

Machine Learning Engineer Nanodegree

Capstone Proposal

Rachael Mahon

June 10th, 2018

Proposal

Domain Background

Being processed by the criminal justice system in the UK is an incredibly arcane and alienating experience for many of those subjected to it. Your barrister is wearing a ridiculous wig and saying words you don't understand. The majority of people making decisions about and advising you on your future are more often than not disproportionately from a different ethnicity and socio-economic background to you. It is, I'm certain, a very disempowering position to find yourself in. You are thrown into a situation in which you have no idea if you will be going to prison, how long you are likely to be there, how much you may be fined.

Historically, defendants simply had to trust that their advocate was giving them sufficient information to make decisions and had no data against which to evaluate whether they felt their lawyer had done a good job or not.

Due to the release of 1.2 million sentencing records by the UK courts, we can train an algorithm to give approximate sentencing results and find patterns based on individual characteristics.

My personal motivation for investigating this issue stems from my previous experience as a criminal defence lawyer and feeling my client's palpable stress and inability to plan their lives because of protracted criminal cases. The number of unknowns in their lives caused by their exposure to the callous machinery of the state hung over them and often led to poor decision-making, further criminal charges or spiraling into substance misuse.

I am also interested in exploring if it can be said that sentencing appears reasonably equal across gender, location and race. As Sir James Matthew once famously stated, "Justice in England is open to all, like the Ritz Hotel".

Many operators in this area are operating commercially and so the specific architecture of their algorithm is not available. However, much research is being done into this area such as:

Richard Berk, "An impact assessment of machine learning risk forecasts on parole board decisions and recidivism" *Journal of Experimental Criminology* 13(2) 193-216, 2017

J. Kleinberg, J. Ludwig, S. Mullainathan, Z. Obermeyer. "Prediction Policy Problems" *American Economic Review: Papers and Proceedings*, 105-110, 2015

Problem Statement

When accused of a crime, you do not have enough data to make decisions about whether you should plead guilty or go to trial, if you need to make arrangements about your employment, property or child care or whether you should appeal a sentence you feel is too harsh. There are too many unknowns and it is very difficult to get an accurate answer on these things.

With the input of the British sentencing data, we can determine an output of both a likely sentence ie regression problem. I had originally planned to also complete a classification problem around whether the sentence was harsh or lenient (ie. above or below average) but I was advised that this would greatly increase my workload. I will likely complete this in my own time out of personal interest and appreciate the advice of the proposal reviewer.

Datasets and Inputs

In November 2011 the British Ministry for Justice released 1.2 million records of criminal sentencing data for the majority of courts in England and Wales. The dataset is anonymised but does contain the age ranges, sex and ethnicities of the subjects. It also contains the sentencing court, the type of offence and the police force who dealt with the matter. The dataset also includes, which is why I am chiefly interested, the ultimate sentence handed down to the defendant.

The release of this data was unprecedented and was seen as a huge leap forward for transparency in a system that has always been opaque and uncertain. I find it a particularly interesting move given that open courts have long been recognized as a cornerstone of the common law. "Open justice is a longstanding and fundamental principle of our legal system. Justice must be done and must be seen to be done if it is to command public confidence," said the justice secretary, Kenneth Clarke. It must be able to be shown that justice is administered in a non-arbitrary manner without prejudice to ethnicity, sex, age or social status.

Some campaigners criticised that the data was released without the names of the defendants but I am cautious about this data as I can see very clearly a possibility of misuse. I will try to be aware of my biases throughout this process.

There are some considerable caveats and limitations to any conclusions we can draw here due to the nature of the data.

Firstly, we do not know if the defendant pleaded guilty or not guilty, and therefore do not know if the matter went to trial or not and hence if there was a plea deal or not.

Furthermore, we do not know if it was the defendant's first offence or if, for instance, the offence was committed while on bail for another offence.

We do not know if the defendant was represented by a legal aid advocate or privately paid for a solicitor or barrister.

We are also unaware of any other mitigating factors that the officer of the court charged with

sentencing the defendant may have had impressed upon them by an advocate for the defendant.

The data is available here:

<http://www.justice.gov.uk/downloads/publications/statistics-and-data/criminal-justice-stats/recordlevel.zip>

Solution Statement

If the problem is that the criminal justice system is stressful and alienating for many people being processed by it, the solution would be to use previous data to provide some insight for managing expectations or decision-making on whether or not to go to trial or to appeal a sentence. The output when this project is completed should be the capacity for someone to plug in their personal details, the offence, the court etc, and receive a reasonably accurate sentencing figure based on a regression analysis.

Benchmark Model

There are some other operators attempting to solve this problem but from a different angle. Some companies have created software for the purposes of recommending sentences to judges based on algorithms which, in my opinion, violate the right to due process as defendants and their advocates are unable to scrutinise or challenge the algorithm due to its protection by intellectual property law.

In the absence of other projects to compare against, I will use a naive model to compare my optimised results against. I think I will use a simple regression around the mean sentence for my regression task.

Cathy O'Neil, author of Weapons of Math Destruction, has roundly pointed out how problematic the use of these completely unknown and sealed algorithms are, particularly in the area of predictive policing software. These predictive models should not be treated as neutral as they are clearly influenced by the systemically problematic underpinning data and the goals and ideology of those who create and commission them.

Evaluation Metrics

I intend to use R squared to evaluate the performance of the regression model along with root mean square deviation for additional evaluation.

Because we have such a large number of records, I think it will be sensible to have a reasonably large testing set.

Project Design

It will be important here to strip out outliers due to the number of limitations of the data as stated in data and inputs section. There may be a number of very harsh sentences or very lenient sentences due to extenuating circumstances and these will need to be ignored.

For the regression problem I would like to use linear regression as the sentence to be predicted is

estimated from continuous variables. For this problem, I would like to explore closely the high bias / high variance trade off to find the best linear regression possible. I would also like to visualise the different options available with complexity curve graphs. If I feel this is becoming overly complex, I may decide to use a grid search to find optimal hyperparameters. If I feel it is necessary, I may cross-validate my linear regression with k-fold cross validation.