Crime Prediction with Various Methods for Limited Time-points

Seyda Betul Aydin [*1] **Csenge Frater** [*2] **Melis Oktayoglu** [*3]

Abstract

Recently, - especially with the advancements of machine learning methods - , crime prediction has acquired a significant research attention [1][2]. Here we are exploring various simpler methods on the dataset of the German Police Crime Statistics. The data covers years 1976 to 2020 with annual aggregation. We are trying to find the best method for the limited amount of data, also considering prediction with simple visualization techniques, Logistic Regression, and a more complex forecasting method (ARIMA), considering the potential pitfalls of working with such a limited amount of data on an extremely complex phenomenon. (We are particularly interested in the extent of violence (involved firearm) occurring during the police interventions across different forms of criminal behaviours, and prediction for the future.)

1. Contribution Plan

Csenge Frater is responsible for the literature research. All members equally contribute to initial exploration of the data and the final evaluation of the methods. Seyda Betul Aydin and Melis Oktayoglu is responsible for time series prediction.

2. Introduction

2.1. Influences on crime

Crime is a complex phenomenon driven by multiple factors. Common sense and much evidence point towards economic indicators as the main candidate. More precisely, for example low legitimate earning prospects, joblessness, and greater inequality seem to be related to higher rates.

But even those can be inconsistent in some cases. One counter example is the Great Recession, which with its financial collapse curbed crime rates in the United States and most developed countries. According to (A. 4.) this could be explained away by inflation as an intermediate variable. If inflation is present in addition to recession, the value of goods rise with an incentive on theft, which introduces a domino-like effect of pushing other kinds of crimes up. This highlights the fact that these factors are not invariant, can have a complex relationships hard to disentangle.

One such is social disorganisation with a high influx of refugees around 2015 (B. 7.) to Germany. (B. 8.) have found a moderate increase in crime rates after the arrival of over one million people.

2.2. Measures to prevent

Legal acts try to minimize crime with respect to these assumed influences. Long-term solutions includes extending labour market opportunities, especially for the less-skilled (A. 1.). One seemingly successful policy in Germany was the investment in education and social system, which provides social support for high-school dropouts and thus prevents the possible downward cycle of the less prospective youth. (B. 3)

Nonetheless, one would think that a "strong hand" helps with keeping crime rates low which was contradicted by the US criminal justice system. After imprisonment the convicted faces a social stigma which could lead to a vicious cycle where being a recidivist seems like an only option. (A .1)

2.3. Looking at the individual level (Rational choice theory)

(B. 1) Ehrlich, Block, Heineke, and Wolpin introduced a theoretical framework for predicting crime rates, which looks at the expected utility of a person allocating time on illegal and legal acts. Each individual is assumed to maximise expected utility, i.e.

[*]Equal contribution [1]6323572, seyda-betul.aydin@student.uni-tuebingen.de, MSc. Quantitative Data Science [2]6639491, csenge.frater@student.uni-tuebingen.de, MSc. Machine Learning [3]6620677, melis.oktayoglu@student.uni-tuebingen.de, MSc. Machine Learning.

Project report for the "Data Literacy" course at the University of Tübingen, Winter 2023/24 (Module ML4201). Style template based on the ICML style files 2023.

$$E(U) = p\{(1 - p_{s|c})U[A + L_b(t_i) + G(t_i)] + p_{s|c}U[A + G(t_i) - F(t_i)]\} + (1 - p)U[A + L(t_e) + G(t_e)], \quad (1)$$

(2)

$$p = p(detection) \cdot p(trial|detection) \cdot p(conviction|trial), \quad (3)$$

(4)

p = probability of detection and punishment of illegal activity
t_l = time allocated for legal work,
t_i = time spent on illegal activity,
A = initial wealth,
$L(t_e)$ = legal income,
$L_b(t_e)$ = legal income after previous conviction(adjusted for non-custodial sentences such as fines),
$G(t_i)$ = illegal gain,
$F(t_i)$ = cost function (monetary equivalent) of prison sentence,
p = probability of detection and punishment of illegal activity
$p_{s|c}$ = custodial sentence,
$1 - p_{s|c}$ = non-custodial sentence (such as probation, fine, social work).

According to the paper replacing custodial sentences (jail or prison time) with non-custodial sentences (like probation or fines) appears to be successful. This approach hasn't led to a significant increase in crime and has benefits like reducing future prison populations, stigmatization, and the accumulation of criminal behavior.

2.4. Problems with the data

One paper (B. 2.) specifically addresses the problems with the German Police Crime Statistics. The main problem with this dataset is that it contains only the officially reported cases. This problem could be solved partly by the so-called "dark-figure" studies with subjective narratives and large sample sizes, where the proportions could be matched to the officially recorded data. However, this method is fairly biased due to subjective experiences of victimization and relies on only the individual's memory. Also these studies have a minimum age requirement with not institutionalized and convicted people, which rules out the possibility of a representative measure.

Thus adding up the total number of cases with equal weights can be misleading as an indicator of true crime risk. For the latter problem the paper examines three types of weight calculations: monetization weighting, opinion-based weighting and data-based weighting. According to the results all methods show similar results. With balancing, the overall crime rate is affected more by more frequent forms of offenses.

2.5. Problems with forecasting crime

One should approach prediction with an heavily critical mindset if it involves such a complex phenomenon as crime prediction. Our method brings assumptions which do not necessarily hold, especially as time progresses. Unexpected events can occur such as the Ukrainian war, also political change can influence the laws and punitive system, causing feedback loops on the observable crime variables.

Even the forecast itself can have an indirect impact on the future correctness of the results by influencing law enforcement. For example predicting the drop of overall crime rates can decrease the budget spent on prevention, thus causing an opposite effect as expected. (B 3.)

3. Methods

References

[1] Saeed, R. M., Abdulmohsin, H. A. (2023). A study on predicting crime rates through machine learning and data mining using text. Journal of Intelligent Systems, 1-17. https://doi.org/10.1515/jisys-2022-0223

[2] Mandalapu, V., Elluri, L., Vyas, P., Roy, N. (2023). Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions. arXiv preprint arXiv:2303.16310.