International Journal of Recent Innovations in Academic Research

E-ISSN: 2635-3040; P-ISSN: 2659-1561
Homepage: https://www.ijriar.com/
Volume-10, Issue-2, April-June-2026: 148-160

Research Article

Crime Rate Prediction by Applying Gradient Boost Regression to Real Estate Data: The Case of Avon, Connecticut

[a]Seongbin Cho and *[b]Bob Nam

[a]Avon Old Farms School, USA, [b]Seoul Innovations Research Institute, South Korea
*Corresponding Author Email: bobyep25@gmail.com

Article History
Received: May 30, 2026
Accepted: June 23, 2026
Published: June 30, 2026

Abstract
In Avon, Connecticut, USA, crime rates can be predicted by a Gradient Boost Regression model using housing advertisement datasets, along with other features such as changes in income and living costs, with an R^2 of 0.99. When living costs are the explanatory variable, the line plot of crime rate exhibits a reversed U-shape, indicating that the crime rate increases, peaks, and then decreases as living costs rise. This characteristic is captured by the linearly decreasing marginal contribution of living costs in the Partial Dependence Plot (PDP) for gradient-boost regression. Moreover, although the numerical data on crime rates are initially classified by location (e.g., Cherry Park, West Avon, Town Center, and Avon East), latitude is a key predictor of crime rates. Although feature importance is low, housing styles and construction materials are associated with crime rates. The linkage among multiple linear regression, PCA Biplot, feature importance, and PDP in Gradient Boost Regression can enhance the interpretability of machine learning outcomes.
Keywords: Avon, Connecticut, Crime Rate, Gradient Boost Regression, Feature Importance, Partial Dependence Plot, Multiple Linear Regression.

1. Introduction

To predict crime rates, many variables have been utilized. Socioeconomic variables like income level, employment conditions, education, housing, and urban form are typically used to predict crime rates. Demographic and social structures, such as age structure, gender composition, race or ethnicity composition, and immigrant or foreign-born share, as well as household and family structure, played an essential role in predicting crime rates. Population density, land-use patterns, and neighborhood characteristics are also used as environmental context.

To test the claim that crime rates can be predicted from daily real data, this study employs housing advertisement sales points, including size, number of bedrooms, number of bathrooms, number of car garages, areas, locations, and altitudes, as well as the styles and construction media. Due to the data availability, the study is limited to Avon, Connecticut, USA. Although there is no direct functional relationship between these housing characteristics and crime rates, even vague and inherent tendencies can enhance predictability. Therefore, these housing styles and construction materials are added as dummy variables after encoding.

After converting the numerical values into ranking statistics, the relationship between the crime rate and real estate price, changes in income level, and living costs in the Avon, Connecticut area is statistically tested. The dependence structure of the crime rate and its associated variables is investigated through multiple linear regression and PCA (Principal Component Analysis). To overcome the limitations posed by non-linear relationships among variables, machine learning-based Gradient Boost Regression is applied to capture feature implications.

2. Literature Review

Before machine learning was introduced for crime prediction, traditional models relied solely on historical crime data, assuming that crime events are near-repeated in space and time. That is, spatiotemporal information about historical crime events is used to predict the distribution of criminal activity at a later time [6, 7] Some examples are: near-repeat prediction model [9] and density estimation model [9,10].

With the rapid development of artificial intelligence (AI) in recent years, numerous machine-learning crime-prediction models have emerged. The representative models are: neural network model [11], a random forest model [12], and a graph convolutional model [13]. Although they may improve prediction accuracy, none of them have systematically considered spatiotemporal environmental variables, nor have they clearly explained the prediction process.

Real estate datasets often include locational, structural, and environmental features that are significant predictors of crime patterns, as they capture neighborhood environments that correlate with crime risk [1]. Several ML algorithms have been applied to crime prediction, including Random Forest, Decision Trees, Support Vector Machines, K-Nearest Neighbors, and deep learning architectures like Convolutional Long Short-Term Memory (ConvLSTM) networks. Ensemble methods, such as Random Forest, tend to perform well because they can capture complex interactions among features [2, 3].

Spatial and temporal elements are crucial in crime forecasting. Deep learning models that capture spatiotemporal dependencies have been proposed, using grids of geographic areas and real estate, sociodemographic, and mobility data to predict crime occurrences. These models show that integrating real estate data reflecting the built environment alongside human activity patterns enhances prediction accuracy [1,4].

Real estate data provides proxies for neighborhood stability, economic conditions, and opportunity structures that influence crime. For Avon, Connecticut, applying these methods would involve integrating local property data with crime records and, potentially, mobility patterns to model crime hotspots and temporal trends using machine learning [1, 4].

Many empirical studies have proved that many business or public facilities people routinely visit can be crime attractors or generators, which provide more potential crime opportunities and increase crime in their vicinity, such as restaurants [14], convenience stores [15], department stores [16], neighborhood parks [17], stadiums [18], alcohol outlets [19] and so forth. For example, schools are generally found to have a significant positive correlation with crime incidents at block-level studies [20].

In addition, scholars found an association between bus stops and street robberies [21], and a correlation between road network densities and property crime [22]. Regarding theft crimes, some scholars also found that motor vehicle theft was concentrated on facilities on commercial land [23]. A study [24] used official crime data from a large Chinese city to verify that theft rates were related to the presence of retail facilities that shape daily activities from the opportunity perspective [24].

Interpretable ML models have also been explored to reveal which real estate and environmental variables most influence crime risk, helping policymakers target interventions effectively [5]. To improve the transparency and interpretability of machine learning, two methods are proposed: a priori and a posteriori [25]. The a priori method uses a simple, interpretable structure to train models [26].

In contrast, the post hoc method allows models to be trained as usual and then assessed for interpretability afterward. Additionally, interpretability in the post hoc method can be divided into global and local interpretability. Global interpretability aims to help the audience understand the overall logic and mechanisms behind complex models [27], whereas local interpretability seeks to explain the decision-making process and the basis of the machine learning model for each input sample [28].

In summary, the literature supports the innovative use of real estate data combined with advanced ML techniques such as Random Forest and ConvLSTM to predict crime rates spatially and temporally. Applying these to Avon could leverage local housing and neighborhood data to understand and forecast crime patterns, enabling more effective law enforcement planning and public safety measures.

3. First Data Inspection

The following scaled data sets of crime rates, living costs, income change, and real estate prices in Avon demonstrate the sectoral classification of Avon.

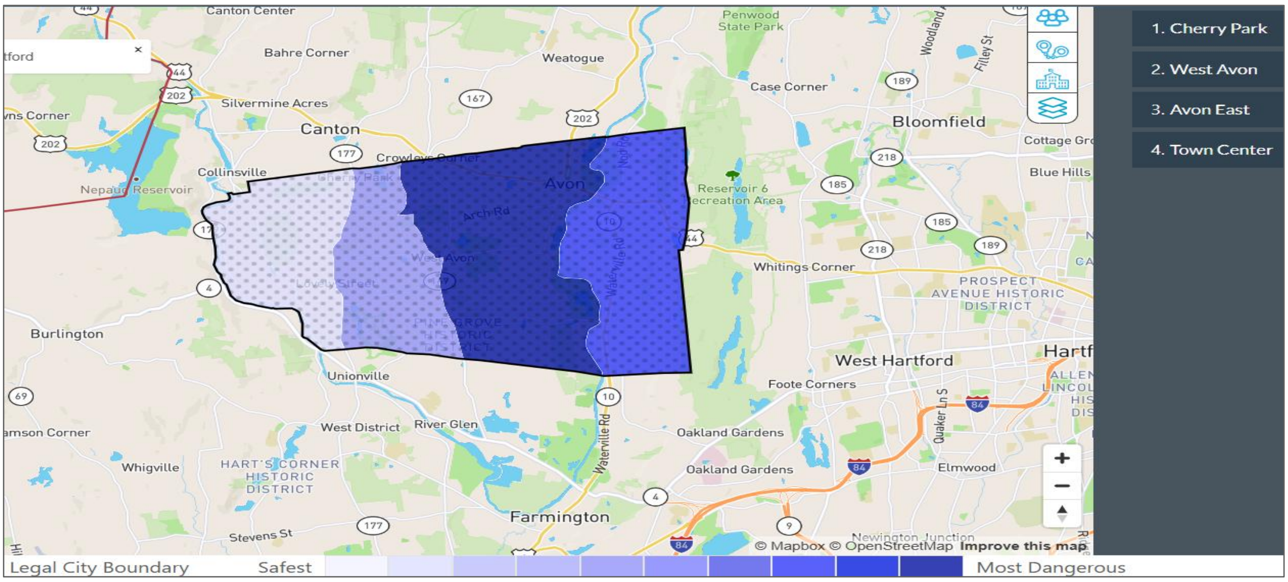

Figure 1. Crime rate in Avon, Connecticut https://www.neighborhoodscout.com/ct/avon/crime

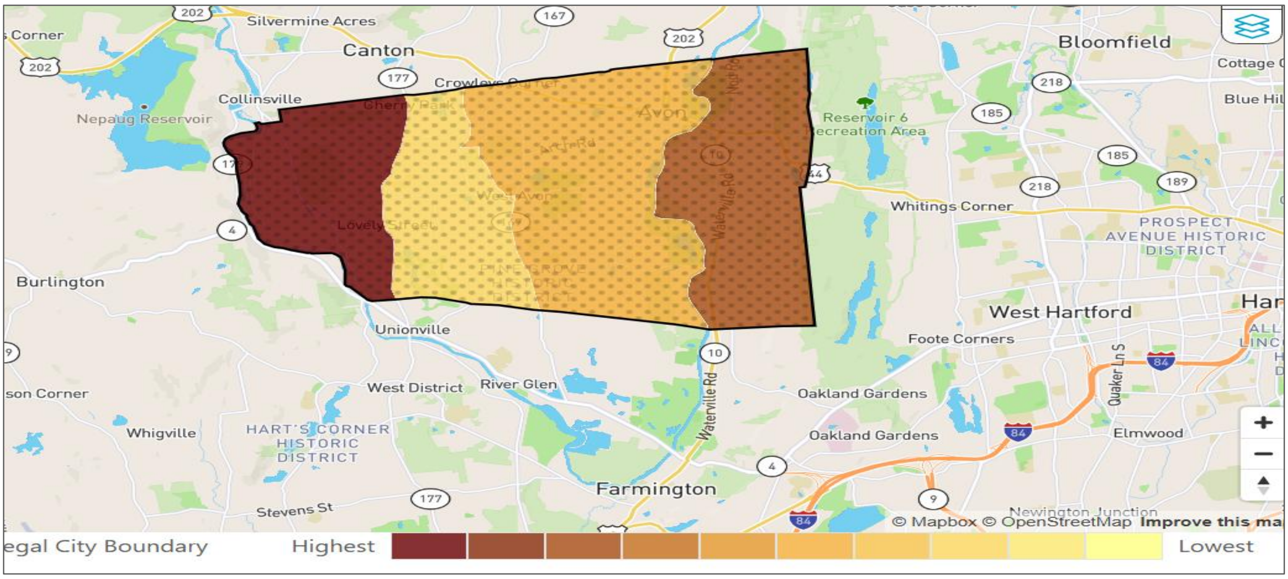

Figure 2. Living costs in Avon, Connecticut https://www.neighborhoodscout.com/ct/avon

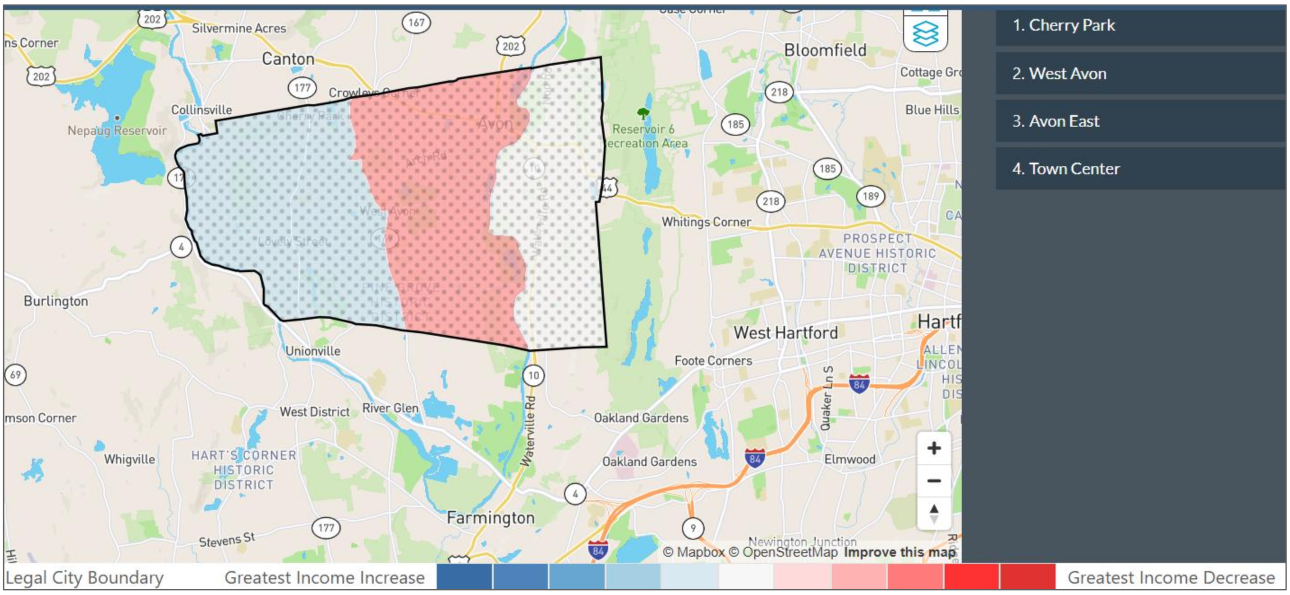

Figure 3. Income changes in Avon, Connecticut
https://www.neighborhoodscout.com/ct/avon/demographics

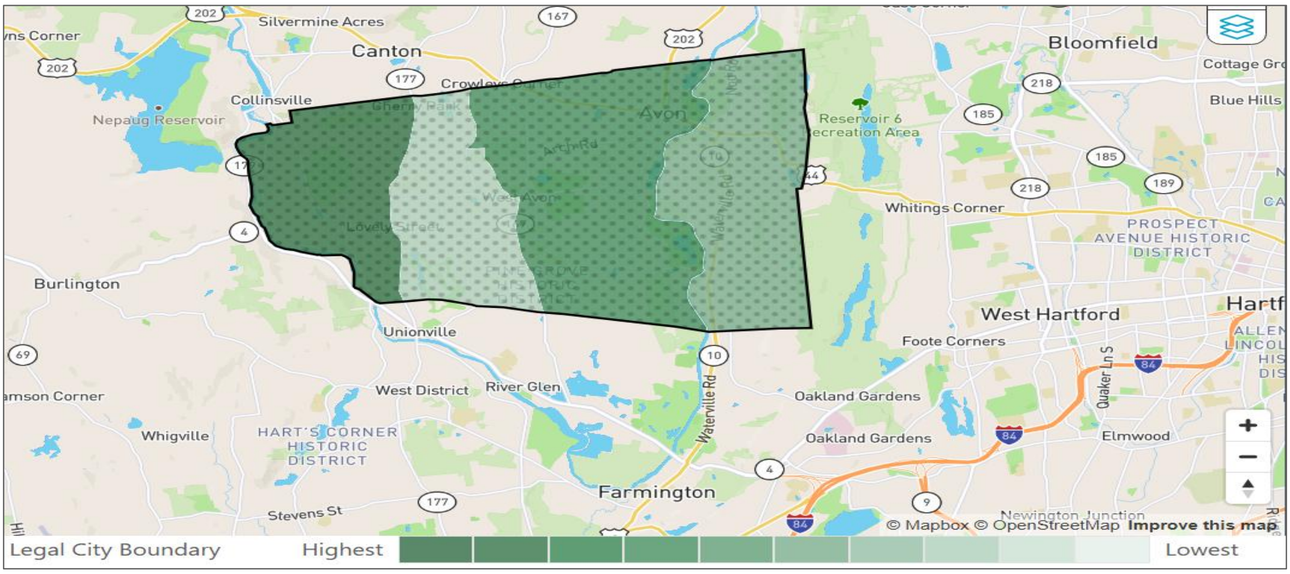

Figure 4. Real estate price in Avon, Connecticut https://www.neighborhoodscout.com/ct/avon/real-estate

Based on the above scaled data sets, the table for each ranking statistics are formed and χ^2 test for independence is conducted.

Where:

$$\chi^2 \; test \; statistic = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Table 1. Crime and real estate price rankings.

	Crime ranking	Real estate ranking	Total
Cheery Park	4	1	5
West Avon	3	4	7
Town Center	1	2	3
Avon East	2	3	5
Total	10	10	20
χ^2 $test\ statistic = 2.47619, p - value = 0.4796$			

Table 2. Crime and income change rankings.

	Crime ranking	Income change ranking	Total
Cheery Park	4	1	5
West Avon	3	1	4
Town Center	1	4	5
Avon East	2	3	5
Total	10	9	19
χ^2 $test\ statistic = 4.780, p - value = 0.19019$			

Table 3. Crime and living costs rankings.

	Crime ranking	Living costs	Total
Cheery Park	4	1	5
West Avon	3	4	7
Town Center	1	3	4
Avon East	2	2	4
Total	10	10	20
χ^2 $test\ statistic = 2.9428, p - value = 0.4005$			

According to the above Tables from 1 to 3, there is insufficient evidence to reject the null hypothesis of independence between the crime rate and other factors such as real estate, changes in income, and living costs. But this ranking data-based χ^2 test is too simplified to support crime prediction. Therefore, real estate advertisement data sets from August to September 2025 are gathered and transformed[1]. From the address of housing, longitude, latitude, and altitude are searched and added along with location dummy variables. Other characteristics, such as housing styles, construction materials, and prices, are included among the explanatory variables. Descriptive statistics of key variables are given in Table 4.

Table 4. Descriptive statistics of key features.

	Year built	bed	bath	area(sqft)	car garage	levels	Longitude	Latitude	Altitude(mete
count	79.000000	79.000000	79.000000	79.000000	79.000000	79.000000	79.000000	79.000000	79.00000
mean	1984.215190	3.670886	3.063291	3193.746835	2.126582	1.822785	41.472630	72.516794	128.2911
std	26.034332	1.058803	1.417309	2165.960262	0.965655	0.473922	0.009784	0.020343	46.4322
min	1890.000000	1.000000	1.000000	808.000000	0.000000	1.000000	41.454500	72.475900	62.00000
25%	1968.000000	3.000000	2.250000	1823.500000	2.000000	2.000000	41.464450	72.502350	93.50000
50%	1984.000000	4.000000	2.500000	2476.000000	2.000000	2.000000	41.473900	72.521100	115.00000
75%	2004.000000	4.000000	3.500000	3754.500000	3.000000	2.000000	41.481050	72.532000	149.50000
max	2025.000000	6.000000	7.500000	13490.000000	5.000000	3.000000	41.491500	72.545900	233.00000

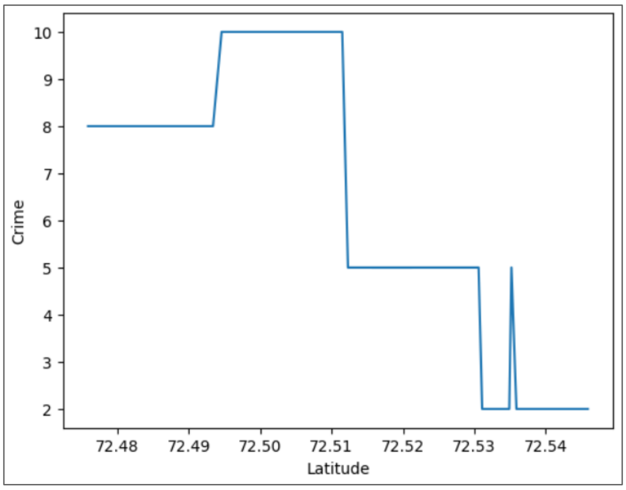

Figure 5. Line plot of crime vs latitude.

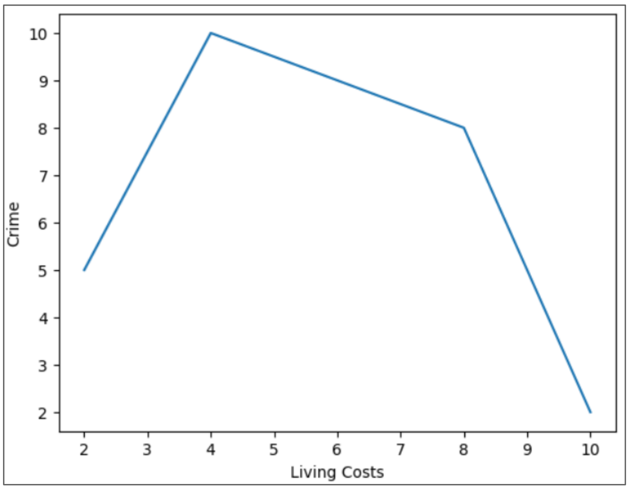

Figure 6. Line plot of crime vs living costs.

The line graphs of two crucial variables are given in Figures 1 and 2. When the latitude lies between 72.495 and 72.515, the crime rate increases, coinciding with the regional change from Cherry Park to West Avon. The interesting fact is that the line plot of the crime rate is a reversed U-shaped, meaning that the crime rate increases when living costs rise from 2 to 4 and then decreases thereafter.

4. Second Data Inspection: Multiple Linear Regression Mode

The Multiple Linear Regression outcome is given in the following equation. Adjusted R^2 is 0.975. The coefficients of all explanatory variables except the intercept are negative.

$$\text{Crime level} = 5185.275_{(211.082)} - 71.406_{(2.911)} Latitude - 1.366_{(0.053)}\, Income\ Change - 0.206_{(0.015)}\, Living\ Costs - 0.264_{(0.138)} Cape\ Cod$$

With 28 explanatory variables, including 18 dummy variables, the adjusted R^2 is remarkably high, although interpretation is not straightforward. Especially, if Cape Cod is incorporated into the housing style, crime rates tend to decrease, which is hard to find logical or theoretical reasons. Therefore, interpretability is not limited to the machine learning methods. A more serious limitation of this regression model is that it does not capture the nonlinear relationships among variables, as shown in the line plots in Figures 10-11 of the appendix.

5. Third Data Investigation: Factor Analysis

To identify the baseline factors governing the variability of all features, Principal Component Analysis (PCA) is applied, yielding the following Biplot and factor loadings.

[1]Data Source: https://www.realtor.com/realestateandhomes-search/Denver_CO/show-newest-listings/sby-6

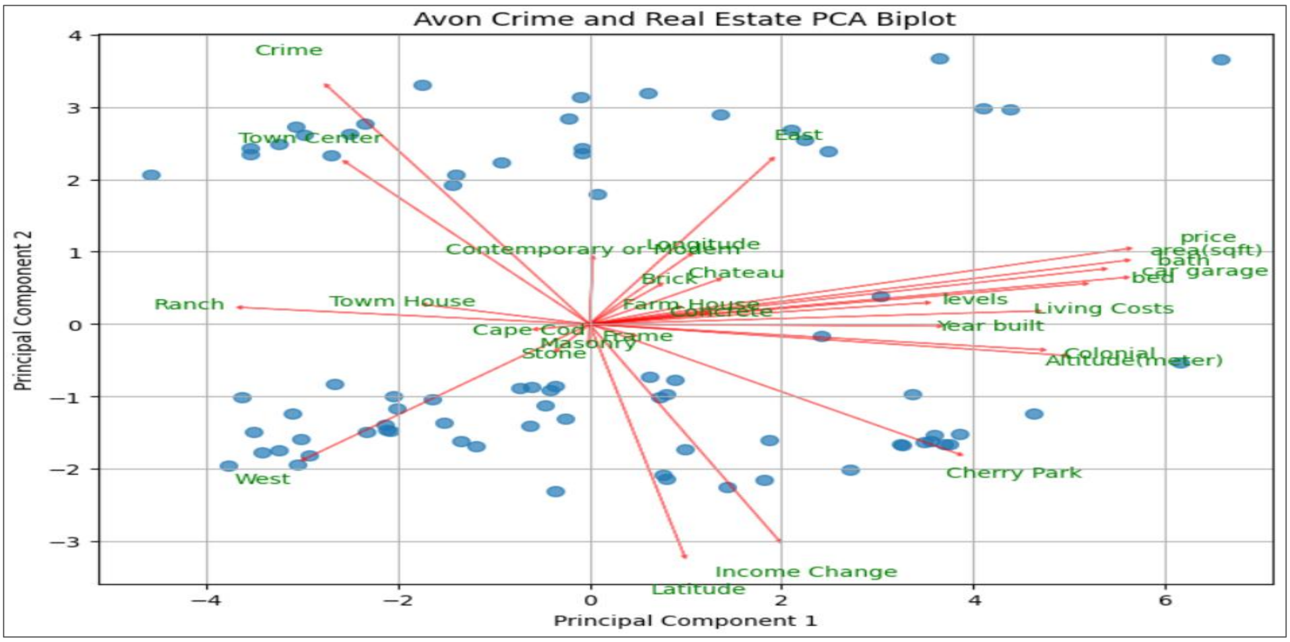

Figure 7. PCA biplot of crime and real estate in Avon, Connecticut.

Table 5. Factor loadings.

	First principal component	**First principal component**
Year built	0.552612383	-0.00724606597
Number of beds	0.784066485	0.151684008
Number of baths	0.813946032	0.207210625
Area (sq ft)	0.850680768	0.239672425
Number of car garage	0.849193049	0.175463462
Number of levels	0.533766440	0.0804670212
Longitude	0.158360139	0.259762782
Latitude	0.150198280	-0.869963743
Altitude	0.753304581	-0.117386090
East	0.288292738	0.618712554
Cherry Park	0.585470526	-0.488537741
West	-0.451757987	-0.507381375
Town center	-0.385381209	0.607497346
Income change	-0.415487989	0.894830866
Living costs	0.299908984	-0.812670531
Colonial	0.709227357	0.0495906459
Cape Cod	0.716963072	-0.0963629929
Ranch	-0.0827392171	-0.0186521388
Farm house	-0.551946135	0.0627401589
Town house	0.139883217	0.0626354246
Contemporary or modern	-0.258775117	0.0726303417
Chateau	0.00521737361	0.245530116
Brick	0.203176480	0.166867852
Masonry	0.110749064	0.147050964
Stone	-0.000556902833	-0.0633989119
Concrete	-0.0507414009	-0.0971542846
Frame	0.182575115	0.0416709621
Price	0.0680477316	-0.0403107330
Crime	0.853455089	0.283956221

The first principal component is positively correlated with the number of beds, the number of baths, area, and the crime rate, suggesting that crime opportunities are associated with wealth. The second principal component is negatively correlated with living costs and latitude, and positively correlated with income change. Therefore, the second principal component may be interpreted as region-specific conditions that are not limited to wealth level. The crime rate is negatively correlated with housing prices, the number of bedrooms, the number of bathrooms, and the number of car garages. Simultaneously, the crime rate is positively associated with the Town Center dummy variable. As is expected, the role of housing styles and construction materials is relatively small. Like the case of Multiple Linear Regression model, the interpretation of outcomes is not straightforward.

6. Final Data Inspection: Gradient Boost Regression

Gradient regression typically refers to using gradient-based optimization to train a regression model, often a linear regression, by adjusting model parameters iteratively to minimize the error between predicted and observed data [24]. It works by computing the gradient of a loss function, which quantifies prediction error, relative to the model's parameters, then moving these parameters in the direction that reduces error most efficiently, like rolling downhill toward the lowest point of a valley [29]. For linear regression, the loss function commonly used is mean squared error, and the updates to parameter values look like: $\theta_{new} = \theta_{old} - \eta \frac{\partial}{\partial \theta} Loss$ where θ is the parameter η is the learning rate, and the partial derivative is the gradient [24]. Gradient descent avoids heavy matrix operations and scales better with high-dimensional data, making it computationally efficient and feasible for big data scenarios [29]. Gradient descent avoids matrix inversion by updating parameters incrementally. To assess gradient boost regression performance, several measures are calculated:

RMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

R^2 (R^2 score) = $1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

MAE score = $\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$

Table 6. Output table of gradient boost regression.

```
MODEL GRADIENT BOOST REGRESSION
RMSE score for test: 6.136856156979707e-05
RMSE score for train: 7.842892374479272e-05
R2 score: 0.9999999991998252
MAE score: 4.911846566951761e-05
Variance score test: 0.9999999991998252
Variance score train: 0.9999999992944921
```

$Importance_j = \sum_{m=1}^{M}\sum_{k=1}^{T_m} I(v_{m,k} = j) \times \Delta L_{m,k} (s_{m,k})$

Where:

M = The total number of trees
T_m = the number of internal nodes in trees m
I (·) = indicator function which is 1 if the node k in tree m splits on feature j, and 0 otherwise.
$v_{m,k}$ = feature used for the split at node k in tree m
$\Delta L_{m,k}$ = reduction in the loss function such as mean squared error achieved by the split $s_{m,k}$

With these procedures, feature importance is calculated in Table 7.

Table 7. Feature importance of gradient boost regression.

Feature name	Feature importance	Feature importance ranking
Latitude	0.8080	1
West	0.0684	2
Living costs	0.0655	3
Cherry Park	0.0362	4
Town center	0.0129	5
Income change	0.0088	6
East	0.0002	7

The dummy variables, such as 'West', 'East', 'Cherry Park', and 'Town Center', essentially serve as the 'Longitude' variable, implying that housing location accounts for approximately 90 percent of the variation in crime rates. Only the remaining 10 percent is related to changes in income and living costs. That is, the outcome of Gradient Boost Regression is remarkably similar to that of Multiple Linear Regression, except for the variable 'Cape Cod'.

For further interpretation, partial dependence plot (PDP) is applied [29].

PDP is defined as:
$\hat{f}_S(x_S) = E_{X_C}[\hat{f}(x_S, X_C)] = \int \hat{f}(x_S, X_C)dP(X_C)$

Where:

x_S = variable to be plotted as PDP to capture its effect on dependent variable
X_C = variables to be held fixed except for x_S after learning is completed

By integrating out and thus marginalizing X_C, we can derive the effect of set S on the dependent variable. Using these procedures, the feature PDPs are derived, as shown in Figures 8 and 9. The marginal impact of latitude is nearly constant until a threshold value of about 72.515, beyond which the crime rate decreases sharply. This pattern is similar to the line plot of Figure 5. Marginal effect of living costs is represented as a downward sloping line in most ranges, which coincides with the reversed U-shape of the line plot of Figure 6. Several other PDPs are given in the appendix.

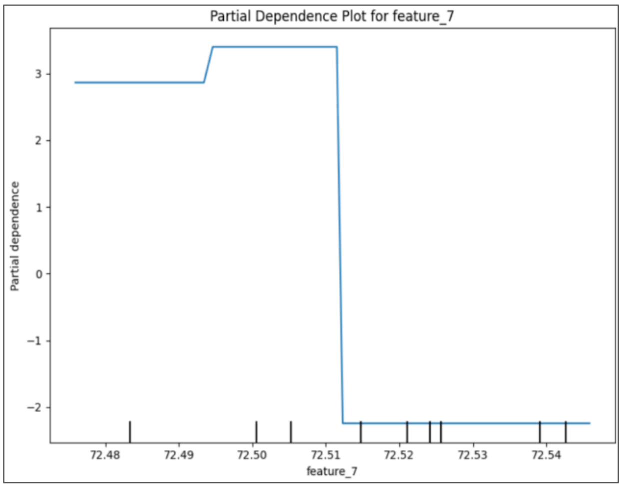

Figure 8. PDP of latitudes.

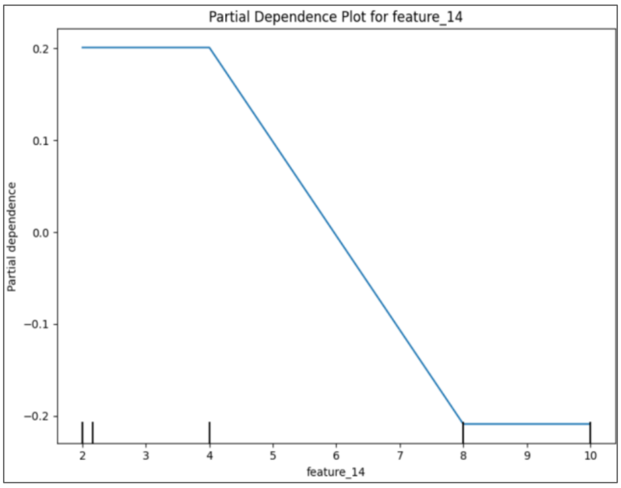

Figure 9. PDP of living costs.

7. Conclusion

According to the study's statistical results, crime rates are primarily determined by housing location. In Avon, Connecticut, USA, crime rates can be predicted using a Gradient Boosting Regression model applied to housing advertisement datasets, along with other features such as changes in income and living costs, with an R-squared of 0.99. When living costs are the explanatory variable, the line plot of the crime rate exhibits a reverse U-shape, indicating that the crime rate increases, peaks, and then decreases as living costs rise. This characteristic is captured by the linearly decreasing marginal contribution of living costs in the Partial Dependence Plot (PDP) for gradient-boost regression. Moreover, although the numerical data on crime rates are originally classified by longitude, such as Cherry Park, West Avon, Town Center, and Avon East, latitude is a key predictor of crime rates both in Multiple Linear Regression and Gradient Boost Regression. Although feature importance is approximately zero in Gradient Boost Regression and insignificant in Multiple Linear Regression, housing styles and construction materials are associated with crime rates in line plots. The interpretability is limited in Gradient descent; however, the linkage among multiple linear regression, PCA Biplot, feature importance, and PDP in Gradient Boost Regression can compensate for this limitation.

Declarations

Acknowledgments: The authors are thankful to the entire staff of Seoul Innovations Research Institute for insightful ideas and invaluable proofreading.
Artificial Intelligence (AI) Use Statement: The authors declare that no artificial intelligence (AI) tools were used to generate, analyze, or interpret the research data or to write the manuscript.

Author Contributions: SC: Design of study, definition of intellectual content, review manuscript, data collection, prepared first draft of manuscript, data analysis, statistical analysis; BN: Statistical interpretation, review manuscript, editing.
Conflict of Interest: The authors declare no conflict of interest.
Consent to Publish: The authors agree to publish the paper in International Journal of Recent Innovations in Academic Research.
Data Availability Statement: The datasets generated and/or analyzed during this study are not publicly available but are available from the corresponding author upon reasonable request.
Funding: This research received no external funding.
Institutional Review Board Statement: Not applicable.
Informed Consent Statement: Not applicable.
Research Content: The research content of this manuscript is original and has not been published elsewhere.

References

1. Utsha, R.B., Alif, M.N., Rayhan, Y., Hashem, T. and Ali, M.E. 2024. Deep learning based crime prediction models: Experiments and analysis. arXiv. https://doi.org/10.48550/arXiv.2407.19324
2. Bindu, K.P., Yaseen, M., Nagendra, B.N., Nitesh, A. and Prajwal, R. 2024. Crime rate prediction using machine learning. Journal of Emerging Technologies and Innovative Research, 11(5): e830–e834.
3. Alsubayhin, A., Ramzan, M. and Alzahrani, B. 2023. Crime prediction using machine learning: A comparative analysis. Journal of Computer Science, 19(9): 1170-1179.
4. Albors Zumel, A., Tizzoni, M. and Campedelli, G.M. 2025. Deep learning for crime forecasting: The role of mobility at fine-grained spatiotemporal scales. Journal of Quantitative Criminology. https://doi.org/10.1007/s10940-025-09629-3
5. Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L. and Chen, J. 2022. Interpretable machine learning models for crime prediction. Computers, Environment and Urban Systems, 94: 101789.
6. Farrell, G. and Pease, K. 1993. Once bitten, twice bitten: Repeat victimisation and its implications for crime prevention (Crime Prevention Unit Series Paper No. 46). Home Office Police Department. https://www.researchgate.net/publication/318710168
7. Sherman, L.W., Gartin, P.R. and Buerger, M.E. 1989. Hot spots of predatory crime: Routine activities and the criminology of place. Criminology, 27(1): 27-56.
8. Townsley, M., Homel, R. and Chaseling, J. 2003. Infectious burglaries. A test of the near repeat hypothesis. British Journal of Criminology, 43(3): 615-633.
9. Chainey, S. and Ratcliffe, J.H. 2005. Identifying crime hotspots. In: GIS and crime mapping (pp. 145–182). John Wiley and Sons. https://doi.org/10.1002/9781118685181.ch6
10. Kalinic, M. and Krisp, J.M. 2018, June. Kernel density estimation (KDE) vs. hot-spot analysis: Detecting criminal hot spots in the city of San Francisco. In: AGILE 2018 – 21st Conference on Geo-information Science (pp. 1–5), Lund, Sweden.
11. Rummens, A., Hardyns, W. and Pauwels, L. 2017. The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. Applied Geography, 86: 255-261.
12. Alves, L.G.A., Ribeiro, H.V. and Rodrigues, F.A. 2018. Crime prediction through urban metrics and statistical learning. Physica A: Statistical Mechanics and its Applications, 505, 435-443.
13. Zhang, W., Wang, Y., Yang, L. and Wang, C. 2020. Suspending classes without stopping learning: China's education emergency management policy in the COVID-19 outbreak. Journal of Risk and Financial Management, 13(3): 55.
14. Bernasco, W. and Block, R. 2011. Robberies in Chicago: A block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. Journal of Research in Crime and Delinquency, 48(1): 33-57.
15. Askey, A.P., Taylor, R., Groff, E.R. and Fingerhut, A. 2018. Fast food restaurants and convenience stores: Using sales volume to explain crime patterns in Seattle. Crime and Delinquency, 64(14): 1836-1857.
16. Carroll, J. and Weaver, F. 2017. Shoplifters' perceptions of crime opportunities: A process-tracing study. In: The reasoning criminal (pp. 19-38). Routledge.
17. Groff, E. and McCord, E.S. 2012. The role of neighborhood parks as crime generators. Security Journal, 25(1): 1-24.

18. Kurland, J., Johnson, S.D. and Tilley, N. 2014. Offenses around stadiums: A natural experiment on crime attraction and generation. Journal of Research in Crime and Delinquency, 51(1): 5-28.
19. Day, P., Breetzke, G., Kingham, S. and Campbell, M. 2012. Close proximity to alcohol outlets is associated with increased serious violent crime in New Zealand. Australian and New Zealand Journal of Public Health, 36(1): 48-54.
20. Weisburd, D., Lawton, B. and Ready, J. 2012. Staking out the next generation of studies of the criminology of place. In: The future of criminology. Oxford University Press.
21. Liu, L., Lan, M., Eck, J.E. and Kang, E.L. 2020. Assessing the effects of bus stop relocation on street robbery. Computers, Environment and Urban Systems, 80: 101455.
22. Brushwood, J., Dhaliwal, D., Fairhurst, D. and Serfling, M. 2016. Property crime, earnings variability, and the cost of capital. Journal of Corporate Finance, 40: 142-173.
23. Kinney, J.B., Brantingham, P.L., Wuschke, K., Kirk, M.G. and Brantingham, P.J. 2008. Crime attractors, generators and detractors: Land use and urban crime opportunities. Built Environment, 34(1): 62-74.
24. Song, G., Liu, L., Bernasco, W., Zhou, S., Xiao, L. and Long, D. 2018. Theft from the person in urban China: Assessing the diurnal effects of opportunity and social ecology. Habitat International, 78: 13-20.
25. Molnar, C., Casalicchio, G. and Bischl, B. 2020, September. Interpretable machine learning–a brief history, state-of-the-art and challenges. In: Joint European conference on machine learning and knowledge discovery in databases (pp. 417-431). Cham: Springer International Publishing.
26. Alvarez Melis, D. and Jaakkola, T. 2018. Towards robust interpretability with self-explaining neural networks. Advances in Neural Information Processing Systems, 31. https://doi.org/10.48550/arXiv.1806.07538
27. Guidotti, R. and Ruggieri, S. 2019, July. On the stability of interpretable models. In: 2019 international joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.
28. Baehrens, D., Schroeter, T., Harmeling, S., Kawanabe, M., Hansen, K. and Müller, K.R. 2010. How to explain individual classification decisions. The Journal of Machine Learning Research, 11: 1803-1831.
29. Ruder, S. 2017. An overview of gradient descent optimization algorithms. arXiv. https://doi.org/10.48550/arXiv.1609.04747

APPENDIX

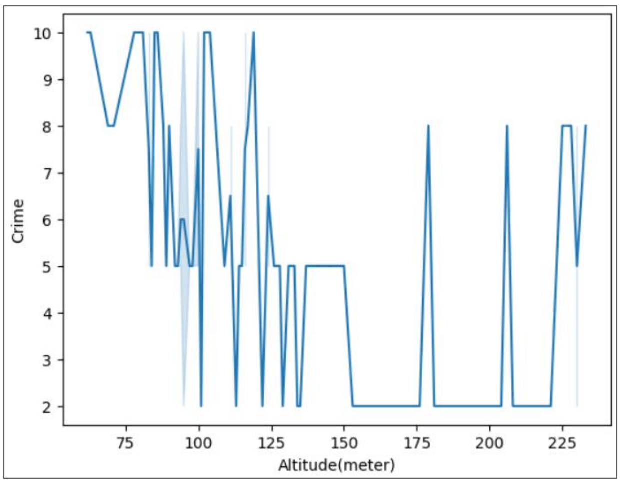

Figure 10. Line plot of crime vs altitude.

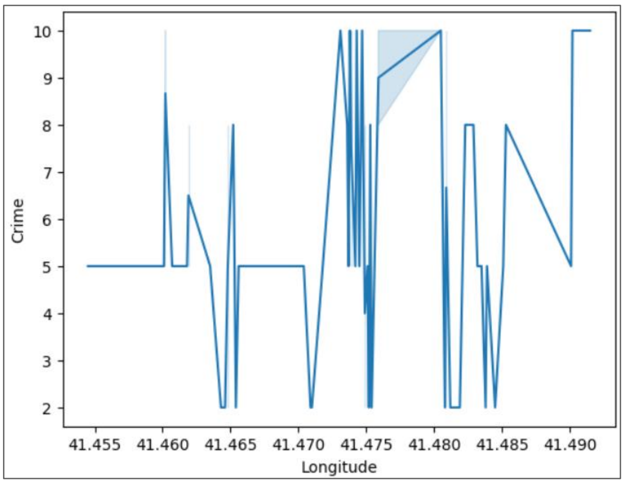

Figure 11. Line plot of crime vs longitude.

Citation: Seongbin Cho and Bob Nam. 2026. Crime Rate Prediction by Applying Gradient Boost Regression to Real Estate Data: The Case of Avon, Connecticut. International Journal of Recent Innovations in Academic Research, 10(2): 148-160.

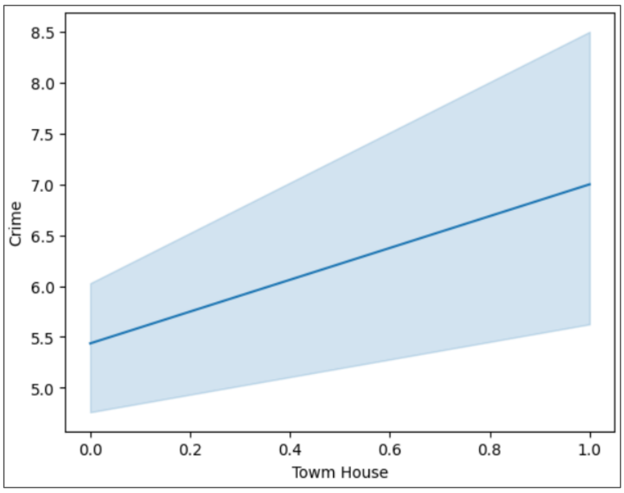

Figure 12. Line plot of crime vs town house.

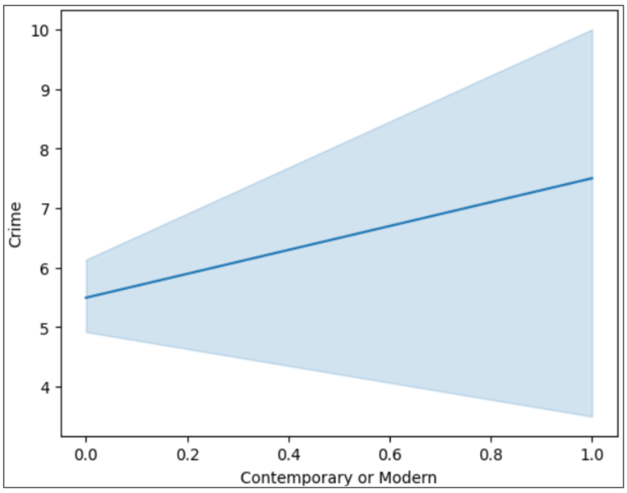

Figure 13. Line plot of crime v contemporary.

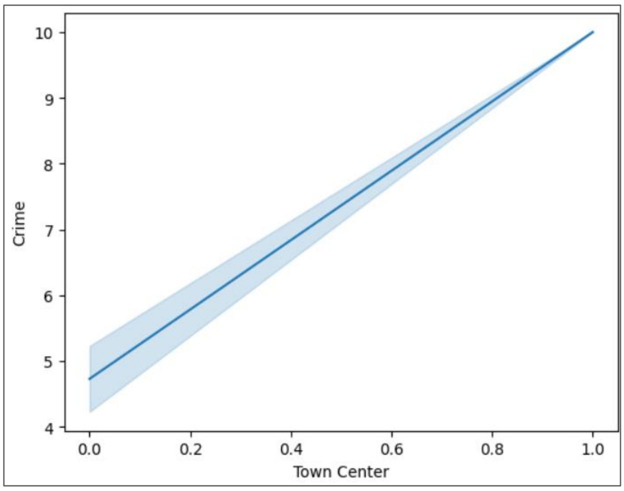

Figure 14. Line plot of crime vs town center.

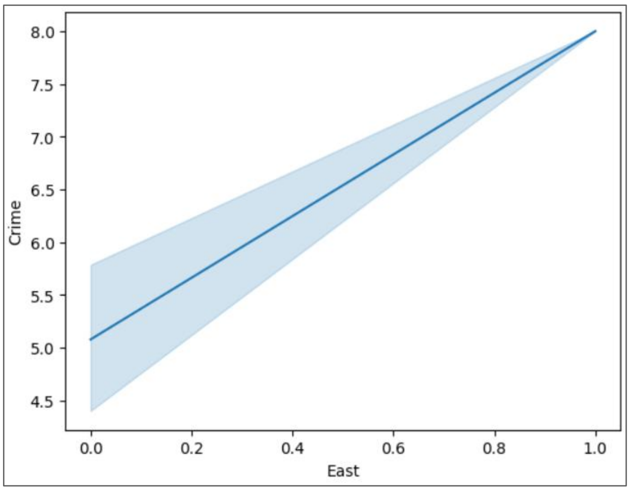

Figure 15. Line plot of crime vs east.

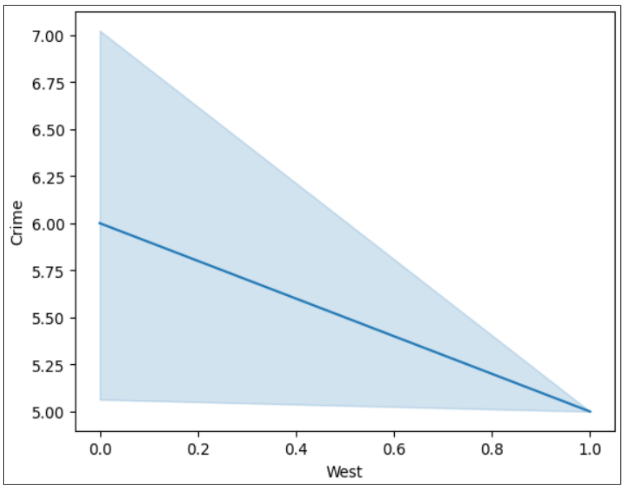

Figure 16. Line plot of crime vs west.

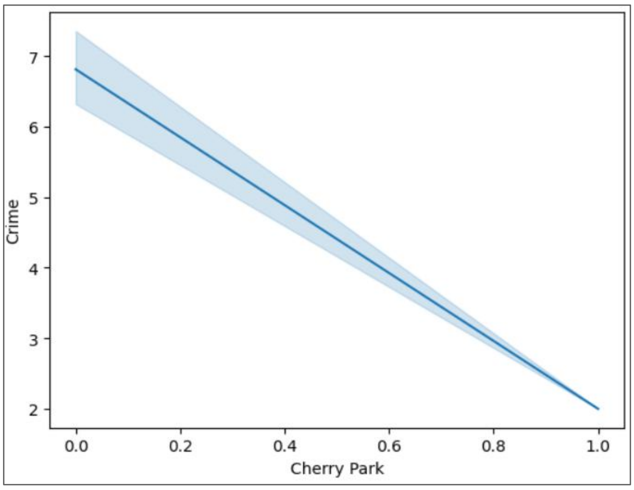

Figure 17. Line plot of crime vs Cherry Park.

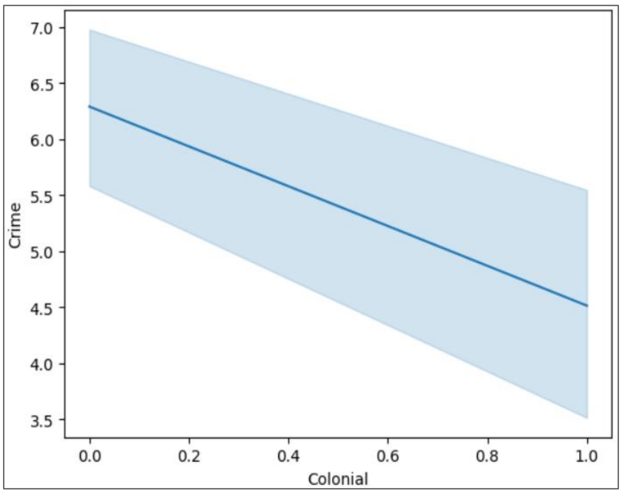

Figure 18. Line plot of crime vs colonial.

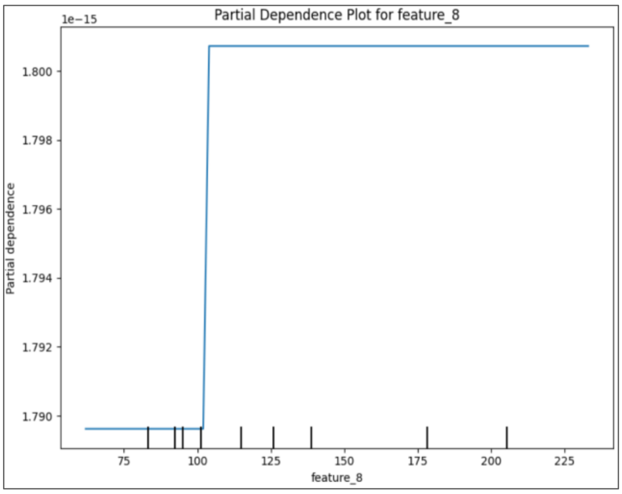

Figure 19. PDP of altitudes.

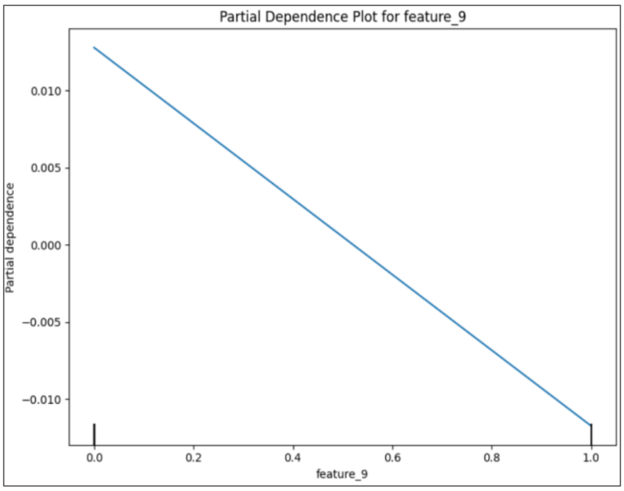

Figure 20. PDP of east dummy variable.

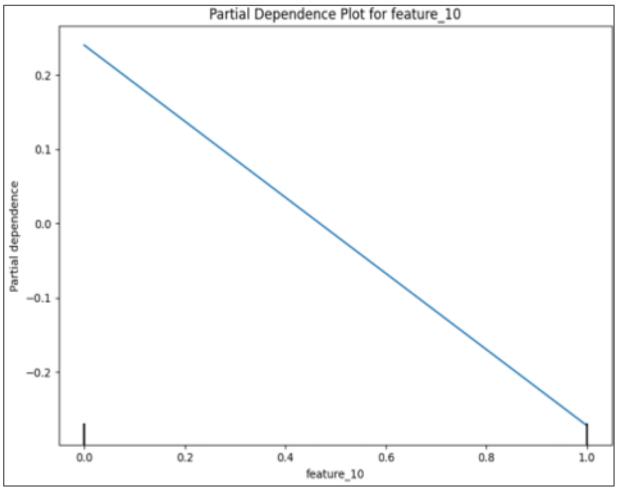

Figure 21. PDP of Cherry Park dummy variable.

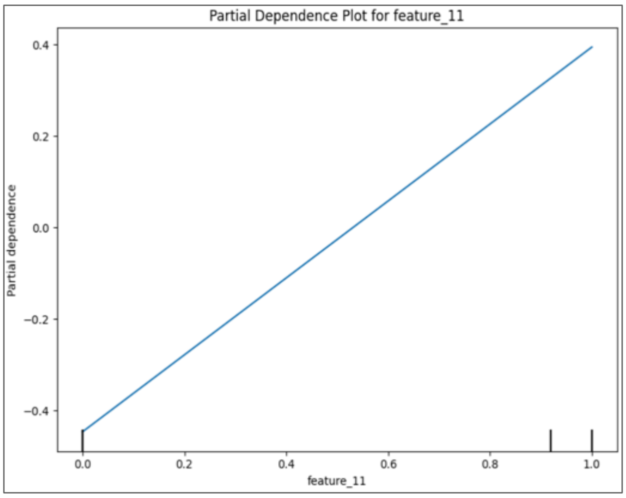

Figure 22. PDP of west dummy variable.

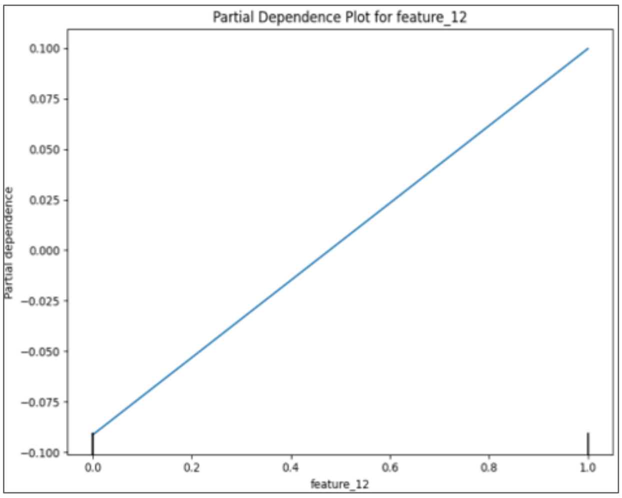

Figure 23. PDP of town center dummy variable.

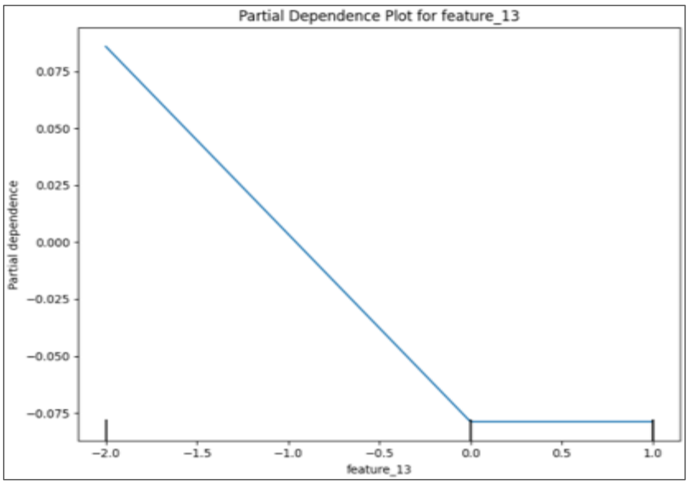

Figure 24. PDP of income change.