



Forecasting Iowa Prison Populations

William Ebert
GT

Greyson Henderson
GT

Shivani Kharbanda
GT

Kristi Rasmussen
GT

Grant Ruedy
GT

Husam Yassin
GT

Summary

The Iowa Department of Corrections needs to quickly identify how potential legislative decision changes will impact Iowa prison populations. Today, forecasts are manually done in Excel and take weeks to build properly. Our goal is to improve this process, not only through automation, but also by enabling insight discovery through an interactive tool leveraging flexdashboard in R.

Data

Public data published by the Iowa Department of Corrections.

Leveraged 3 separate datasets:

- Prison Admissions (16 features, 49k records)
- Prison Releases (18 features, 56k records)
- Current Prison Population (18 features, 8k records)

Using these three datasets, we derived an “exploded” dataset with a monthly population counts for each offense code, from 2010 through current. This is the dataset from which we created our forecasts.

Processing Method

The process of exploding the data was done using vectorized operations to augment the data in parallel allowing for efficient, concise, easier to read code. This way enabled us to account for each year-month someone spent in prison. This method was also leveraged in the forecasting process, as well, to compute across all subpopulations.

Impacts of Covid-19

On the data:

Following the February 3rd, 2020 declaration of a public health emergency by the US:

- Iowa prison admissions decreased
- Iowa prison releases increased (parole in particular)
- Net result: significant decrease in Iowa prison populations

On our forecasts:

Our initial forecasts appeared to be impacted by the overall prison reduction, and resulted in significantly decreased population forecasts.

How we handled it:

Models were trained with Covid-19 timeframes both included and excluded, and then compared to determine the optimal model.

The Dashboard

The interactive dashboard allows a user to drill down to the offense classification and offense type granularity to visualize specific forecasts and various demographic breakdowns. A major feature within this tool is enabling the user to adjust forecasts based on expected impact of legislative decisions. Once adjusted, the user can infer the potential influence a decision may have on certain demographics.

The tool is built in R using the flexdashboard package with a shiny runtime. Using this framework allows for a high degree of interactivity for the user and enables us to dynamically control the visualizations based on user input. RStudio has a service, shinyapps.io, which gives us a place to host the dashboard. Once published to their server, the entire process can be automated on a scheduled job.

This will be adopted by the Iowa Department of Corrections in place of their current process, which takes weeks to build each time they need to reforecast. Not only will this save time, but it also reduces the risk of error by ensuring reproducibility each run without a decrease in performance.

Forecasting Populations

The forecasting models were built at the offense classification and offense type level of granularity. This decision was made based on the requirements from the Corrections team and data sparsity concerns of going into deeper detail across subpopulations. Much like how we augmented the training data, we used a similar vectorized approach to train, evaluate, and score each model.

The models we evaluated included different variations of ARIMA, exponential smoothing, and linear regression. We selected these models based on their simplicity to implement and understand, but also because the overall distributions of the data suggested simpler models could likely achieve good performance. Ultimately, we chose to move forward with ARIMA for our forecasts .

Model performance varied across subpopulations, as expected, but overall our approach holds up well relative to baseline metrics and the current approaches by the Corrections team.

Model	MAE	RMSE	MSE	AIC
ARIMA	75.4	110.8	12296.8	1734.4
Exponential Smoothing	76.3	108.5	11775.1	2049.8
Regression	708.0	966.0	934044.0	2425.7

