

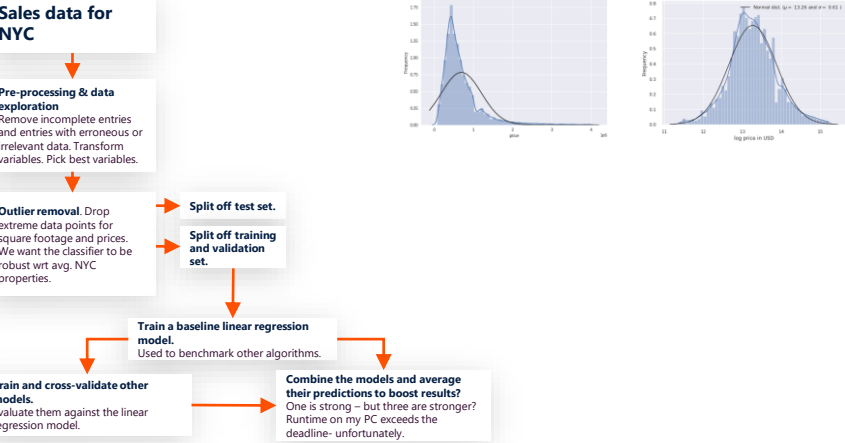
Building a robust regression model on New York City property prices

PRESENTER
Matthaeus Hilpold

Background

Can AI outperform junior appraisers in predicting selling prices of NYC properties? If so, time and money can be saved by developing a reliable automated solution. Many have tried, few have succeeded. Here, I discuss my approach.

Methods



Challenges, Conclusion and Limitations

NYC sale data is very noisy: many properties are sold at 0\$, data pertaining to the date and area of the property is often infeasible or missing. Appropriate data cleaning and the right model nonetheless achieve a **better** and **more robust** average prediction **than** the commonly used **linear regressors**. **Human junior appraisers** are –however- **still better** at mitigating these outliers. They are doubly as accurate, with an average 70000\$ error, according to REIT (sources on the website – scan the QR code).



Algorithms can robustly **outperform** common approaches on **NYC property sale price** predictions.

Model	R ²	MAE (mln. USD)
Linear baseline regressor	0.23	288.05
Linear Kernel Ridge regressor	0.56	189.99
RBF Kernel Ridge regressor	0.66	193.12
Random Forests regressor	0.53	195.24
Combined average boosted regressor	??	??

