

# Chicago Crime Rates Prediction

Anton Kamenov, Cristian Guba, Stefanos Agelastos

May 30, 2022

## Abstract

TODO

## 1 Introduction

Crime has always been a major problem in small and big cities around the world. Throughout the years, society has been trying to reduce the number of unlawful acts committed by people. By establishing new rules and punishments for those who do not obey the law, people have found a way to reduce the crime rate. Nations have evolved their legal systems during thousands of years of trial and error, but that still does not stop citizens from committing a crime.

There are many reasons why people still tend to participate in unlawful activities. Whether it be because of poor living conditions, psychological disorders or other, the world is always going to be a witness of crimes, no matter how scary the consequences of committing them are. For this reason the safety of each city is handed over to the police departments.

Police departments are another tool for lowering the amount of crimes. They operate by answering to people's calls for help and by patrolling around the city, split into designated areas. Sometimes, however, there are not enough patrolling cars in the area to react quickly to the amount of crimes happening at the same time.

## 2 Research Question

By collecting data from one of the U.S's largest city - Chicago, where the murder rate has remained persistently high throughout the years[2], this research is going to use a data-driven approach to examine

if:

- Crime is influenced by the date and time;
- Crime is influenced by the districts.

Using the analysis the research is going to try to answer the question:

- Can crime rate be predicted based on the date, time and district?

## 3 Methods

This research objectives necessitate a multi-methodological approach that integrates theory building, systems development, observation and experimentation, as described by Nunamaker et al[3]. Rather than a linear research method, this approach can be considered an agile research model, due to the continuous going back and forth between theory building, systems development, observation, and experimentation.

### 3.1 Theory building

Starting with the dataset[1], the research is going to begin with data exploration, in order to see how the data looks like. Viewing the different features the dataset has and the missing values, it will help to formulate a theory that will need to be proven. After removing or extrapolating the missing values, different parts of the data can be plotted in order to gain more insight into how the data is looking.

## 3.2 Systems development

Next part is the System development, which involves creating new features to the dataset (feature engineering), converting the one type of data to another (for example date and time to UNIX timestamp). After that models are selected for experimentation with the newly build data.

## 3.3 Observation

The Observation happens after each model is trained. By examining the result a decision needs to be made to choose which model performs the best in the given conditions with the given parameters.

## 3.4 Experimentation

After observation the research is continued by making small adjustments to the model parameters and the dataset in order to see what changes will take place and how will they influence the results.

# 4 Analysis

## 4.1 Lists

You can make nice lists in L<sup>A</sup>T<sub>E</sub>X:

- First item
- Second item

Or like this:

1. First item
  - (a) Sub element
2. Second item
  - Another sub element

## 4.2 Text Decoration

Different text decorations like *textit*, *emph* (which might render differently than here), **textbf** are possible, and it might be a good idea to find a L<sup>A</sup>T<sub>E</sub>Xcheat-sheet to find out more.

You can put things in quotes like "Hey, look at that!", or "Hey, look at that!". Notice these are not identical.

## 4.3 Quoteblocks

Quote blocks look nice:

*I believe that in about fifty years' time it will be possible, to programme computers (..) to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning.*  
[turing1950computing]

## 4.4 Tables

Things like tables are obviously supported in L<sup>A</sup>T<sub>E</sub>X, see table 1.

Table 1: Likert scale and scoring

Score	Answer
-2	Strongly disagree
-1	Disagree to some extent
1	Agree to some extent
2	Strongly agree

## 4.5 Equations

The cost function we most often use in classification is shown in Equation 1.

$$J_{\theta}(X) = -\frac{1}{m} \sum (Y \cdot \log(h_{\theta}(X)) + (1-Y) \cdot \log(1-h_{\theta}(X))) \quad (1)$$

## 4.6 Figures

# References

- [1] Spiros Politis (2018). "Chicago crimes 2001-2018 (November)". In: (2001-2018).
- [2] Matthew Friedman, Ames C Grawert, and James Cullen. *Crime Trends, 1990-2016*. Brennan Center for Justice at New York University School of Law New York, NY, 2017.

- [3] Jay F Nunamaker Jr, Minder Chen, and Titus DM Purdin. “Systems development in information systems research”. In: *Journal of management information systems* 7.3 (1990), pp. 89–106.