Crime Rate Prediction using Socioeconomic Indicators

Tristan Jorge Gapasin Gonzalez

1. Overview

A real-world application of economics has been extended to the field of criminology since a few decades ago. The economic analysis of crime suggests that even criminals are rational agents, so they weigh the expected costs and benefits before deciding whether to commit a crime. The expected costs in this context would be the severity of the punishment multiplied by the probability of getting such punishments. The studies conducted before the 1980s mainly focused on identifying an optimal punishment based on detection probability and severity. Nevertheless, the economists in the 1980s and 1990s, with the rising importance of empirical research and econometrics, started to shift their focus to a new topic—an analysis of economic determinants of crime. These determinants include but are not limited to national income inequality, unemployment, inflation, wages, race, and the level of urbanization (Kizilgol & Selim, 2017).

Crime, however, is an extremely complex aspect of human behavior. There are multiple variables interplaying with each other, either directly or indirectly influencing criminal behaviors. Aside from the microeconomic and macroeconomic factors, the sociocultural environment as well as the events and history specific to certain countries or communities are likely to influence criminal behaviors (Cerulli et al., 2018).

Some of these elements can be totally unexpected; for instance, Levitt & Dubner (2005) suggests that a single event, the legalization of abortion in 1973 in Roe v. Wade has largely reduced the crime rates in the following generation. Upon the legalization, a larger fraction of women, mainly from disadvantaged and unstable social environments, started to get abortions. In the early 1990s, the time when these never-born children would have been teenagers and entered their criminal prime, the US crime rate suddenly began to fall. As there was no explicit link between the abortion rate and crime rate prior to the late 1980s, which is when the post-Roe cohort was reaching its criminal prime, the paper argues that this fall in crime rates can be attributed to the legalization of abortion. The causal relationship is implied to some extent since the states with higher abortion rates had lower crime rates in general. This strong piece of evidence alludes that the examination of the relationship among various socioeconomic factors and crime rates should be done with a more holistic perspective and a multidisciplinary approach should be applied.

However, the common research approach of economists until recent days was the use of classic econometrics, measuring the effect of change in a single variable or a specific set of variables, whether it be a set of macroeconomic variables, microeconomic variables, or even a development economic variable. For example, Lochner (2004) explores the role of education, Fowles & Merva (1996) investigates the effect of changes in the distribution of wage income, and Cui & Hazra (2018) examines how a set of key macroeconomic variables such as real GDP per capita, unemployment and inflation determine the crime rates.

Unlike conventional and common research approaches, this paper examines the relative variable importance of a huge set of features on crime rates using machine learning techniques in order to gain a holistic view of the overall influences and interactions of variables. In other words, several different regression algorithms will be used to predict the crime rate based on a large variety of features, which deliver the prior information on the economic and socio-cultural environment of the communities.

This paper has two main research objectives. First, I want to investigate which socioeconomic features have the highest contribution to crime rate prediction. A corresponding research question focusing on an economic aspect would be: to what extent can the economic variables predict the crime rates of the communities? Secondly, the paper aims to find out which machine learning model has the highest prediction accuracy. Various regression algorithmic models including LASSO, elastic net, decision tree with ensemble methods, and artificial neural network will be evaluated for performance.

Investigating this topic endorses a significant social value. Crime rates are indirect measures of the rule of law in a country, which serve as an indicator of national development and determinant of economic growth and prosperity. There is evidence that higher crime rates incur social costs, deterring economic development (Mehlum et al., 2005). Some examples of the social costs would be the government budget spent on crime prevention, lower productivity among workers, a decline in FDI, loss of competitiveness and tourism, and many more (Kizilgol & Selim, 2017). The rule of law provides a sense of stability and uncertainty, fostering an ideal environment for capitalism (Blackwell, 2009). Among several components of the rule of law, the mitigation of violence is one of the major theoretical routes connecting the rule of law to economic growth (Haggard & Tiede, 2011). The measures of violence are also known to influence the volatility of growth. This may explain why some of the developing countries that did not manage to control social violence are exhibiting low levels of growth but relatively high volatility.

The following analysis will serve as preliminary work for further research, providing insights into the main factors potentially contributing to violent crime rates. Additional work will be needed before proposing any policy responses in order to establish a clear causal relationship. The result will contribute to the literature by either reinforcing or limiting the findings of the previous researchers. It is extremely important to cultivate a crime-deterring socioeconomic environment, and identifying the key predictors and determinants of crime rates will be the first step to achieving this. Also, specifying a high-accuracy crime prediction model will help law enforcement and the efficient allocation of resources in the area of crime prevention or patrol.

In fact, applying data science and machine learning techniques to crime prediction is getting increasing attention in recent years. A large volume of data-driven empirical studies is being published, examining several factors influencing criminal behavior. Some papers such as Liao et al. (2010) and Lin et al. (2017) even take spatial-temporal features into consideration, aiming to predict crime rates and incidence within specific regions, cities, and countries. A new algorithm developed by the University of Chicago, for instance, has shown a surprising performance, making a 1-week ahead prediction of future crimes with approximately 90% accuracy (Wood, 2022). I hope my analysis builds a foundation for this applied field and gives motivated scholars insightful guidance.

2. Data Review

2.1 Data Source

The data source is the Communities and Crime Data Set from the UCI Machine Learning Repository, and data is collected from "communities within the United States. The data combines socio-economic data from the 1990 US Census, law enforcement data from the 1990 US Law Enforcement Management and Administrative Statistics (LEMAS) survey, and crime data from the 1995 FBI Uniform Crime Reporting (UCR) program." (Redmond, 2009) In this dataset, all predictive numeric data was normalized into the decimal range of 0 to 1 using an "unsupervised, equal-interval binning method", where features "retain their distribution and skew". (Redmond, 2009)

2.2 Data Pre-processing

For this data set, there are 1994 observations and 128 features. Of these 128 features, 5 are not predictive, namely the state, county, community code, community name, and the fold number for non-random 10-fold CV. This means that, excluding the goal feature which is the violent crimes per population, there are 122 predictive features. There are 22 features with 1675 missing rows out of the total 1994 rows (84% missing). As such, these missing rows have to be imputed with missing data.

To transform the data, the five non-predictive features are first deleted using Microsoft Excel. Afterwards, it can be observed that some portions of the dataset have a lot of unknown values "?", as can be seen in the following figure:

?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
0.06	0.3	0.93	0.38	0.05
?	?	?	?	?
?	?	?	?	?
?	?	?	?	?
0.03	0.08	0.96	0.12	0.05

Figure 1. Depiction of missing values in the dataset

To manage the missing values, the question mark symbols are first replaced with NA in RStudio. Before moving on to impute the missing values, all features in the dataset must have the data class of type "numeric". This is checked in RStudio, as seen in the following sample figure:

```
     PctPolicWhite.numeric        PctPolicBlack.numeric         PctPolicHisp.numeric
               "character"                  "character"                  "character"
     PctPolicAsian.numeric        PctPolicMinor.numeric   OfficAssgnDrugUnits.numeric
               "character"                  "character"                  "character"
  NumKindsDrugsSeiz.numeric     PolicAveOTWorked.numeric             LandArea.numeric
               "character"                  "character"                    "numeric"
           PopDens.numeric         PctUsePubTrans.numeric            PolicCars.numeric
                 "numeric"                    "numeric"                  "character"
      PolicOperBudg.numeric   LemasPctPolicOnPatr.numeric   LemasGangUnitDeploy.numeric
               "character"                  "character"                  "character"
LemasPctOfficDrugUn.numeric        PolicBudgPerPop.numeric   ViolentCrimesPerPop.numeric
                 "numeric"                  "character"                    "numeric"
```

Figure 2. Depiction of predictor variables with the wrong class "character"

All features with the "character" class are then changed to "numeric" class:

```
     PctPolicWhite.numeric        PctPolicBlack.numeric         PctPolicHisp.numeric
                 "numeric"                    "numeric"                    "numeric"
     PctPolicAsian.numeric        PctPolicMinor.numeric   OfficAssgnDrugUnits.numeric
                 "numeric"                    "numeric"                    "numeric"
  NumKindsDrugsSeiz.numeric     PolicAveOTWorked.numeric             LandArea.numeric
                 "numeric"                    "numeric"                    "numeric"
           PopDens.numeric         PctUsePubTrans.numeric            PolicCars.numeric
                 "numeric"                    "numeric"                    "numeric"
      PolicOperBudg.numeric   LemasPctPolicOnPatr.numeric   LemasGangUnitDeploy.numeric
                 "numeric"                    "numeric"                    "numeric"
LemasPctOfficDrugUn.numeric        PolicBudgPerPop.numeric   ViolentCrimesPerPop.numeric
                 "numeric"                    "numeric"                    "numeric"
```

Figure 3. Depiction of corrected classes

Since all features have the same numeric class, I can proceed with imputing the NA values. Imputing was accomplished with the rfImpute() function in the randomForest library. The function imputes missing values in the predictive features by means of the proximity matrix from the library. In this case, the imputed values of my predictive variables are the weighted average of non-missing rows, where the proximities are the weights. Now, the dataset can be used to train and test the models I will use in this paper.

2.3 Construction of the Response Variable

My response variable is the total number of violent crimes per 100k people. In other words, it is the number of violent crimes per capita. This was calculated using the sum of crime variables considered as violent crimes in the United States, namely murder, rape, robbery, and assault. The sum is then divided by the population, thereby yielding the final goal variable of violent crimes per capita. For the sake of simplicity, I renamed it as the acronym VCPP.

VCPP is the target variable because my interest in this paper is to find out the characteristics that have the greatest impact on the rate of violent crimes committed.

2.4 Elaboration of Predictive Features

There are 122 predictive features in the dataset, which are both economic and socioeconomic in nature. Economic features include income and employment, while socioeconomic characteristics include race, marital status, illegitimate birth, language, homelessness, and population density, to name a few. Similar to the goal variable VCPP, the predictive features are also normalized to an interval between 0 and 1. Normalizing the variables is important because I do not want a single variable to drive model performance in a particular direction simply because variable ranges are different. The loss of interpretive ability is not particularly significant within the scope of this paper, because my interest is to find the most important predictive features that potentially influence VCPP, as well as to compare model performance relative to each other.

3. Baseline Results

3.1 Methods and Packages Used

The baseline models that I will evaluate are the 10-fold CV LASSO and elastic net.

Package	**Description**
glmnet()	To create LASSO and elastic net models for evaluation
vip()	To create variable importance plots for the models in order to evaluate the most important factors that contribute to the violent crime rate

Figure 4. Table of packages used for section 3 and their descriptions

3.1.1 LASSO

The LASSO model shrinks the coefficients as shown in Figure 5.

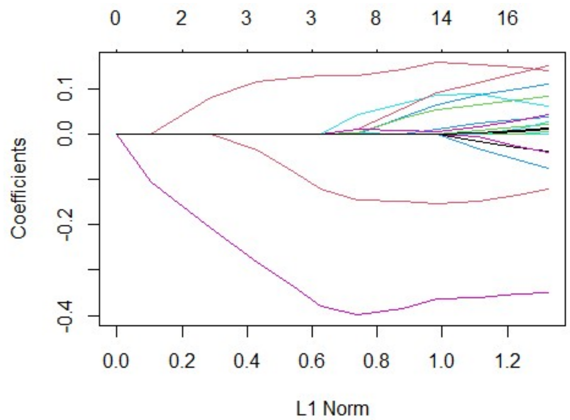
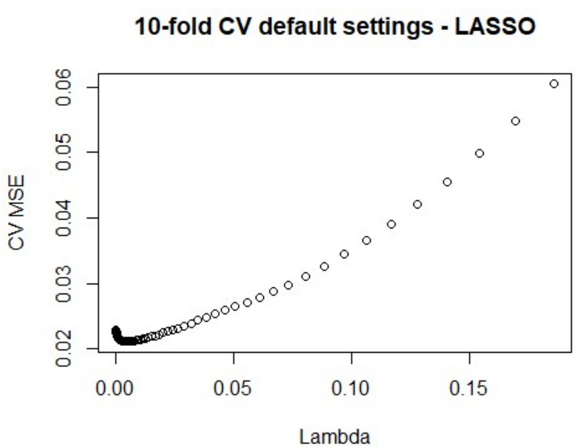


Figure 5. Shrinkage of coefficients by L1 norm



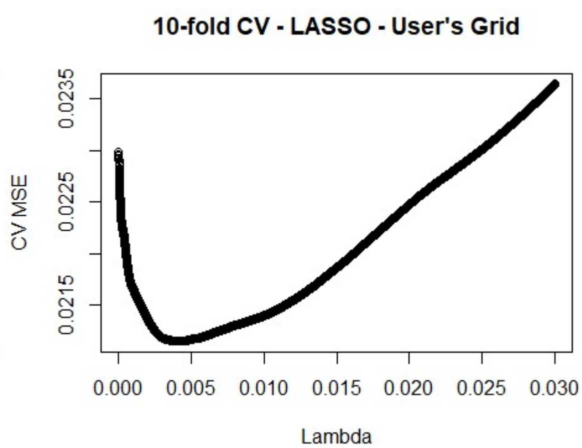
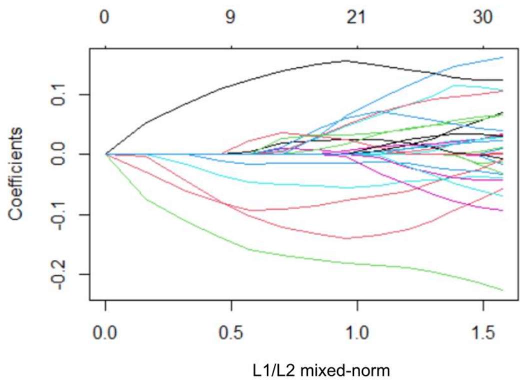


Figure 6. 10-fold CV default settings - LASSO (left); Figure 7. 10-fold CV user's grid - LASSO (right)

Using the default grid, I find that the lowest lambda that minimizes CV MSE is 0.003493706, as shown in Figure 6. The corresponding CV MSE is 0.01761090. Using the user's grid, I find that the lowest lambda that minimizes CV MSE is 0.003492349, as shown in Figure 7. The corresponding CV MSE is 0.01761090, which is the same as the result under the default grid.

3.1.2 Elastic Net (alpha=0.5)



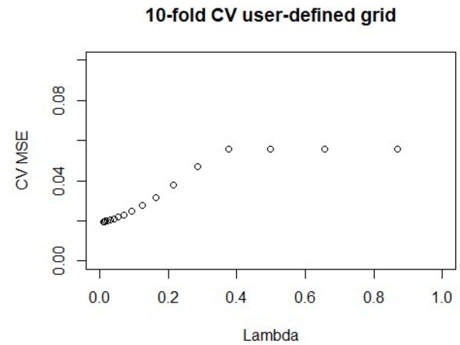


Figure 8. Shrinkage of coefficients by L1/L2 mixed-norm (left); Figure 9. 10-fold CV user's grid (right)

Using alpha = 0.5, the Elastic Net model shrinks the coefficients as shown in Figure 8. Using 10-fold CV, I find the lowest lambda that minimizes CV MSE is 0.1, as shown in Figure 9. The corresponding CV MSE is 0.01948314. When I apply the Elastic Net (alpha=0.5) to the test sample, I get an MSE of 0.01719054.

3.2 Results and Discussion

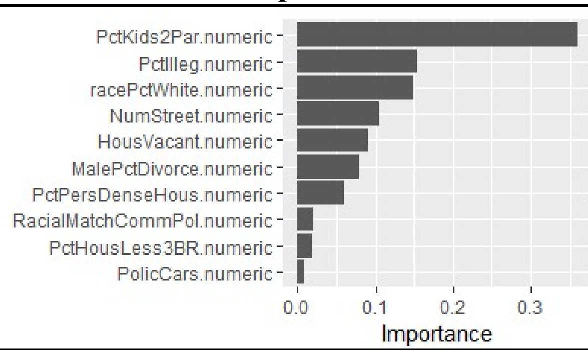
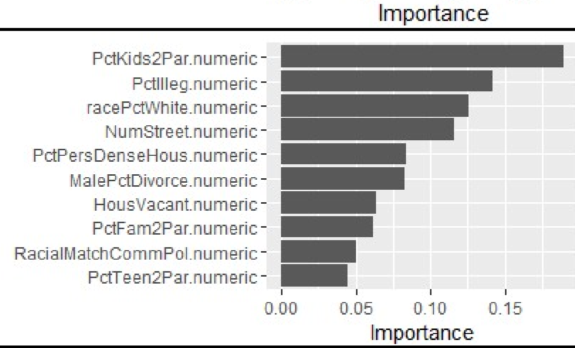
Model	Test MSE	Variable Importance Plot
LASSO 10-fold CV	0.01761090	
Elastic Net (alpha=0.5)	0.01719054	

Figure 10. Table of test MSEs and VIPs for LASSO and Elastic Net (alpha=0.5)

From Figure 10, I can see that the Elastic Net (alpha=0.5) model is the better model since it has the lowest out-of-sample MSE. This is in line with the theory of elastic net being the better model when compared to LASSO.

In terms of the importance of the features, both models selected PctKids2Par as the most important feature to forecast VCPP, being slightly more than 35% for LASSO 10-fold CV, and about 18% for Elastic Net (alpha=0.5). Both models also select PctIlleg, racePctWhite and NumStreet as the 2nd, 3rd, and 4th most important features respectively.

4. Ensemble Learning Results

4.1 Methods and Packages Used

The ensemble learning methods that I will evaluate are Bagging and Random Forest, Gradient Boosting, and Artificial Neural Network.

Package	**Description**
randomForest()	To create random forest models for evaluation
gbm()	To create gradient boosting models for evaluation
nnet()	To create artificial neural networks for evaluation
vip()	To create variable importance plots for the models in order to evaluate the most important factors that contribute to the violent crime rate
caret()	To streamline model training process for complex regression and classification problems
dplyr()	For data manipulation on my datasets
ggplot2()	For data visualization

Figure 11. Table of packages used in section 4 and their descriptions

4.1.1 Bagging and Random Forest (RF)

For my paper, two types of ensemble algorithms will be evaluated, namely Bagging and RF with the default choice of mtry (40). Mtry refers to the number of predictors randomly selected from all 122 predictors at each split. Because the problem in this paper is a regression problem, the default mtry is 122/3, or approximately 40. This means that 40 variables will be randomly selected from the 122 predictive features at each split as the individual trees grow.

For both methods, the error reaches its asymptotic minimum when the number of trees approaches 500. Hence, for the sake of consistency, the number of trees that will be used for each model is 500. The error plots for the respective random forests used are as follows:

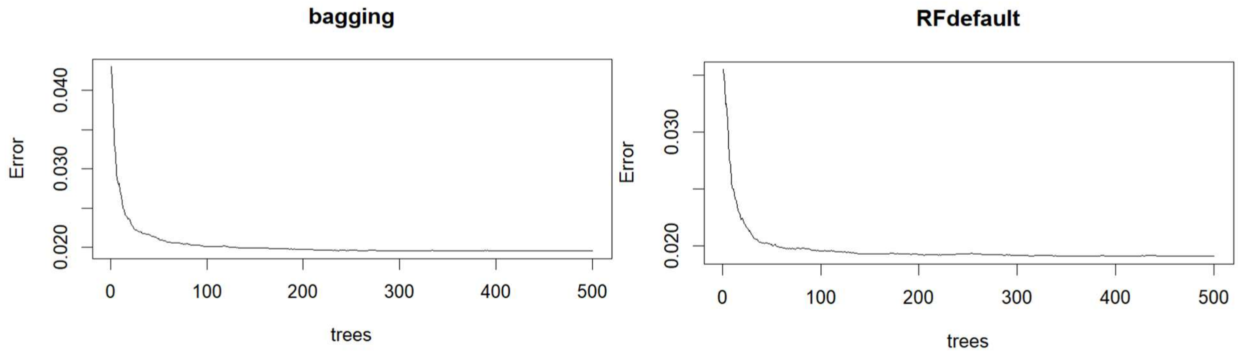


Figure 12.1 Error plot for Bagging (left); 12.2 Error plot for RF with default mtry=40 (right)

After training both models with the training dataset and testing with the test set, the following results are obtained:

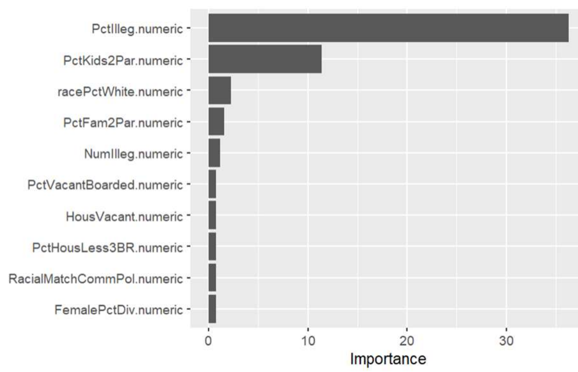
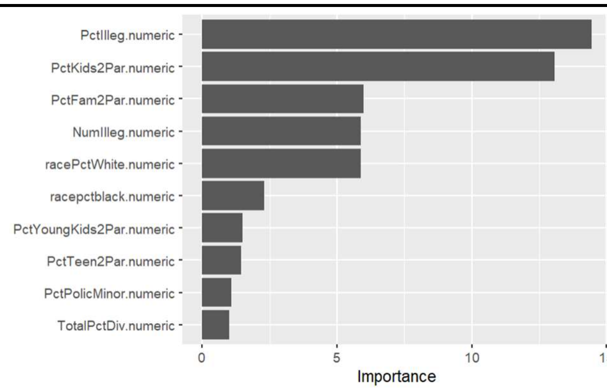
Random Forest	Test MSE	Variable Importance Plot (VIP)
mtry = 122 (bagging)	0.01750951	
mtry = 40 (default)	0.01693416	

Figure 13. Table of test MSEs and VIPs for their corresponding values of mtry

Random Forest mtry	Test MSE	Test MSE Relative Rank
40 (default)	0.01693416	1
122 (bagging)	0.01750951	2

Figure 14: Table of ranked MSEs

From the results above, bagging performs worse in terms of MSE for the test dataset. These results are not surprising, given that the trees in RFs with a lower choice of mtry are more decorrelated than the trees in bagging, which leads to a more optimal bias-variance tradeoff, thereby leading to a lower test MSE.

It can be noted from the VIPs above that the most dominant variable in influencing the violent crime rate is the percentage of illegitimate children (Pctilleg) in that area. This finding corroborates the findings in the literature review, which show a strong link between crime rates and out-of-wedlock births. Notably, as mtry decreases, the previously unimportant variables become more important relative to the most dominant variable. This is because trees become more de-correlated the more mtry decreases, which leads to a "more thorough exploration of model space" (James et al., 2021), thereby increasing the relative importance of previously less-dominant predictive features.

For the bagged RF, virtually all of the trees will use the dominant variable Pctilleg in the top split, which leads to trees in the bagged RF being rather similar to each other, because many (if not all) of the splits in each tree are influenced by Pctilleg. In contrast, due to the decorrelation of the trees in the default RF, Pctilleg loses its influence and other variables have more opportunity to take part in splits down each tree. This effect of decorrelation can be prominently observed in the VIP for the tuned random forest, where Pctilleg appears to be less distinguished from the other variables.

<u>4.1.2 Gradient Boosting</u>

I first ran my first Gradient Boosted model (gbm.fit) with generic hyperparameters (particularly, n.trees=10000, interaction.depth=1, λ=0.001) and 10-fold CV, and found that optimal n.trees = 10000 with a 10-fold CV MSE of 0.01926316. The graph below shows that the CV error is still decreasing even at 10,000 trees, which shows that the very small learning rate resulting in very small incremental improvements requires n.trees > 10,000 to minimize the CV error.

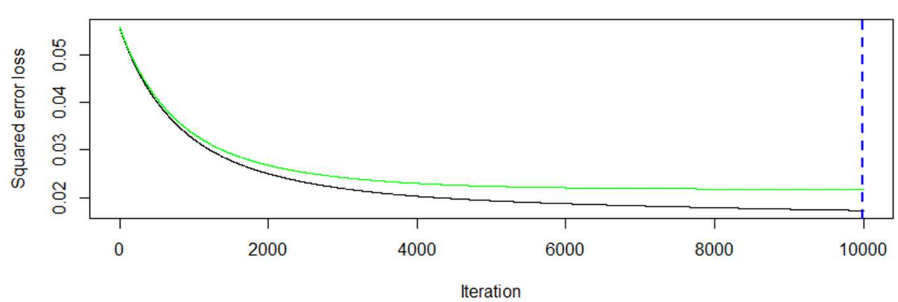


Figure 15. Plot of Loss Function for gbm.fit

After tuning the model by reducing the number of trees, increasing the learning rate to take larger steps down the gradient descent, and increasing the depth of each tree from a single split to 3 splits (n.trees=5000, interaction.depth=3, λ=0.01) for my second gradient boosted model (gbm.fit2), I arrived at optimal n.trees = 1472 with a 10-fold CV MSE of 0.01893943.

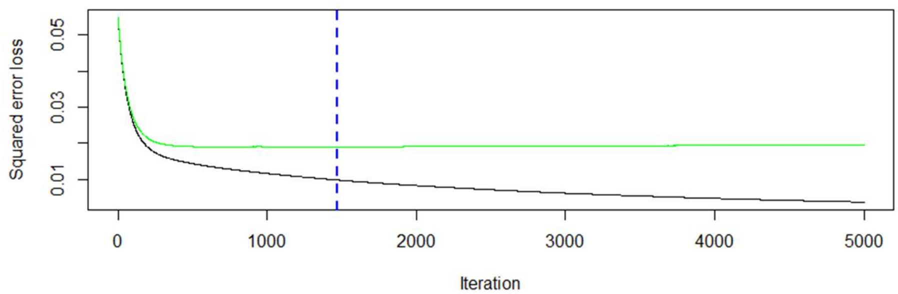


Figure 16. Plot of Loss Function for gbm.fit2

I then created a hyperparameter grid (hyper_grid) to assess different combinations of hyperparameters, as shown below, and ran a for-loop to train my tuning model (gbm.tune) on a total of 81 combinations of hyperparameters. This hyper_grid was continually adjusted to allow us to look into regions of values that produced the best results in previous grid searches. The final adjusted hyper_grid as well as the top 10 models generated from gbm.tune are shown in the table below.

```
hyper_grid <- expand.grid(
    shrinkage = c(.01, .05, .10),             # determine optimal learning rate
    interaction.depth = c(3, 5, 7),           # highest level of variable interactions
    n.minobsinnode = c(5, 10, 15),            # vary min. no. of obs. in tree terminal nodes
    bag.fraction = c(.65, .8, 1),             # introduce stochastic gradient descent
    optimal_trees = 0,                        # a place to dump results
    min_MSE = 0                               # a place to dump results
    )
```

#	λ	interaction.depth (d)	n.minobsinnode	bag.fraction	optimal_trees	min_MSE
1	0.01	7	5	0.65	668	0.01707578
2	0.01	7	10	0.65	670	0.01719975
3	0.01	7	15	0.65	1237	0.01728900
4	0.10	3	15	0.65	222	0.01731691
5	0.10	3	10	0.65	179	0.01736683
6	0.01	5	5	0.65	775	0.01739482
7	0.01	7	10	0.80	1341	0.01742191
8	0.01	5	15	0.65	1237	0.01747583
9	0.01	5	10	0.65	1083	0.01751727
10	0.01	3	10	0.65	2002	0.01753708

Figure 17. Top 10 gbm.tune models ranked by min_MSE

I learned a few things from the top models of gbm.tune: (1) Majority of the top models used shrinkage = 0.01, suggesting that smaller incremental steps down the gradient descent is still preferred over shrinkage = 0.05, 0.10, with a few exceptions; (2) Few of the top models used interaction.depth=1, indicating there are likely some important interactions that the deeper trees are able to capture; (3) Adding a stochastic component with bag.fraction < 1 introduces randomness into my model fit, which seems to help as there may be some local minimas in my loss function gradient; (4) The optimal n.trees shows a bias-variance tradeoff, where increasing N reduces the error but could lead to overfitting.

Using the hyperparameters from my top gbm.tune model, I created gbm.fit.final and predicted my results using my test set, which gave us a test MSE of 0.01690916. The variables of importance can be seen in the Figure below. PctIlleg is my standout variable accounting for more than 28% of the result, with PctKids2Par at ~18%, racePctWhite and NumIlleg at ~9%, and the remaining 7 variables around 2-5% importance.

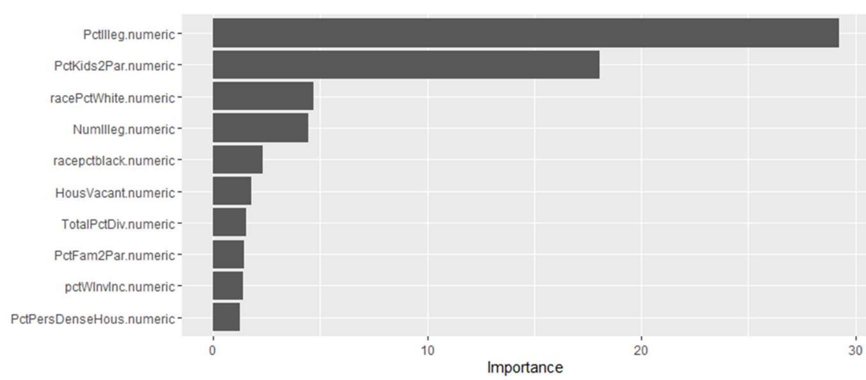


Figure 18. Variable importance plot of gradient boosted model

4.1.3 Artificial Neural Network

I first normalize the dataset with the Min-Max Scaling method. This suppresses the effect of outliers on the data values to a certain extent, and helps us to have a smaller value of the standard deviation of the data scale. Then, I run the neural network through a list of decay constants to determine the optimal decay constant, selecting the best network with the best MSE result. For each neural network model, I used 10 hidden units (size=10), 1,000 maximum number of iterations (maxit=1000), swapped to linear output (linout=TRUE), with 10,000 maximum number of weights (MaxNWts=10000). My best MSE was reached when $\lambda = 3$, with MSE = 0.01637489. The results for each decay constant can be seen in Figure 19.

Decay Constant, λ	Test MSE
0	64.82866183
0.001	0.08789873
0.003	0.08717548
0.01	0.04689725
0.03	0.02779873
0.1	0.02017240
0.3	**0.01639020**
1	0.01660042
3	0.01734771
10	0.01902379

Figure 19. Decay constants (λ) and resulting MSEs in neural networks

The Variable Importance Plot below shows the 10 most important variables as identified by the Artificial Neural Network model. Here, racepctblack is the most important variable accounting for over 20% of the result, with PctIlleg at around 18% and the remaining 8 variables at only around 9-10% each.

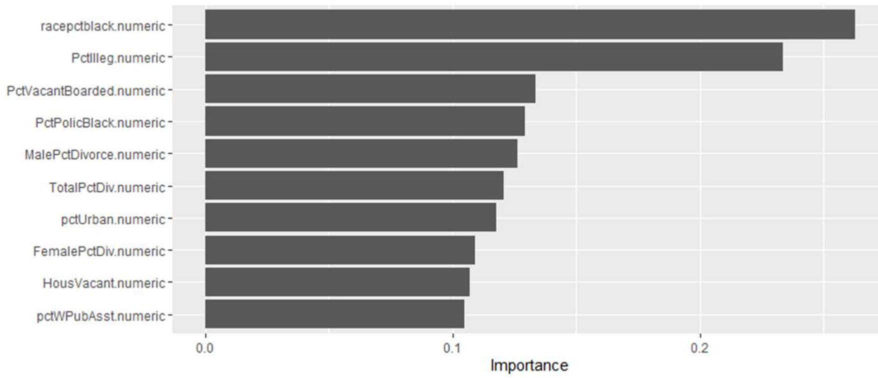


Figure 20. Variable importance plot for the artificial neural network model

5. Overall Results and Discussion

Model	Test MSE	Ranking (best to worst)
Artificial Neural Network (maxit=1000, λ=0.3)	0.01639020	1
Gradient Boosting (d=7, λ=0.01)	0.01690916	2
Default RF (ntrees=500, mtry=40)	0.01693416	3
Elastic Net (alpha=0.5)	0.01719054	4
Bagging (ntrees=500, mtry=122)	0.01750951	5
LASSO 10-fold CV	0.01761090	6

Figure 21. Test MSEs by ensemble learning method

From the table above, the ANN performs the best in terms of Test MSE, out of all the six models considered in this paper. This could be explained by the ANN's ability to evaluate nonlinearities in the dataset as well as capture the various interaction effects. Gradient boosting also performs better than the default random forest and bagging, potentially due to the slow-learning characteristic of boosting, which involves fitting small trees using current residuals. In this case, the results imply that gradient boosting did not overfit, thereby leading to a better Test MSE performance than default RF and bagging. Unsurprisingly, LASSO performed the worst out of all six models, due to its low flexibility in selecting variables of a dataset that demands nuance and careful treatment of its interactive variables.

6. Conclusion

Based on the VIPs of the models evaluated in this paper, I conclude that economic variables such as wage and employment play a virtually insignificant role in predicting community crime rates. This implies that taking a more holistic approach, by incorporating sociocultural features into a model with purely economic variables, can significantly improve predictive performance. The major predictors I identified are PctKids2Par and PctIlleg. Respectively, these variables refer to the percentage of kids in family housing with two parents, as well as the percentage of kids born to parents that were never married, surprisingly supporting what Levitt & Dubner (2005) has suggested. Other noticeable predictors include race and the number of vacant households. Further studies on the causal relationship between these variables and violent crime rates may guide policy decisions to help prevent crime and enhance law enforcement.

By comparing the predictive accuracy of my forecasting models according to their test MSE, I concluded that my best model by far in terms of predicting my target variable (violent crimes per population) is the Artificial Neural Network, with the lowest test MSE of 0.0163902. The nonlinearity of ANN's sigmoid activation function allows the model to capture the complex nonlinearities and interaction effects (James et al., 2021) within my 122 predictors.

While I acknowledge that the majority of the predictors exhibit multicollinearity, this is not a problem for neural networks. Due to their tendency to be overparameterized, neural networks tend to be fairly insensitive to problems of multicollinearity (De Veaux & Ungar, 1994). The extra learned weights generated in neural networks create redundancies that make things that affect any subset of features unimportant. It is worth noting that if I were to use multi-layer networks and deep learning (unlike single-layer in this case), I could obtain far better test predictive accuracy through the higher-order interactions of original predictors, albeit at the expense of interpretability.

Overall, the results that I have obtained reinforce the complex relationship between criminal behavior and its component predictive features. Additionally, my results also demonstrate why I need to account beyond the scope of just economic variables.

7. References

Blackwell, J. K. (2009). *Rule of Law: Economic Prosperity Requires the Rule of Law*. Heritage Foundation. Retrieved from http://www.heritage.org/research/projects/indivisible

Cerulli, G., Ventura, M., & Baum, C. F. (2018). The Economic Determinants of Crime: an Approach through Responsiveness Scores. *Boston College Working Papers in Economics*. Retrieved from http://fmwww.bc.edu/EC-P/wp948.pdf

Cui, Z., & Hazra, D. (2018). Macroeconomic determinants of crime: Evidence from India. *Journal of Quantitative Economics*, *16*(1), 187–198. https://doi.org/10.2139/ssrn.3005019

De Veaux, R., & Ungar, L. (1994). Multicollinearity: A tale of two nonparametric regressions. https://www.cis.upenn.edu/~ungar/Datamining/Publications/tale2.pdf

Fowles, R. & Merva, M. (1996), Wage inequality and criminal activity: An extreme bounds analysis for the united states, 1975–1990, Criminology 34(2), 163–182.

Haggard, S., & Tiede, L. (2011). The rule of law and economic growth: Where are I? *World Development*, *39*(5), 673–685. https://doi.org/10.1016/j.worlddev.2010.10.007

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning. In Springer Texts in Statistics. Springer US. https://doi.org/10.1007/978-1-0716-1418-1

Kizilgol, O., & Selim, S. (2017). Socio-economic and demographic determinants of crime by Panel Count Data Analysis: The case of EU 28 and Turkey. *Business, Economics and Finance JBEF* , *6*(1), 31–41. https://doi.org/10.17261/Pressacademia.2017.383

Levitt, S. D., & Dubner, S. J. (2005). Freakonomics: A rogue economist explores the hidden side of everything. William Morrow & Co.

Liao, R., Wang, X., Li, L., & Qin, Z. (2010). A novel serial crime prediction model based on Bayesian learning theory. 2010 International Conference on Machine Learning and Cybernetics. https://doi.org/10.1109/icmlc.2010.5580971

Lin, Y.-L., Chen, T.-Y., & Yu, L.-C. (2017). Using machine learning to assist crime prevention. *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*. https://doi.org/10.1109/iiai-aai.2017.46

Lochner, L. (2004), 'Education, work, and crime: A human capital approach', *International Economic Review* 45(3), 811–843.

Mehlum, H., Moene, K., & Torvik, R. (2005). Crime induced poverty traps. *Journal of Development Economics*, *77*(2), 325–340. https://doi.org/10.1016/j.jdeveco.2004.05.002

Redmond, M. A. & Baveja, A. (2002). A Data-Driven Software Tool for Enabling Cooperative Information Sharing Among Police Departments. European Journal of Operational Research 141, 660-678.

Redmond, M. (2009, July). *Communities and Crime Data Set*. UCI Machine Learning Repository: Communities and Crime Data Set. Retrieved October 2022, from https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime

U. S. Department of Commerce, Bureau of the Census. (1990). Census Of Population And Housing 1990 United States: Summary Tape File 1a & 3a. (Computer Files)

U.S. Department Of Commerce, Bureau Of The Census. (1992). Producer, Washington, DC and Inter-university Consortium for Political and Social Research Ann Arbor, Michigan. (Computer File)

U.S. Department of Justice, Bureau of Justice Statistics. (2020). Law Enforcement Management And Administrative Statistics (Computer File)

U.S. Department of Justice, Federal Bureau of Investigation. (1995). Crime in the United States (Computer File)

Wood, M. (2022, June 30). *Algorithm predicts crime a week in advance, but reveals bias in police response*. Biological Sciences Division | The University of Chicago. Retrieved November 10, 2022, from https://biologicalsciences.uchicago.edu/news/features/algorithm-predicts-crime-police-bias