Neighborhood types relate to income levels in subdistricts of Tallinn, Estonia

Oleksandr Karasov, Evelyn Uuemaa, Tiit Tammaru. Department of Geography, University of Tartu, Estonia

INTRODUCTION

We **aim** to predict income for subdistricts of Tallinn – the largest Estonian city – by using remote sensing (RS) and Google Street View (GSV) panoramic imagery.

RESEARCH QUESTION

- what are the optimal RS- and GSV-based proxies for income predicting?

DATA

Sentinel-2 cloud-free summertime 2018 mosaic, 21 896 **GSV** photographs (2018-2020). Average monthly salary (2018), Fig. 1.

METHODS

RS predictors:

- 327 indices: shape, textural, spectral.
- Random Forest pixel- and object-based (91-96% CV accuracy) Google Earth Engine classification
- Share of RS-based classes per subdistrict.

GSV predictors:

share of GSV-based classes per subdistrict from TensorFlow classification model (94% CV accuracy):

Modeling:

Extreme Gradient Boosting (XGBoost) model to predict income.

RESULTS

Both RS and GSV data improve the income prediction quality. The identified best predictors were:

1. Built-up intensity/shape complexity have negative relationship with income.
2. Contrasting/concentrated vegetation has positive relationship with income.
3. Presence of private houses and greenery in GSV photographs – positive predictors.

The overall model perfomance was good (Fig. 2; Table 1) and the spatial distribution of residuals revealed that model underestimated the income in historical and very green areas and overestimated mosty in the open mining and industrial areas (Fig. 3).

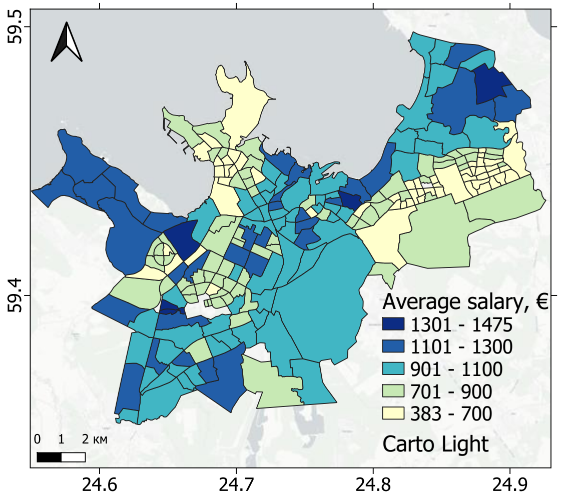

Fig 1. Actual income levels

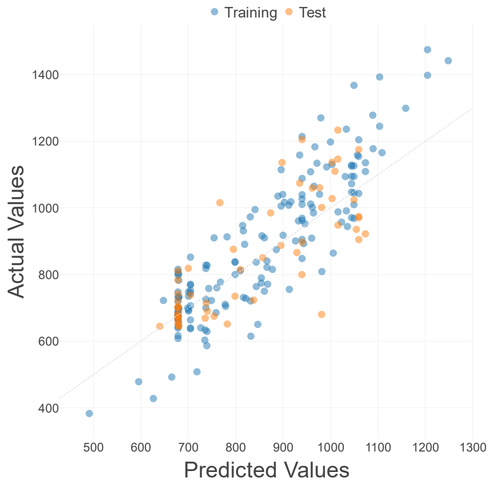

Fig 2. Actual vs. predicted income

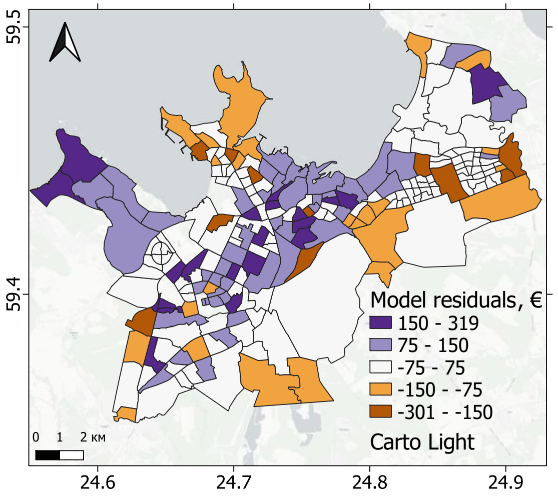

Fig 3. Spatial distribution of residuals

Table 1. The goodness of XGBoost model

Cross-validation	R-Squared	RMSE	Subdistricts
Training	0,74	108,21	165
Test	0,62	108,94	55

CONCLUSIONS

Machine learning algorithms enable application of Sentinel-2 imagery for spatial income modeling.