Case description

Overview

- Amsterdam has limited living space for both citizens and visitors.

- Citizens who want to rent out their home to tourists must meet certain requirements.

- A maximum of 30 nights per year and 4 people at a time is allowed, and it must be reported to the municipality.

- The municipality receives reports of possible illegal holiday rentals and investigated with the help of an algorithm pilot program starting July 2020.

- The algorithm analyzes data from related illegal housing cases of the past 5 years to calculate the probability of an illegal holiday rental situation.
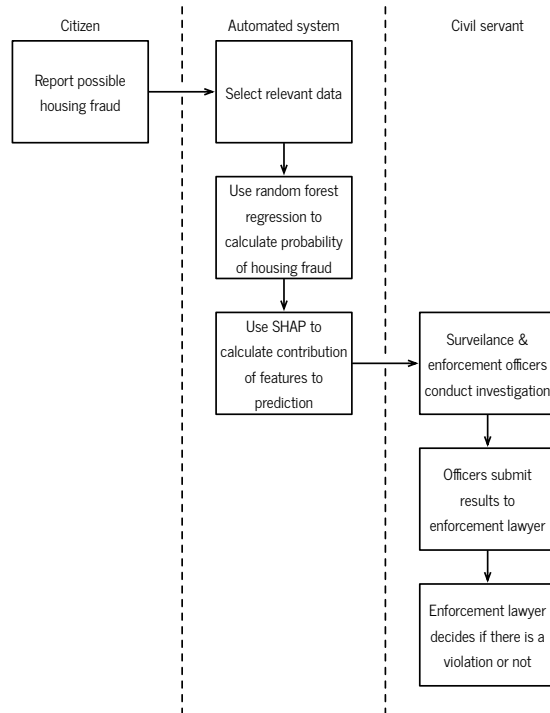

Figure 1: Flowchart of algorithmic system used for enforcement of illegal vacation rentals. Adapted from the original by Linda van de Fliert.

Functional description

Data

Identity and housing rights data: Minimized dataset from the Personal Records Database (BRP), showing information about the identity and housing rights of the residents; specifically: name; date of birth; gender; date of residence in Amsterdam date of residence at the address; family composition; date of death.

Buildings data: Minimized dataset from the Registry of Addresses and Buildings (BAG), showing information about the building; specifically; address, street code, postal code; description of the property; Amsterdam BAG-code, national BAG-code; the type of home (rent, social rent / free sector, owner-occupied); number of rooms; floor surface area; floor number on which the front door of the apartment resides; number of building layers; description of the floor of the residential property.

Prior illegal housing cases: Data from any related illegal housing cases; specifically: starting date of investigation / report; stage of investigation; report code number; violation code number; investigator code number; anonymous reporter yes/no; situation sketch; user that created the report (including date), or edited the report (including date); handling code number (type of case, allocation to team); date when case closed; reason why case closed.

Model

- A "random forest regression" algorithm has been developed to find patterns in a large amount of information about illegal housing.

- The algorithm calculates the probability of illegal holiday rental at an address based on related illegal housing cases from the past 5 years.

- The algorithm uses the probability tree principle to perform mathematical calculations and take an average to generate the expectation of illegal holiday rental at an address.

- The "SHAP" (SHapley Additive exPlanations) method is used to explain the features in the data that resulted in high or low suspicion of illegal housing so that employees can make a well-considered decision.

- The algorithm must be carefully calibrated to avoid overfitting and categorizes continuous data points to better reach a conclusion.

Non-discrimination

- The algorithm was developed using a privacy impact assessment to ensure that sensitive information is not included.

- The dataset used in the algorithm only includes critical information to determine if the Housing Act is violated.

- The data used for the algorithm comes from previous illegal holiday rental cases to ensure good-quality data.

- The algorithm may indirectly lead to undesirable differences in treatment between cases, so the AI Fairness 360 toolkit is used to address algorithmic bias during the pilot.

- Further research will be conducted to ensure that the algorithm is fair and unbiased.

Human oversight

- Automated decision-making is not used in the investigation of suspected illegal holiday rentals.

- The algorithm assists the employee in identifying the most probable cases of illegal holiday rental, which they can then prioritize for field investigation.

- A visualization of the algorithm's risk assessment is provided to the employee to help them decide whether to follow its recommendation or not.

- The supervisor and project enforcer are responsible for determining if there is actually a case of illegal housing.

- The algorithm has a significant influence on the planner, but it does not make independent decisions. Employees receive training to recognize the opportunities and risks of using algorithms.

More information

- City of Amsterdam private vacation rental service (in Dutch): `https://www.amsterdam.nl/wonen-leefomgeving/wonen/vakantieverhuur/`

- This system's Algorithm register entry: `https://algoritmeregister.amsterdam.nl/en/illegal-holiday-rental-housing-risk/`

Appendix

Amsterdam tourism policy

- The city of Amsterdam considers tourism as an important source of income and employment, but also acknowledges its negative impact on the livability of the city.

- The city is working on a sustainable tourism policy to ensure that tourism contributes to the quality of life in the city without causing overcrowding, nuisance, or damage to the environment.

- The measures being taken by the city include limiting the growth of hotels and holiday rentals in busy areas, promoting lesser-known neighborhoods and attractions, and stimulating sustainable modes of transport.

- The city also aims to tackle the nuisance caused by large groups of tourists and monitor the impact of tourism on the city to adjust policy if necessary.

- More information about the city's tourism policy can be found here (in Dutch): `https://www.amsterdam.nl/bestuur-organisatie/volg-beleid/toerisme/`

Random forest regression

- Random forest regression is a type of machine learning algorithm used for regression tasks, which involve predicting a continuous numerical value.

- It works by constructing many decision trees and aggregating their predictions to make a final prediction.

- Each decision tree is trained on a random subset of the available features and a random subset of the training data.

- The randomness introduced in building the trees helps to reduce the risk of overfitting and improves the accuracy of the model.

- Random forest regression is a popular and powerful tool for predictive modeling in many fields, including finance, healthcare, and engineering.
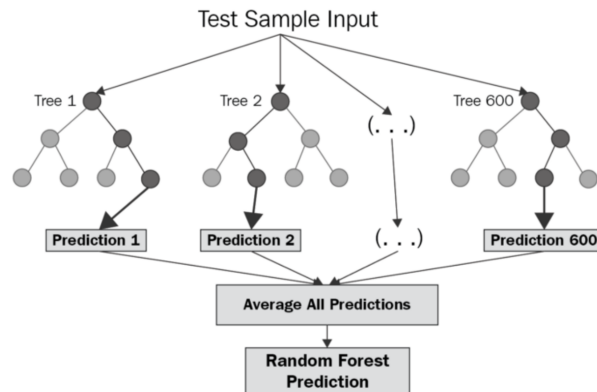

Figure 2: A random forest regressor is constructed from multiple decision trees, the predictions of which are averaged. (Source: Keboola.)

SHAP

- SHapley Additive exPlanations (SHAP) is a method for explaining the output of machine learning models.

- It is based on game theory and uses the concept of Shapley values to assign importance to the input features that contribute to the output of the model.

- The method calculates the contribution of each feature by comparing the model's predictions with and without that feature.

- SHAP produces a set of explanations for each data point, which can help to interpret and understand the model's behavior.

- It is a flexible and model-agnostic method, meaning it can be applied to a wide range of machine learning models.

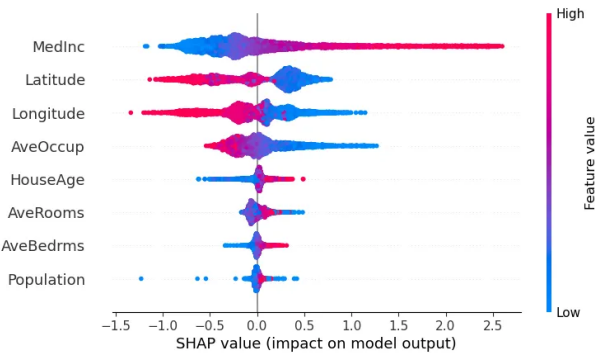- More on SHAP: `https://shap.readthedocs.io`

Figure 3: Explaining feature impacts on model output of a housing price prediction random forest regressor using a SHAP beeswarm summary plot. (Source: Towards Data Science.)

Shapley values

- Shapley values are a concept from cooperative game theory.

- They measure the marginal contribution of each player in a coalition game.

- Shapley values have been adapted to machine learning as a way of assigning importance to input features in a model.

- In this context, the input features are treated as players in a game, and the output of the model is treated as the payoff.
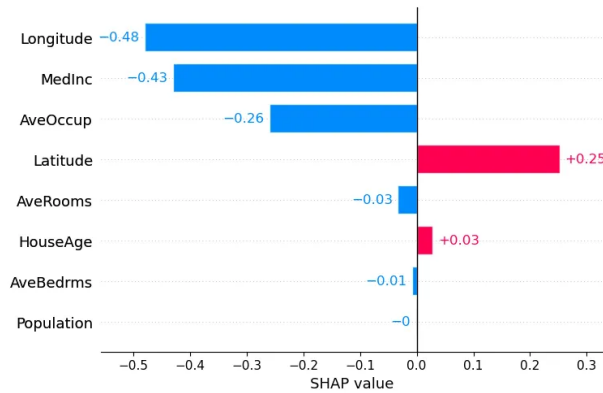
Figure 4: Explaining feature impacts on a single housing price prediction by a random forest regressor using a SHAP bar plot. (Source: Towards Data Science.)

AI Fairness 360 toolkit

- The AI Fairness 360 (AIF360) is an open-source toolkit developed by IBM to help detect and mitigate bias in machine learning models.

- The toolkit includes a comprehensive set of metrics, algorithms, and tutorials that can be used to analyze and mitigate various forms of bias.

- The AIF360 toolkit can be used across various industries and domains to promote fair and trustworthy AI systems.

- The toolkit includes various components such as data preprocessing, bias detection, bias mitigation, and bias visualization.

- The AIF360 toolkit has been used in various real-world applications, including hiring and lending decisions, to ensure that AI systems are fair and unbiased.

- More on AIF360: `https://aif360.mybluemix.net`