise Source Type	Building Site
	Commercial Premises
	Property Alarm
	Residential Premises
	Street
B. Historical Complaint Count	
Historical Complaints of Address	Address Complaints Same Day
	Address Complaints Last 2-7 Days
	Address Complaints Last 7-30 Days
	Address Complaints All
Historical Complaints of LSOA	LSOA Building Site Complaints Same Day
	LSOA Commercial Premises Complaints Same Day
	LSOA Property Alarm Complaints Same Day
	LSOA Residential Premises Complaints Same Day
	LSOA Street Complaints Same Day
C. COVID-19 Impact	
Complaint's relation with Lock-down	Complaint is in First Lockdown
	Complaint is in Second Lockdown
	Complaint is in Third Lockdown
LSOA Complaint Trend Clusters	LSOA is in Complaint Trend Cluster 0
	LSOA is in Complaint Trend Cluster 1
	LSOA is in Complaint Trend Cluster 2
D. LSOA Spatiotemporal Context	
LSOA Socio-Economic Features	Mean Residential Mobility
	Mean Housing Price
	Median Income
	Median Population
LSOA Built Environment Features	Building Volume Density
	Road Length Density
	Airbnb Density
	POI Transport & Infrastructure Density
	POI Food & Drink Density
	POI Education & Health Density
	POI Commercial & Retail Density

Table S3 Dependent Variables

Variables to Predict	Value / Label (Thresholds)	Count
Repeat Complaints within 0-2nd days	True (Complaint Count > 0)	17274
	False (Complaint Count = 0)	64233
Repeat Complaints within 0 -7th days	True (Complaint Count > 0)	26284
	False (Complaint Count = 0)	55223
Repeat Complaints within 0 -30th days	True (Complaint Count > 0)	36790
	False (Complaint Count = 0)	44717

Table S4 Variables to Predict in the Random Forest Model

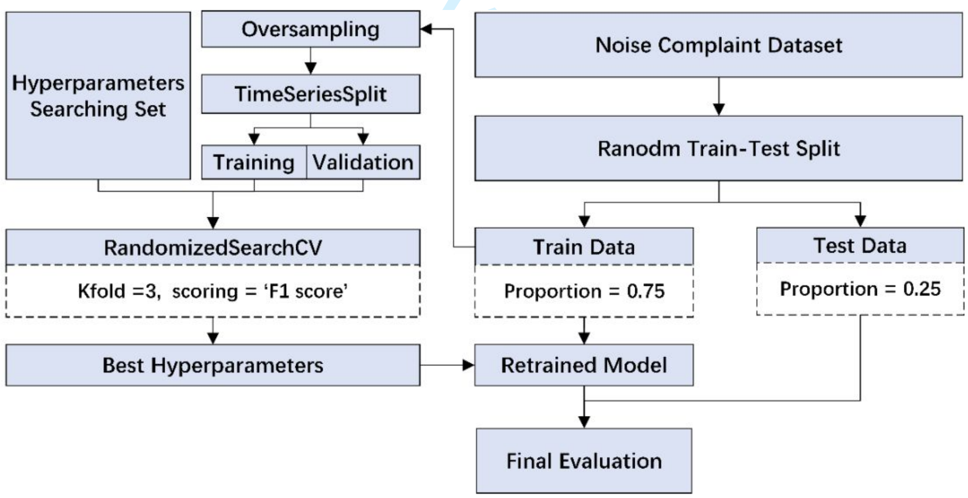


Figure S3 Workflow of Model Tuning with RandomizedSearchCV

Temporal - Spatial Distribution of Noise Complaints

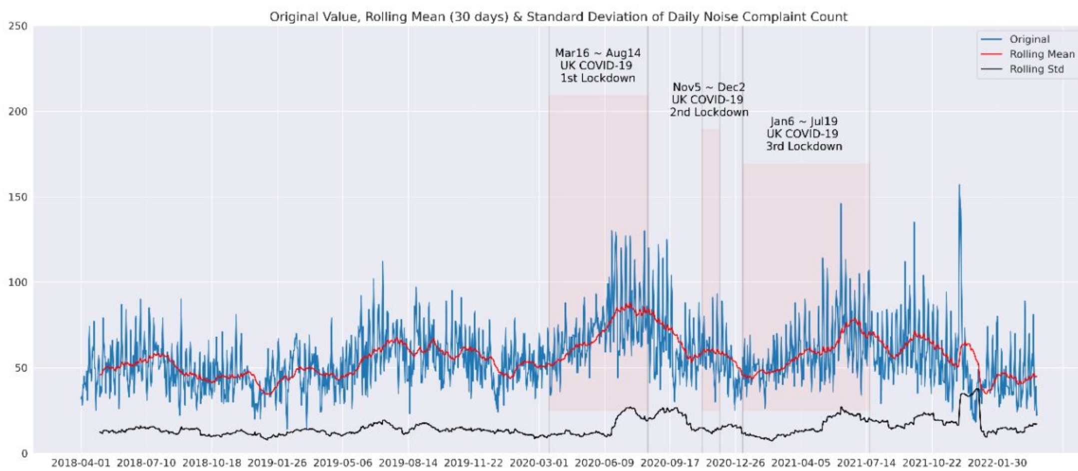


Figure S4 Count of Noise Complaint by Date and

the Rolling Mean and Standard Error values with a Fixed Subset Size of 30 Days

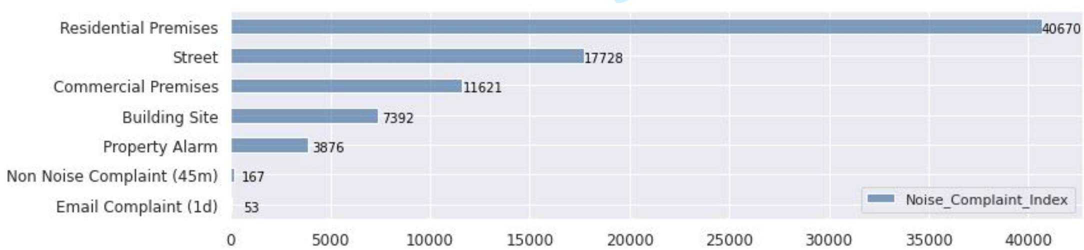
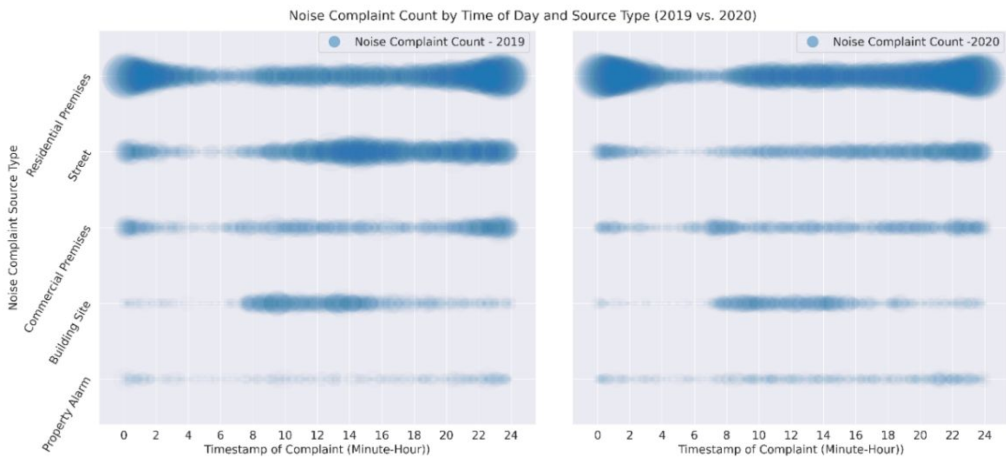
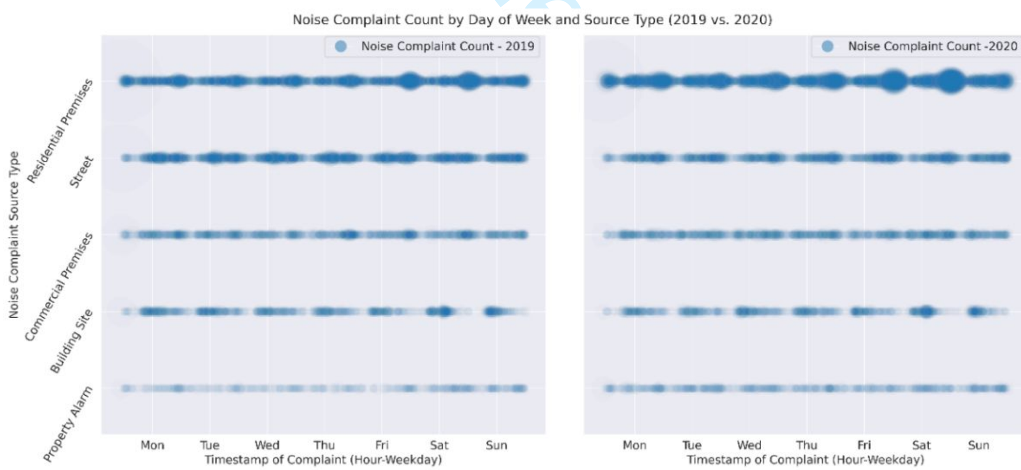


Figure S5 Cumulative Distribution of Noise Complaints in Different Types



a. Complaints in 1440 mins(24h) One Day



b. Complaints in 168 hours(7days) One Week

Figure S6 Cumulative Distribution of Different Noise Complaints in One Day and in One Week

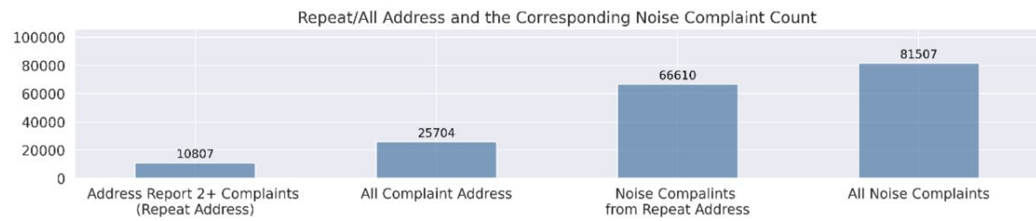


Figure S7 Count of All/ Repeat Complaints and Complaint Addresses

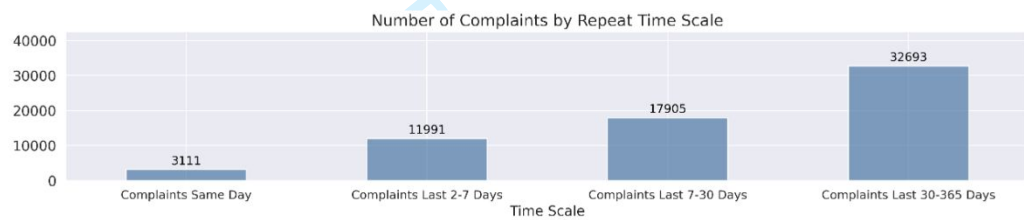


Figure S8 Counts of Complaints with other complaints within Earlier Time Scales

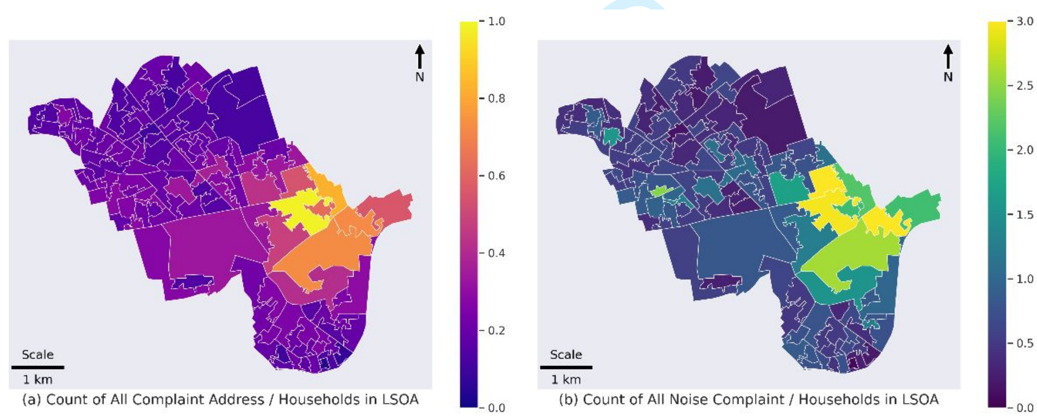


Figure S9 Spatial Distribution of Complaints / Complaint Address based on LSOA

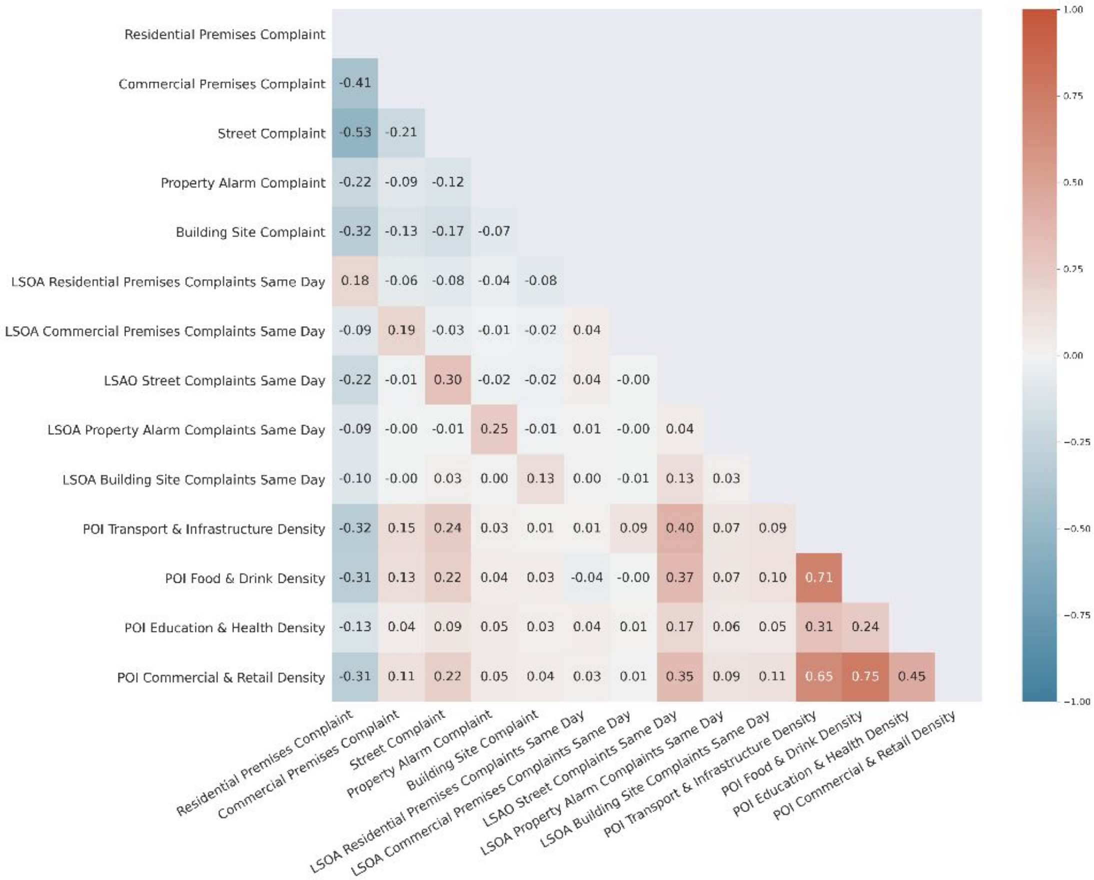


Figure S10 Correlation Matrix of Noise Source Types, LSOA Noise Complaint History and LSOA POI Count

Time Series Clustering of Noise Complaints

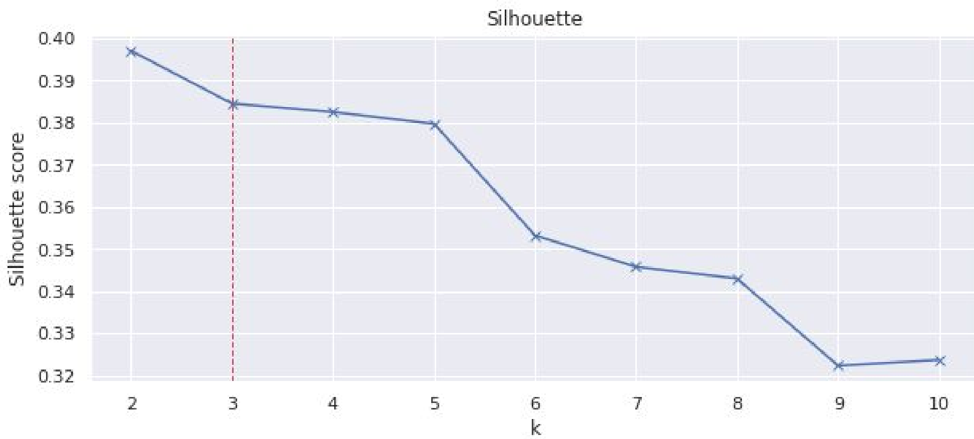


Figure S11 Silhouette Analysis for K-Means (DTW) Clustering based on the Time Series of Monthly Noise Complaint Count in Each LSOA

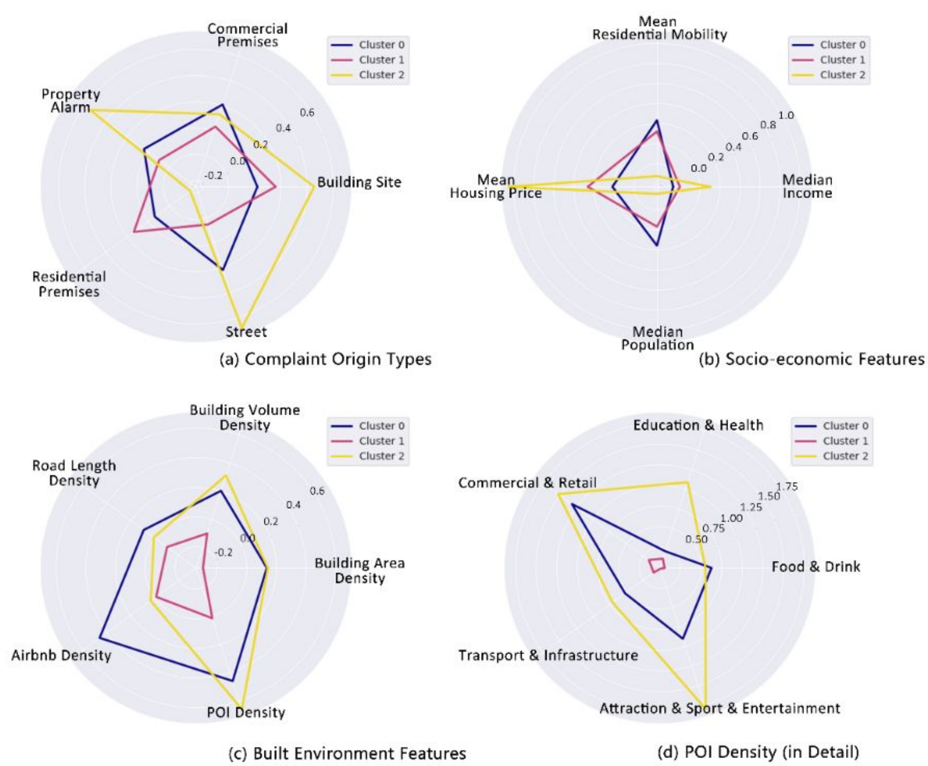


Figure S12 Cluster Centroid Comparison

Predicting Repeat Noise Complaints

Hyperparameters	Value	0-2 Days	0-7 Days	7-30 Days
bootstrap	True, False	False	False	True
max_depth	10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, None	30	30	60
max_features	'sqrt', 'log2', None	'log2'	'log2'	None
min_samples_leaf	1, 2, 4	1	1	2
min_samples_split	2, 5, 10	2	2	2
n_estimators	200, 400, 600, 800, 1000	400	400	1000

Table S5 Hyperparameter range and Best Hyperparameters Selected with TimeSeriesSplit Cross-Validation

Tasks	Classes	Performance Metrics			
0-2 Days Prediction		Precision	Recall	F1-score	Support
	0 (Non-repeat)	0.79	0.8	0.8	11123
	1 (Repeat)	0.76	0.75	0.75	9254
0-7 Days Prediction		Precision	Recall	F1-score	Support
	0 (Non-repeat)	0.83	0.86	0.84	13776
	1 (Repeat)	0.68	0.64	0.66	6601
0-30 Days Prediction		Precision	Recall	F1-score	Support
	0 (Non-repeat)	0.88	0.89	0.88	16088
	1 (Repeat)	0.56	0.55	0.55	4289

Table S6 Comparison of Model Performance between Different Tasks with Tuned Hyperparameters and on Test Dataset

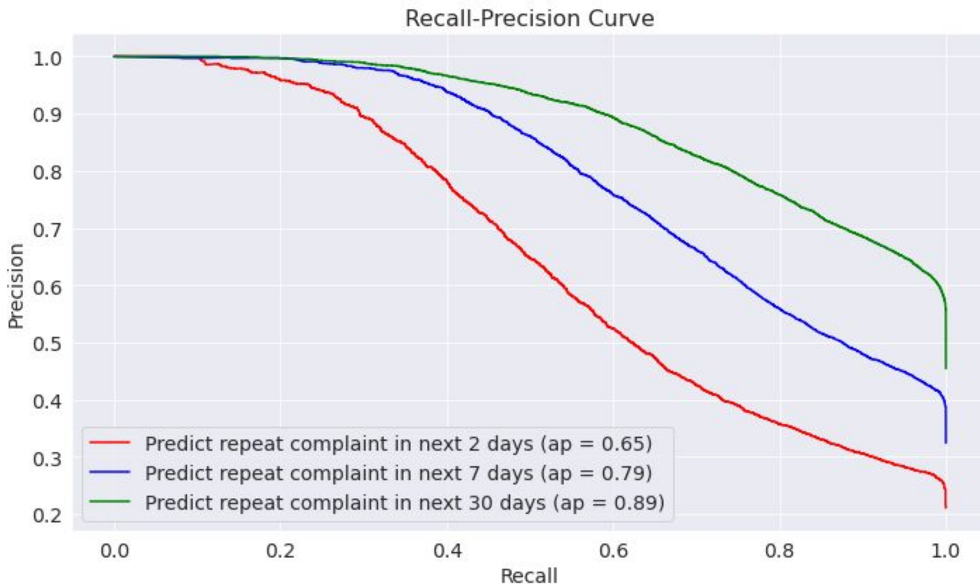


Figure S13 Recall-Precision Curve for tuned Model

Code

The code and analysis process in this project are available on GitHub: https://github.com/fzc961020/CASA0004