

# ESTIMATING BOLT USER PENETRATION IN CITY DISTRICTS

## **Members:**

Rasul Agharzayev, Orkhan Jamaladdinov, Taavi Pipar, Kristjan Ševerev

## **Project owner:**

Polina Shevchenko



# PROBLEM STATEMENT

- **Why?**

- Better resource distribution between city districts in Tallinn
- Improve our skills by solving real life problem with machine learning

- **How?**

- Bolt data and data from external sources
- Machine Learning model (regression)



# APPROACH

- Official statistics
- Land use
- Social media usage
- Bolt usage



# DATA

## Spatial data

- Grid 1x1 km

## Descriptives

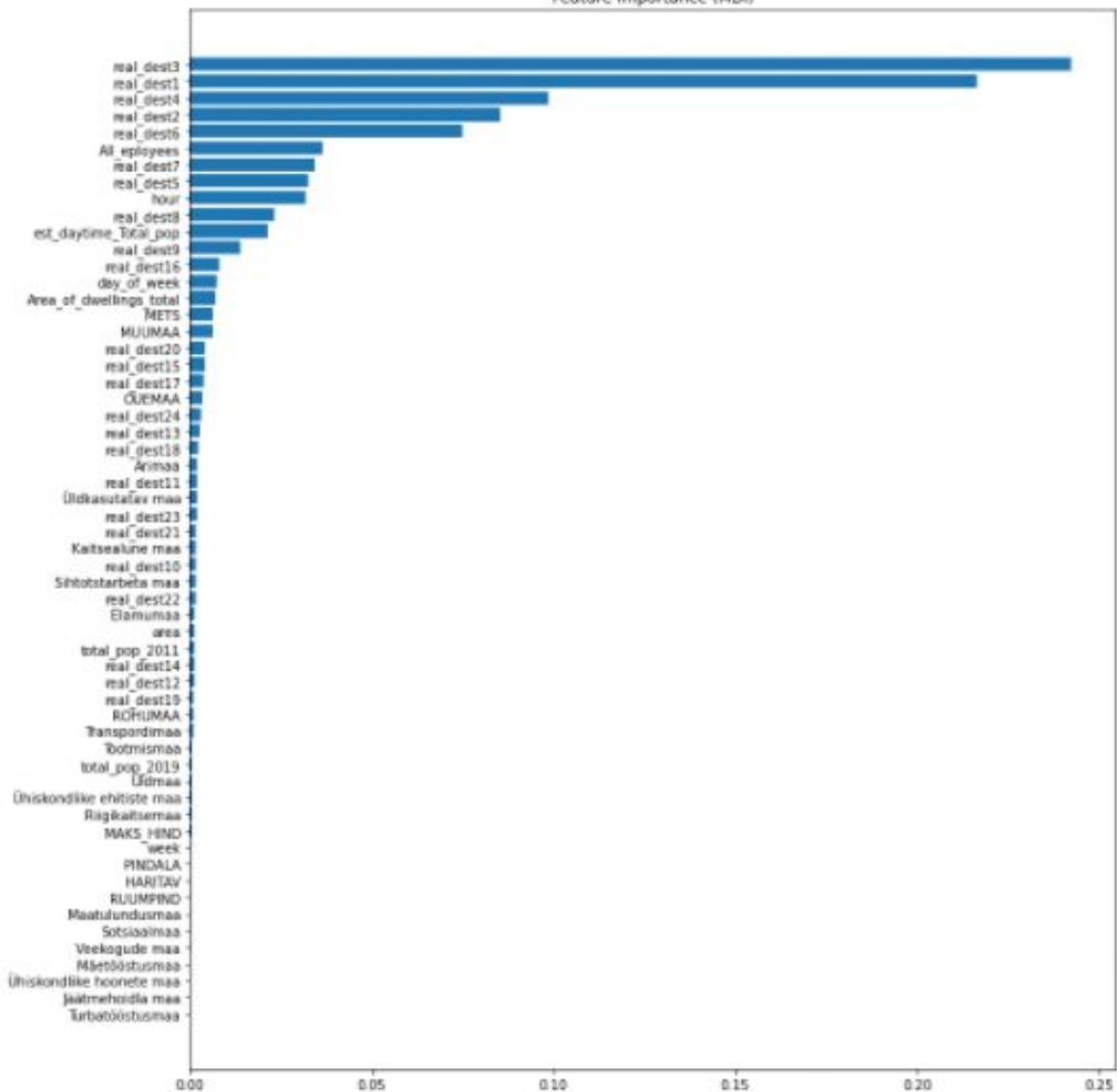
- Time unit
- Population
- Land usage
- Official statistics



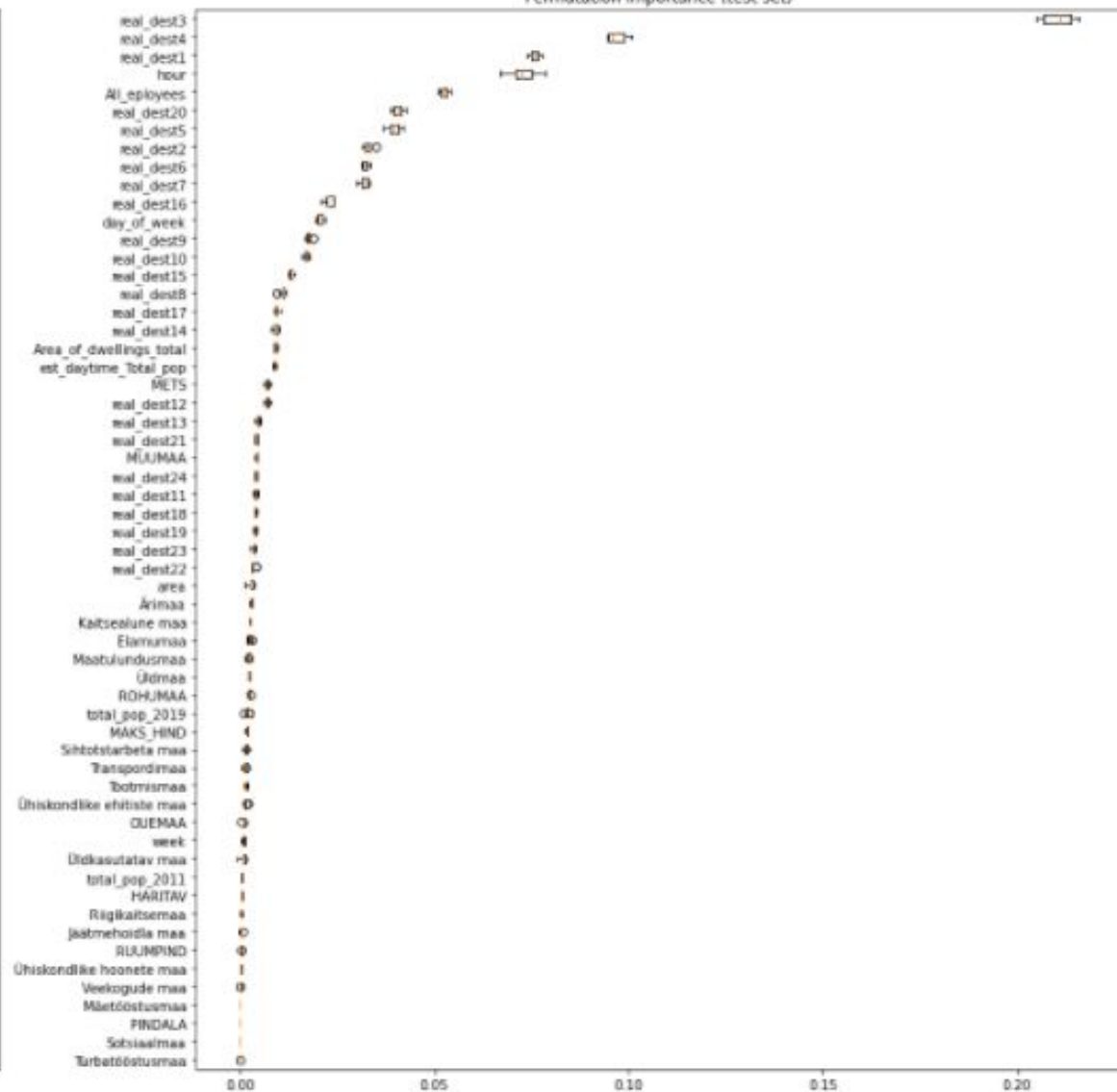
| Model  | Performance   |
|--|---|
| <ul style="list-style-type: none"><li>▪ Time unit 1h</li><li>▪ Gradient Boosting Regressor<ul style="list-style-type: none"><li>'n_estimators': 450,</li><li>    'max_depth': 4,</li><li>    'min_samples_split': 22,</li><li>    'learning_rate': 0.2</li></ul></li></ul> | <ul style="list-style-type: none"><li>▪ RMSE: 10.3164</li><li>▪ Score: 0.9116</li></ul> |



Feature Importance (MDI)



Permutation Importance (test set)



# OUTCOME

- In order to predict penetration, we need to know where people are located in a particular area and time
- The smaller the area and time, the fewer factors affect it and the easier it is to give accurate results



**THANK YOU/AITÄH**

