Roadmap for AI-driven Policy Dashboard

Due to this course (EPA221 Societal Challenge Project) being only 5 ECTS the end product cannot be created in full due to a lack of time. However, we as group 4b want to make a roadmap for the people who are interested in this project and want to continue with it. Now that the problem of youth criminality with explosives is evident and the possible innovative way to help combat this problem has been addressed this paragraph will ensure the longevity of the proposed product.
To ensure this AI-driven Policy Dashboard is properly implemented by future researchers the following roadmap needs to be followed. Note that this roadmap is a general roadmap, once future researchers start working on the AI-driven Policy Dashboard they will need to make their own assessments as well. Figure 1 displays the roadmap conceptually, with text and explanations in the different paragraphs of this chapter.

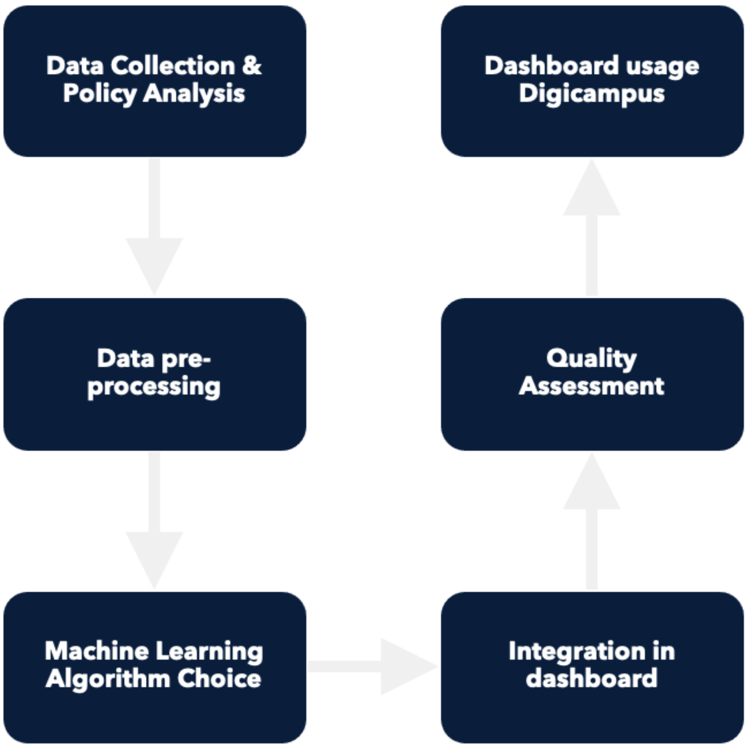

Figure 1: Conceptual roadmap for further analysis

Purpose of the product

This product is made to analyse youth criminality with explosives which has been a growing problem in the Netherlands. Especially, the four largest cities in the Netherlands (Amsterdam, Rotterdam, The Hague and Utrecht) suffer from explosive related youth criminality. Acknowledging this problem we tried to find an innovative way to address and possibly tackle explosive youth criminality. We tried to find answers in publicly available data from multiple ministries like Zicht op Wijken, Zicht op Ondermijning, Onderzoek en Statistiek Amsterdam and Den Haag in Cijfers. To address the issue of fragmented policy interventions at the municipal level, we also analyzed neighborhoods in Utrecht and The Hague, examining their specific policy interventions to facilitate mutual learning and collaboration.

With this data one can create a data-driven AI policy tool which gives neighborhood specific policy interventions based on publicly available data and insights of experts on youth problems and policies. In other words, this product aims to automate policy interventions using Machine Learning (ML) algorithms based on socio-economic features of the neighbourhood. When implemented correctly, policy makers will see actionable policy interventions without having to analyse complex data structures and relationships.

Statistical Analysis and Policy Analysis

The first step in preparing ML models is to collect data from multiple sources (like the ones mentioned in "purpose of the product") and preprocess this data. The types of data that will need to be gathered are explosives being placed by youths and socio-economic variables, known as 'Onderscheidende kenmerken' in the *Zicht op Ondermijning* dashboard. Missing values need to be handled and certain categorical variables will need to be (dummy)-encoded.
Since DigiCampus works together with Dutch ministries, one might refrain from using black box ML models due to their lack of transparency. White box models like a DecisionTree are more transparent, however they are more prone to overfitting. One of the challenges therefore will be the consideration of less transparency but a better generalizable (less overfitted) model.
Since ML models are different from classical statistical models there does not need to be causation of independent variables on the dependent variables, but the weights are not interpretable (Cranenburgh, 2023). However, the ML algorithm can provide valuable insights into patterns that are currently undiscovered. The outcomes of the ML model will need to be validated by a proper classical statistical model.
Furthermore, the researchers must conduct interviews with experts in the field of youth criminality. It is crucial to get a full overview of why interventions do or do not work. The knowledge gained by this qualitative analysis will ultimately lead to an algorithm, which determines which intervention is recommended to each neighbourhood.

One way this algorithm can work is via a decision tree. By answering a sequence of many simple questions, the AI will find a fitting intervention. The following questions would be examples of such questions:

- Which determining factor is most occurring in this neighbourhood?
  - Determining which social problem seems to lead to criminal activity in this neighbourhood
- How many youths are in the target group for a possible intervention?
  - Determining the scale of the intervention, based on predictors for youth criminality
- What budget is available for a possible intervention?
  - Determining whether a possible intervention can actually be executed.

The actual insights gained from qualitative analysis will determine how the algorithm would look, and whether a decision tree is fitting or not.

So next to the future analyst needing insights into ML models he also needs to conduct a proper policy analysis of different municipalities. If a municipality has a certain policy intervention in place the policy must be listed to later be used in the ML model.

Training the ML model

Like previously stated the data must be pre-processed by eliminating or filling missing values. Different categorical variables will need to be encoded using either dummy encoding or one-hot side encoding. Apart from the socio-economic variables the model will need to be trained on policy interventions. One potential challenge lies in effectively encoding policy interventions. Once the link between socio-economic variables and policy interventions is established, the choice of ML model becomes crucial. Starting with a Decision Tree could be an appropriate approach to identify which policies might be effective in a given neighborhood. The choice that the future analyst must make is whether to make a Classifier ML model or a Regression ML model. The regression model will be useful for predicting whether an explosive can happen in the neighborhoods, while a classifier model will be useful for identifying if a neighbourhood has a high risk or low risk of explosive youth criminality. If a neighborhood is initially classified as "high risk" and is reanalyzed after the implementation of policy interventions, and it is subsequently classified as "low risk," this could indicate the effectiveness of the policy. In order to achieve this the dashboard must be regularly fed with new data.

Stakeholder Involvement

One of the most important necessities to effectively create this product is good cooperation between the four key players of the Quadruple Helix Model of DigiCampus. First of all, DigiCampus must find a motivated master student or employee with proper background to further analyse the product. Thereafter, DigiCampus must closely work together with certain ministries that store the information necessary for analysing neighborhoods with regard to explosives in the Netherlands. If proven useful the dashboard can result in more citizen participation. For example, policy interventions like “Bondgenoten” in Utrecht increase the participation of citizens in the neighbourhood which might eventually lead to less explosive youth criminals. Integration of citizen participation of the Quadruple Helix Model can therefore be essential in order to tackle youth criminality. Furthermore, private authorities like creditors or sport clubs can also provide useful policy interventions to tackle explosive youth criminality. Therefore, it is essential for DigiCampus to stay proactive in the collaboration between actors.

Potential pitfalls

Uncertainty
The structure of the algorithm has a big impact on the results. It is therefore important to write multiple algorithms, and analyse how structural changes lead to different results. The final decisions on the algorithm should be clearly described in a separate section of the dashboard. As this dashboard will be used by governments, the decision making process must be open.

Example of Roadmap

Let's assume for simplicity that all the data has been gathered and that the analysis can begin. Let’s further assume that the fact that someone is an “early school dropout” has a positive influence on him or her placing explosives. Let's also assume that Halt is considered an important actor with proper policy interventions like the “school dialogue” to prevent youths from going into criminality.

The proposed product should indicate on the dashboard which neighbourhoods are affected most (see Figure 2). Based on this result, and the fact that Halt has proper policy intervention measures, the dashboard should show the “school dialogue” of bureau Halt as a proper policy intervention measure.

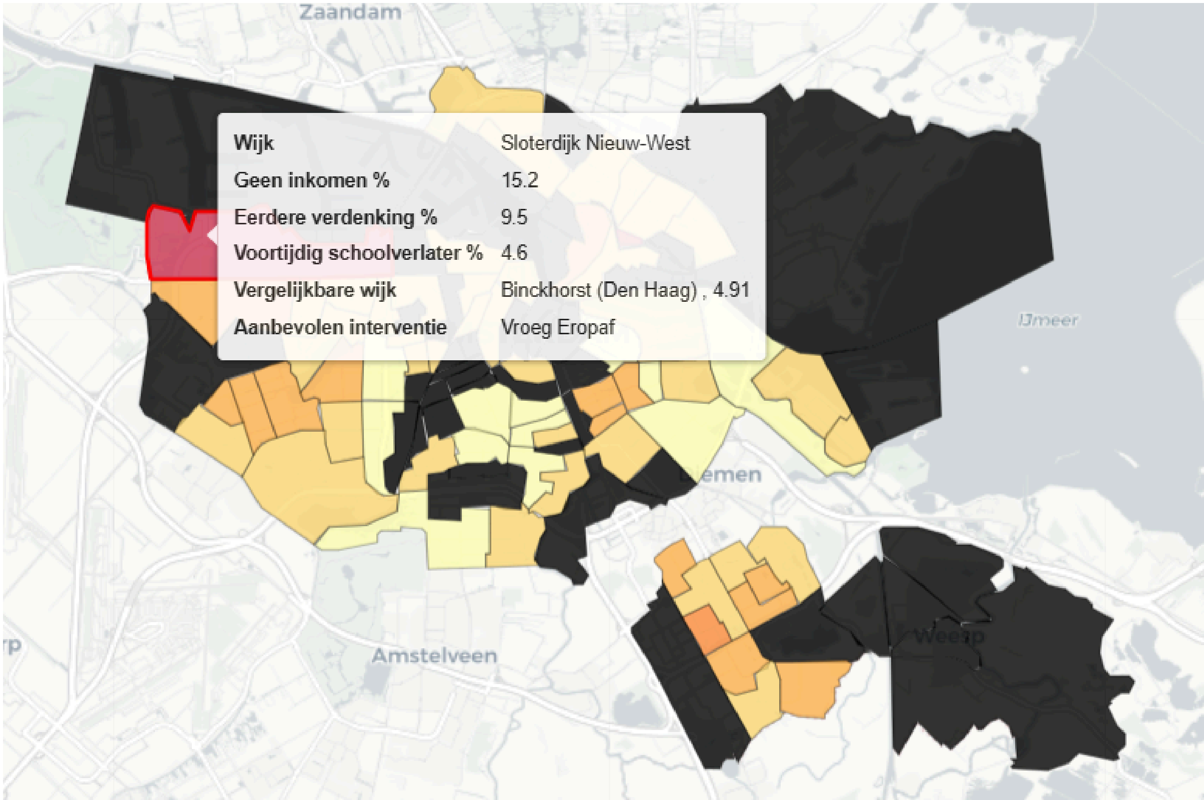

Figure 2: Example of what the dashboard shows based on different data

Figure 2 also highlights that "no income" has a higher percentage value compared to "early school leavers." This difference leads to a distinct policy intervention, shifting the focus from school-related measures, such as those seen in the Halt case, to addressing "no income" issues, for which the "Vroeg Eropaf" policy is designed. Bear in mind that the above figure is a proof of concept, the values will need to be substantiated even further.

Requirements for future researchers

- Have a good understanding of Machine Learning Algorithms and how they can be used to assess data in combination with policy interventions
- Have a good understanding of public policy of municipalities to train the AI model on both data and policy interventions
- Have good communication skills to interview relevant stakeholders

Literature

Cranenburgh, S. (2023). Machine Learning for Socio-Technical Systems [Slide show; Brightspace]. https://brightspace.tudelft.nl/d2l/le/content/683054/viewContent/4043137/View