



DEVELOPMENT OF REGRESSION MODELS FOR NATIONAL CRIME FORECASTING

Nikola Petroski, Aleksandra Dedinec, Sonja Filiposka, Anastas Mishev,
Faculty of Computer Sciences and Engineering,
Macedonian Academy of Sciences and Arts, Skopje, N. Macedonia



Content

- Introduction
- Data
 - Analyzing data
 - Structuring the data
- Methodology
 - Classification model (in previous study)
 - Regression model
- Results and conclusions
 - From the regression model
 - Comparison of classification and regression models



Introduction

- Crime rates forecasting
- Patterns in crime
- Classification problem
- Regression problem

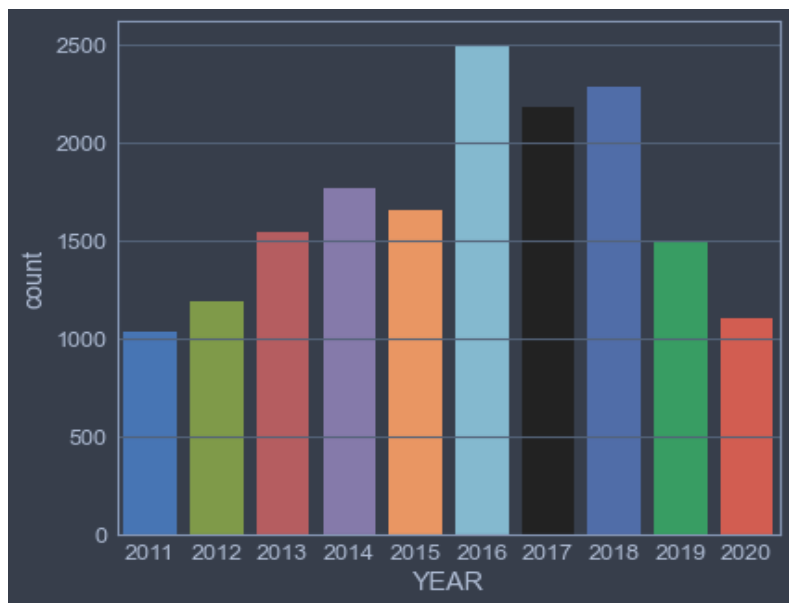


Data

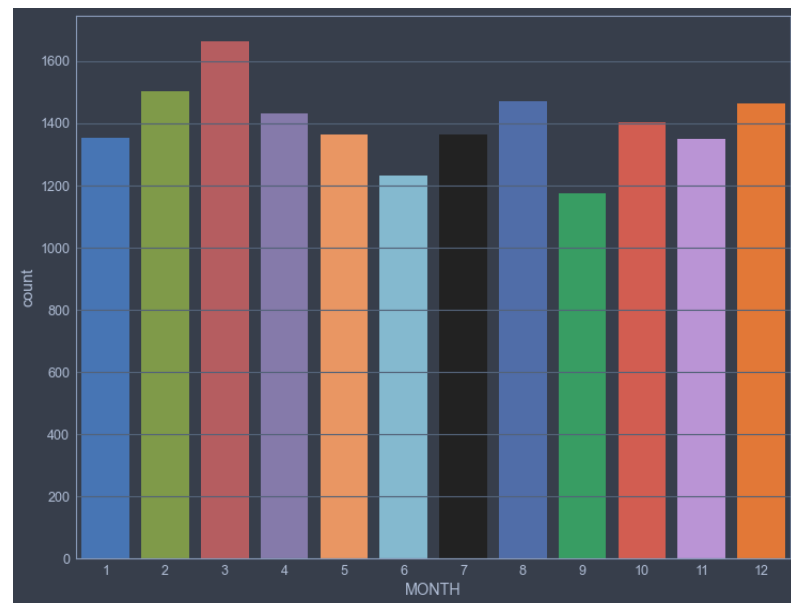
- Data from 2011 until now
- Analyzing data

Analyzing Data

- Distribution of crimes per year

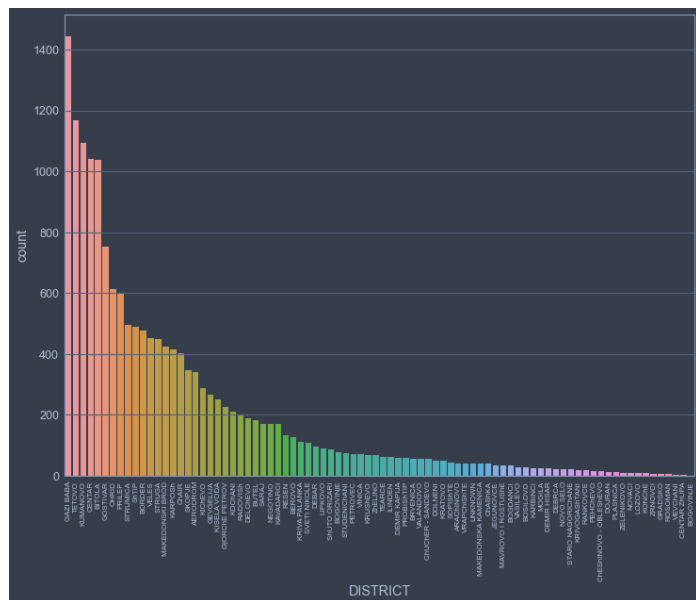


- Distribution of crimes per month

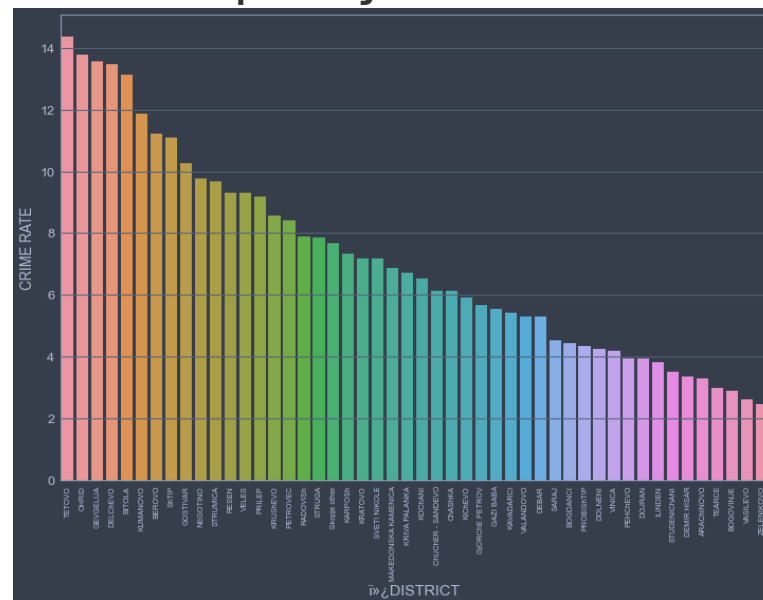


Analyzing Data

- Distribution of crimes per municipality

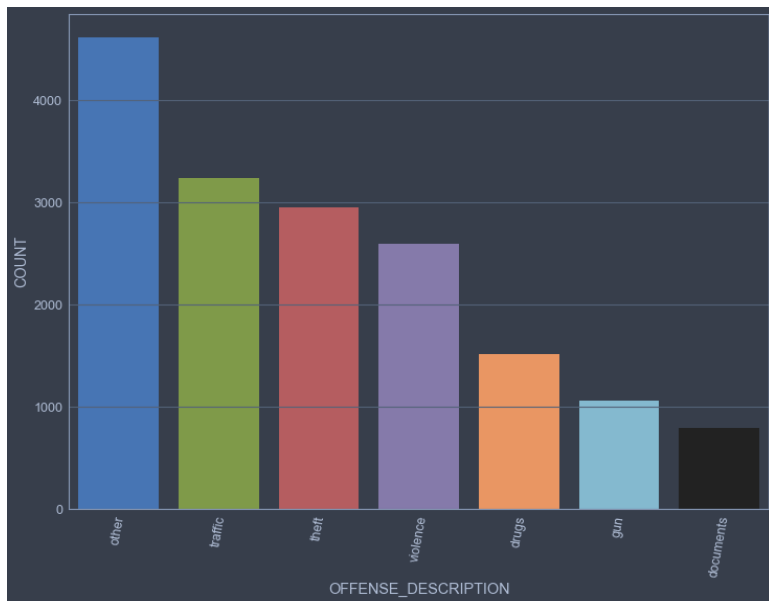


- Number of crimes/population per municipality



Analyzing Data

- Distribution of the crime per category



- Seven categories:
 - Traffic
 - Theft
 - Violence
 - Drugs
 - Gun
 - Documents
 - Other



Data

- Data from 2011 until now
- Analyzing data
- Grouping crime categories
- Clustering spatial data
- Aggregated historical data



Previous Study

- Classification problem
- Gradient boosted decision trees model
- Positive data threshold – 30% of predictions positive
- Score

Score - Classification

- Confusion matrix of the training data

	Negative Pred.	Positive Pred.
Negatives	16575	4243
Positives	21	326

- Confusion matrix of the test data

	Negative Pred.	Positive Pred.
Negatives	6741	2272
Positives	26	174

Score - Classification

		precision	recall	f1-score	support
• Classification report for training data	0.0	1.00	0.80	0.89	20818
	1.0	0.07	0.94	0.13	347
	accuracy			0.80	21165
	macro avg	0.54	0.87	0.51	21165
	weighted avg	0.98	0.80	0.87	21165
		precision	recall	f1-score	support
• Classification report for test data	0.0	1.00	0.75	0.85	9013
	1.0	0.07	0.87	0.13	200
	accuracy			0.75	9213
	macro avg	0.53	0.81	0.49	9213
	weighted avg	0.98	0.75	0.84	9213



Regression Model

- Using same data
- Testing out models and tuning parameters
- Models with multiple outputs



Score - Regression

- Scoring the model's performance
- Mapping regression results to binary results
- Setting threshold – 30% of predictions positive
- Scoring the model with classification metrics

Score - Regression

- RMSE, MAE, R2 are scored on the continuous results
- recall and precision are scored on the mapped binary results

	rmse	mae	r2	recall	precision
0	0.145602	0.030707	0.090336	0.925	0.066691

- Confusion matrix of the test data

	Negative Pred.	Positive Pred.
Negatives	6424	2589
Positives	15	185



Analyze Results

- Observe negatively predicted positive cases
- Compare models
- Where each model goes wrong

Wrong Predictions

- Wrong predictions by each model, grouped by the predicted outcome

	Wrong Predictions	Wrong Negative Predictions	Wrong Positive Predictions
classification	2298.0	26.0	2272.0
regression	2604.0	15.0	2589.0



Multiclass Classification

- Map regression predictions to classification
- Cases: 0 crimes, 1 crime, >2 crimes
- Choose threshold – 28% for 1 crime, 2% for >2 crimes
- Observe results

Observe Results

- Results from multiclass classification mapping

	Zero Pred.	One Pred.	Two or More Pred.
Zero	6424.0	2464.0	125.0
One	14.0	108.0	38.0
Two or More	1.0	16.0	23.0



THANK YOU FOR YOUR ATTENTION!