

D10 - Crime reporting

Predicting crimes using machine learning and classifiers

LTAT.02.02 course project

Rasmus Saame, 6. rühm

Toomas Roosma, 6. rühm

Table of Contents

1. Setting up.....	3
2. Business understanding.....	4
Identifying your business goals.....	4
Background.....	4
Business goals.....	4
Business success criteria.....	4
Assessing your situation.....	4
Inventory of resources.....	4
Requirements, assumptions and constraints.....	4
Risks and contingencies.....	5
Terminology.....	5
Costs and benefits.....	5
Defining your data-mining goals.....	5
Data-mining goals.....	5
Data-mining success criteria.....	5
3. Data understanding.....	6
Gathering data.....	6
Outline data requirements.....	6
Verify data availability.....	6
Define selection criteria.....	6
Describing data.....	6
Exploring data.....	6
Verifying data quality.....	6
4. Project plan.....	7
Plan.....	7
Tools and techniques.....	7

1. Setting up

Our repository is hosted in GitHub and is shared with our instructor Anna Aljanki. Repository can be found in [this link](#).

2. Business understanding

Identifying your business goals

NB! Since we chose this project simply out of interest we don't have a big company behind us that's depending on our work. Our background analysis is made up in one possible way what the business logic and requirements for the project might be.

Background

Every day multiple public order breaches are committed. So that the crimes wouldn't go unpunished, police have to counter actions. One of the options is to punish criminals for their crimes. Another option is to prevent crimes by increasing the presence of the police officers in places where breach of the public order takes place most often.

Business goals

The Police expects to reduce the crime rates in the next 10 years by trying different methods. Police would like to reduce crimes by 5% each year and by 40% for the end of 10 year period.

Business success criteria

The success will be measured by counting the times and places where public order breaches take place. Crime rates change will be measured in each location separately and in Estonia total as well. The goal has been achieved when there are some locations where crime has dropped by previously mentioned amounts.

Assessing your situation

Inventory of resources

In our team we have 2 very dedicated data scientists Rasmus and Toomas who are going to work on the project.

We have data about public order breaches in the last 10 years (starting from 2012) from [this source](#).

For software we are going to use Python, Jupyter Notebook and different Python libraries for data mining, such as pandas, numpy and sklearn, and Google.

For hardware we are going to use our HP laptops from Institute of computer science to do all of the calculations and presentations.

Requirements, assumptions and constraints

The data is distributed with Creative Commons 3.0 license and is available for sharing and modifying if properly referenced. At the end of the project we need to have a working model predicting what type of crime is most probable in certain location and statistics about public order breaches in the last 10 years.

Risks and contingencies

Possible risks include hardware failures, if our laptops break we can't work on the project. In this case we can find substitute computers and restore our progress from GitHub.

Terminology

Public order breach (Avaliku korra rikkumine) – Action against the law that take place in public room.

Petty theft (Pisivargus) – Small theft where stolen value is less then 40€.

Misdemeanor (Väärtegu) – Smaller offense for which the usual punishment is fine or arrest.

Crime (Kuritegu) – More significant offense for which the usual punishment is fine or imprisonment.

Costs and benefits

Possible costs include data scientists' time and effort. Different nutrients and drinks for developing also require some amount of money. Possible benefits include deeper knowledge about public order breaches, statistics about breaches, model predicting breaches, lower crime rates and better course project and grade.

Defining your data-mining goals

Data-mining goals

In the end of the project we'd like to have different models which predict the likeliness of a crime on a given time and location. In addition to that we'd like to have statistics and graphs about last 10 years offenses that happened in Estonia. Map plot about the likeliness of different offenses in Estonia.

Data-mining success criteria

Predicting crimes on a test set with a greater then 0.6 AUC. Statistical knowledge about different offenses that took place. Graphs that illustrate on a interactive map offenses in the last 10 years.

3. Data understanding

Gathering data

Outline data requirements

Verify data availability

Define selection criteria

Describing data

Exploring data

Verifying data quality

4. Project plan

Plan

1. Investigate data – Purpose of this task is to get familiar with the data
2. Clean data – Purpose of this task is to select only relevant fields from the data and verify that the data is usable and make possible corrections.
3. Decide on different approaches – In this task the target is to talk and discuss about different approaches for this project and choose models that we are going to try to implement.
4. Implement different models – In this task we are going to implement different models chosen in previous task and test with parameters to find the best models.
5. Analyze built models – Find how well the models behaved and calculate their efficiency. This can be done in parallel with task 4 when certain model is ready.
6. Build graphs (interactive) to visualize the results – This task's purpose is to make our results presentable.
7. Present the project – This task consists of presenting the project at project session on December 16th.

Tools and techniques

We are planning to use Python3 and different package made for manipulating data, such as numpy, pandas and sklearn. We plan to consider using different methods learned in this course such as KNN, RF and much more for best possible predictions.