

Reasons for Implementing the System

The Artificial Intelligence system helps prioritize hundreds of reports coming from neighbors or rental platforms so that the limited enforcement capacity can be used **efficiently** and **effectively**.

Role of the System - Workflow

1. A citizen or rental platform **submits a report**
2. The AI system calculates the **probability of housing fraud**.
3. A **visualization is given of the features** that resulted in high or low risk of fraud.
4. The **responsible supervisor** determines if there is a case of illegal housing through a **preliminary research and field investigation**.

Potential Harms

Good-quality data has been used, ensuring that it does **not contain biases**. The system naturally has an impact on the alleged offender, as the report on their offense get more priority, Risk mitigation has been performed through **continuous monitoring** in the pilot phase.

Data sources

1. **Identity and housing rights data** from the Personal Records Database
2. **Buildings data** from the Registry of Addresses and Buildings
3. **Data from any related illegal housing cases**

Data Features

1. **Identity and housing rights data:** Name, date of birth, gender, date of residence in the city, date of residence in the address, family composition, date of death
2. **Building data:** Address, street code, postal code, description of the property, type of home, number of rooms, floor surface area, floor number, number of building layers, description of the floor
3. **Related illegal housing cases:** Starting date of report, stage of investigation, report code number, violation code number, investigator code number, anonymous reporter yes/no, situation sketch, user that created/edited the report, handling code number, date when case closed, reason why case closed.

Model Architecture

The system relies on a model that **finds relationships and patterns** in a large amount of information about illegal housing. The model calculates which information can be associated to illegal housing and to what degree. This type of model is called “**random forest regression**”

Decision Explanation

Data features that contributed to the decision (positive (+) means that it contributed to high fraud risk):

- **Street code +3.87**
- **Anonymous reporter yes/no +2.5**
- **Description of the property +0.95**
- **Floor surface area +0.63**
- **Type of home +0.62**
- **Number of rooms +0.61**
- **Date of residence in the address +0.52**

Performance

The “random forest regression” model is a fairly complex model that can **approximate reality quite well**. In order for the model to remain **generic** (not overfit), research has been conducted to know how many layers the model needs to have.