

Eksploracja danych

Zastosowanie algorytmów

Gradient Boosted Decision Trees

do prognozowania szeregów czasowych

Sprawozdanie z projektu - 18.01.2019

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

Bartłomiej Bukowski

Ryszard Sikora



Na ile dni do przodu mogę przewidzieć kurs euro

- Jak dużo danych historycznych potrzebuję, żeby przewidzieć przyszłe wartości?
- Na ile dni do przodu mogę przewidywać przyszłe wartości?
- Z jaką dokładnością jestem w stanie przewidzieć kurs za X dni mając Y dni danych historycznych?
- Jak częstotliwość danych wpływa na te przewidywania?



Gradient Boosted Decision Trees

Gradient Boosting to technika uczenia zespołowego, w której klasyfikatory uczone są sekwencyjnie - kolejny klasyfikator uczy się na błędach poprzednich klasyfikatorów

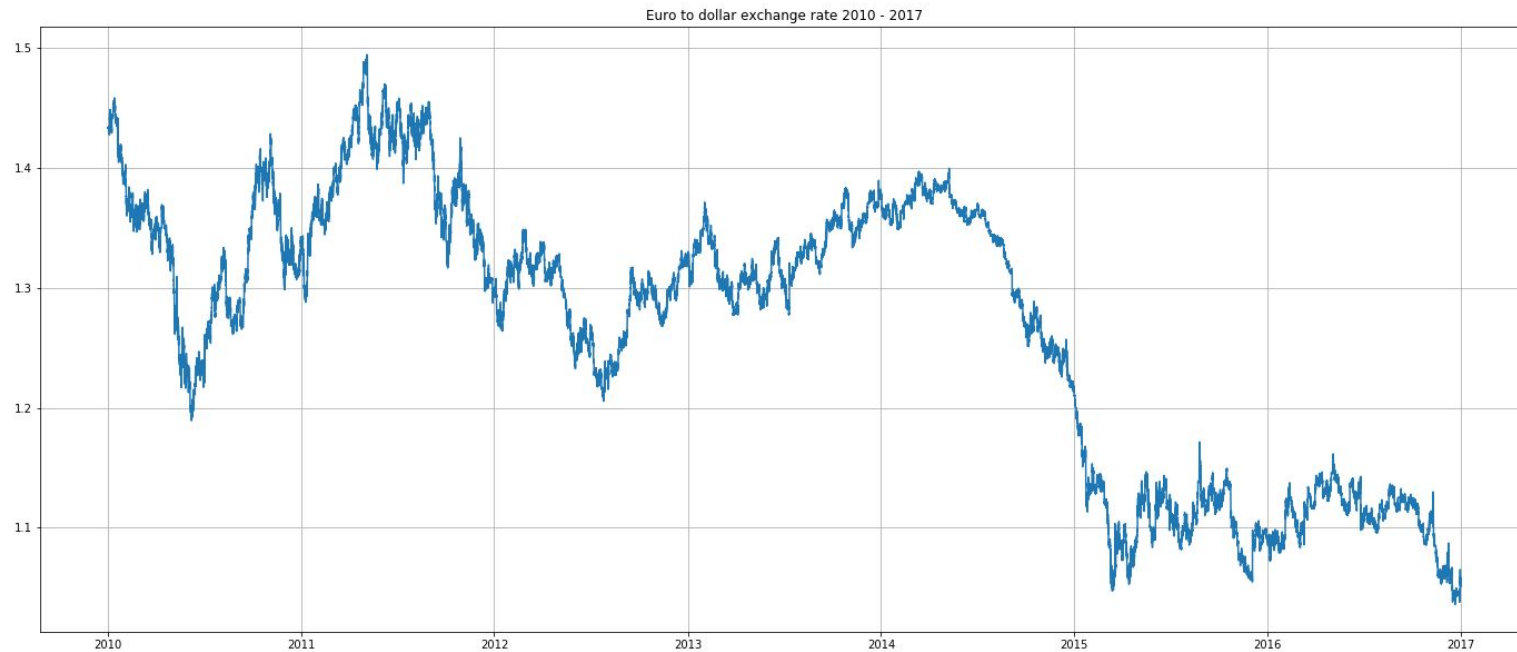
- Zmniejsza stronnicość (bias) i wariancję
- Może przeuczać (overfitting)
- Najczęściej używany z drzewami decyzyjnymi



XGBoost - eXtreme Gradient Boosting



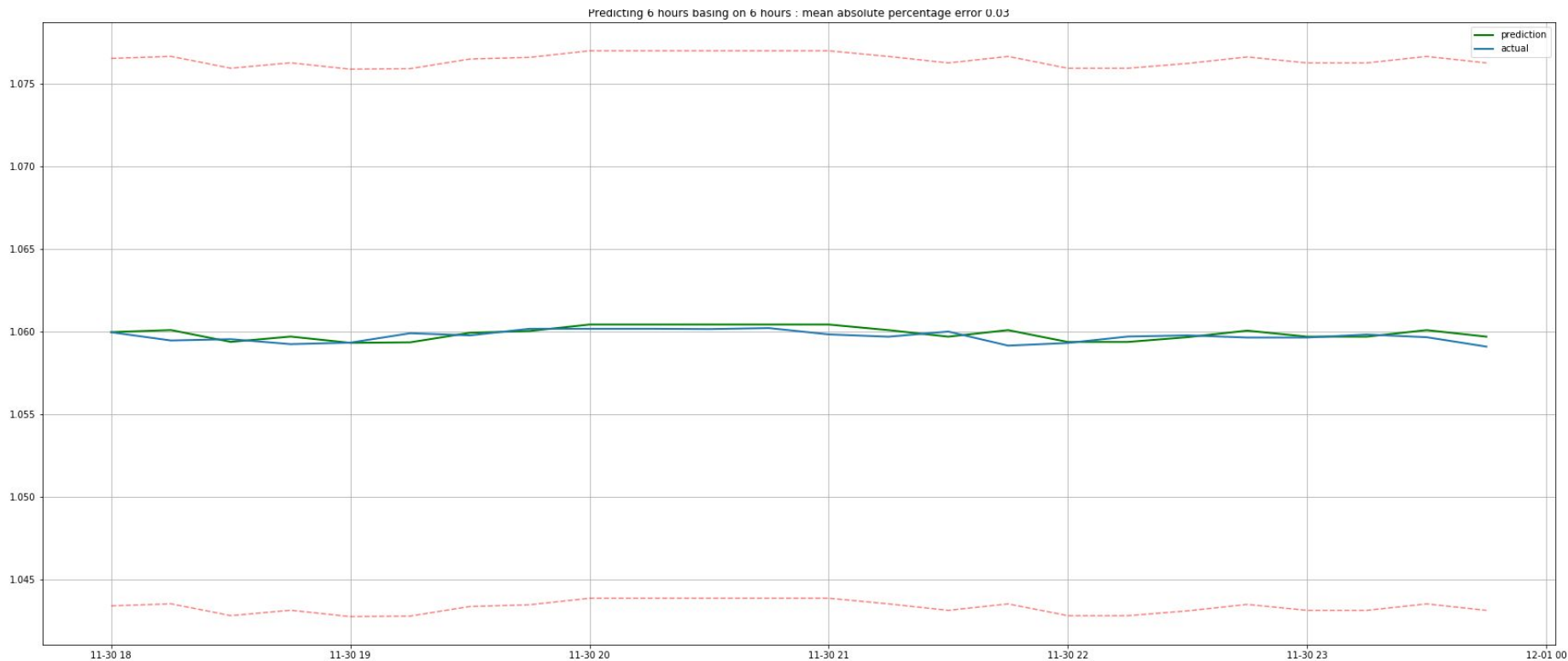
Kurs EURO do USD 2010-2017



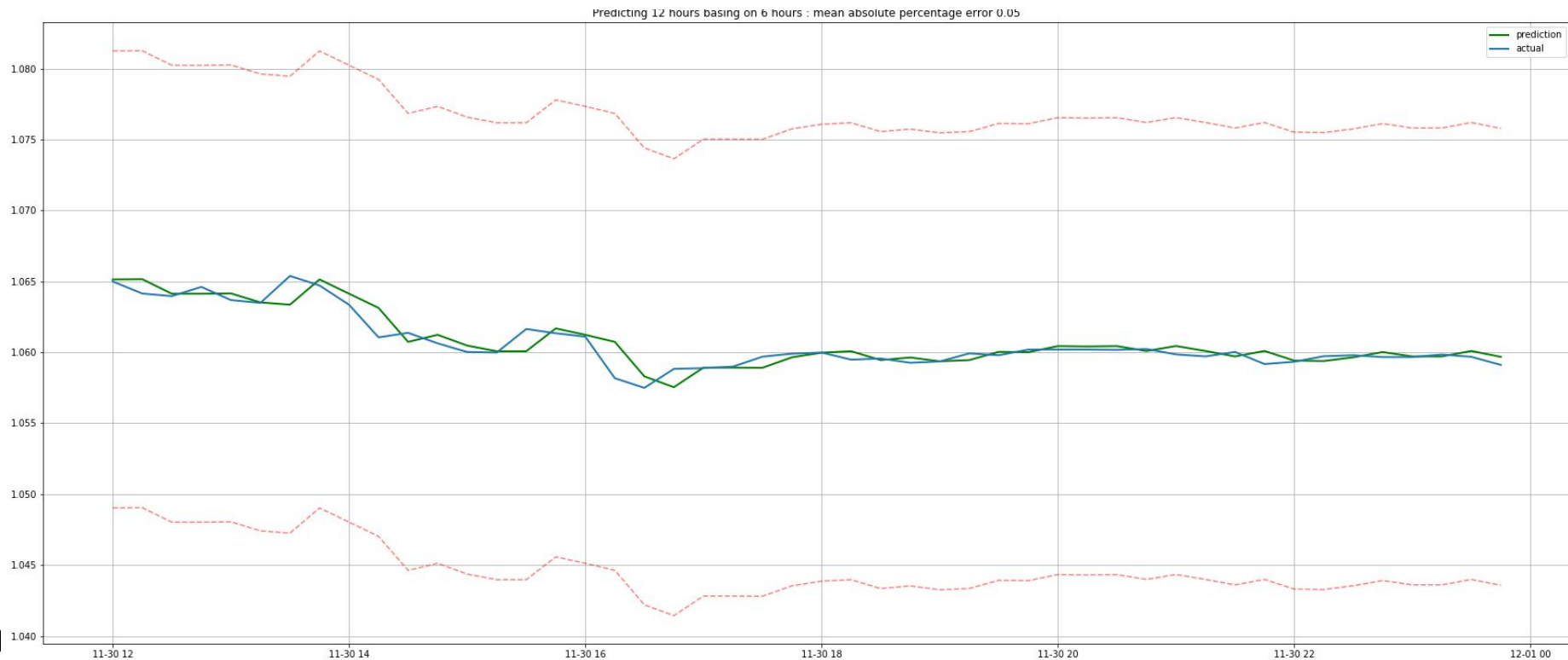
Kurs EURO do USD 2016.10.01 - 2016.11.30



Przewidywanie 6 godzin → 6 godzin



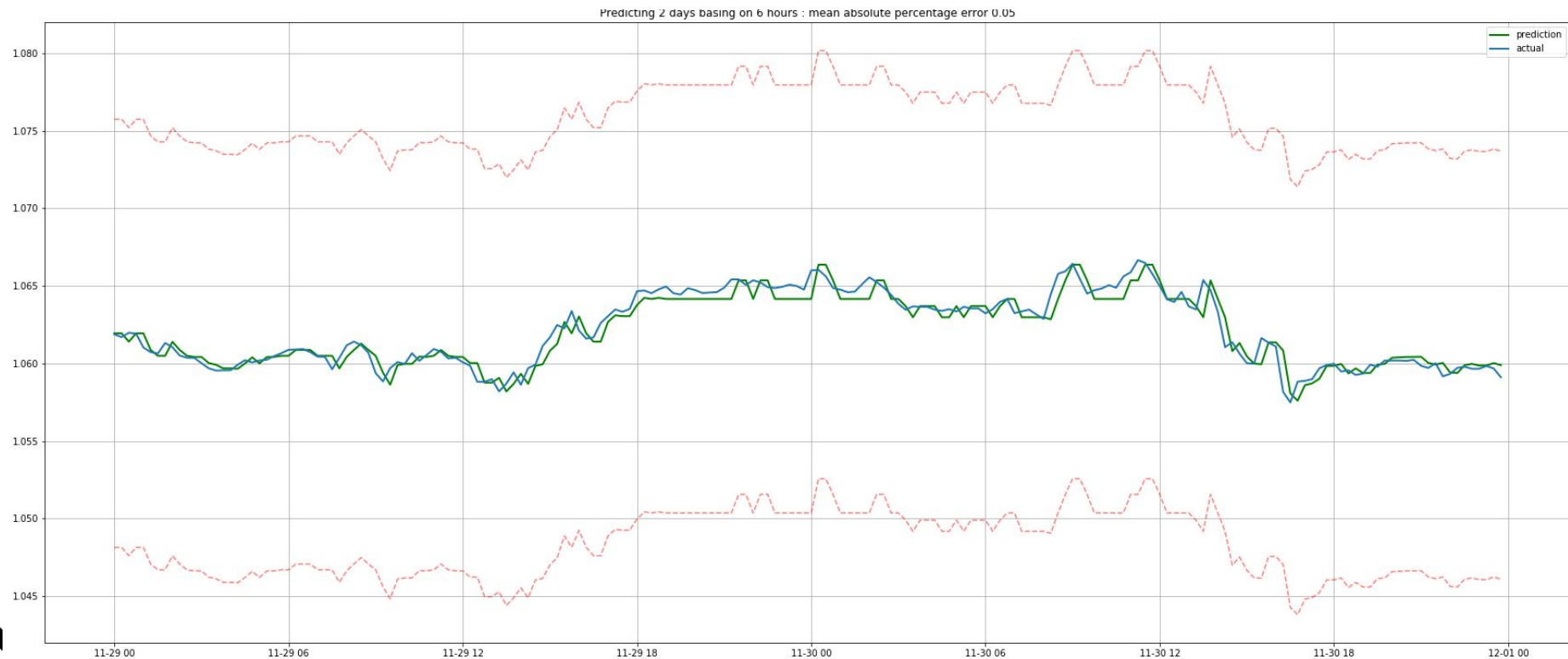
Przewidywanie 6 godzin → 12 godzin



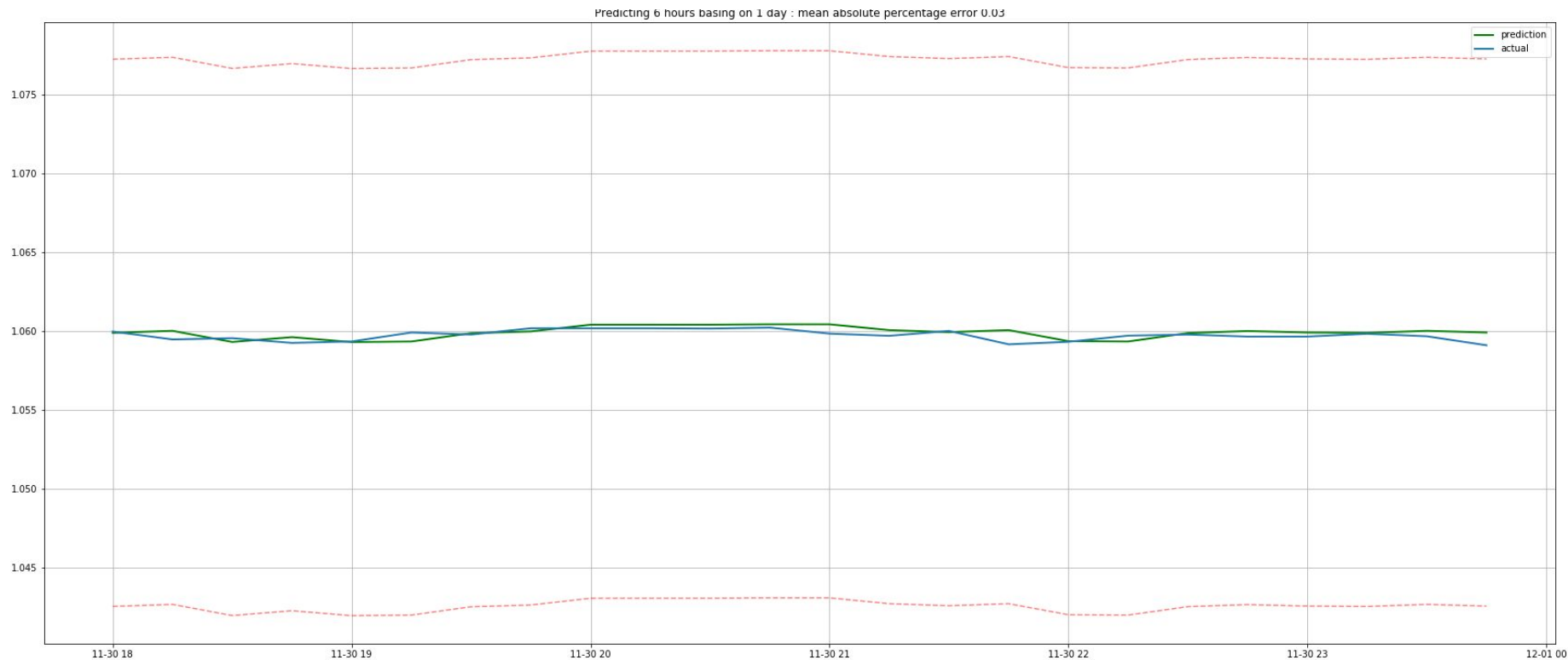
Przewidywanie 6 godzin → 1 dzień



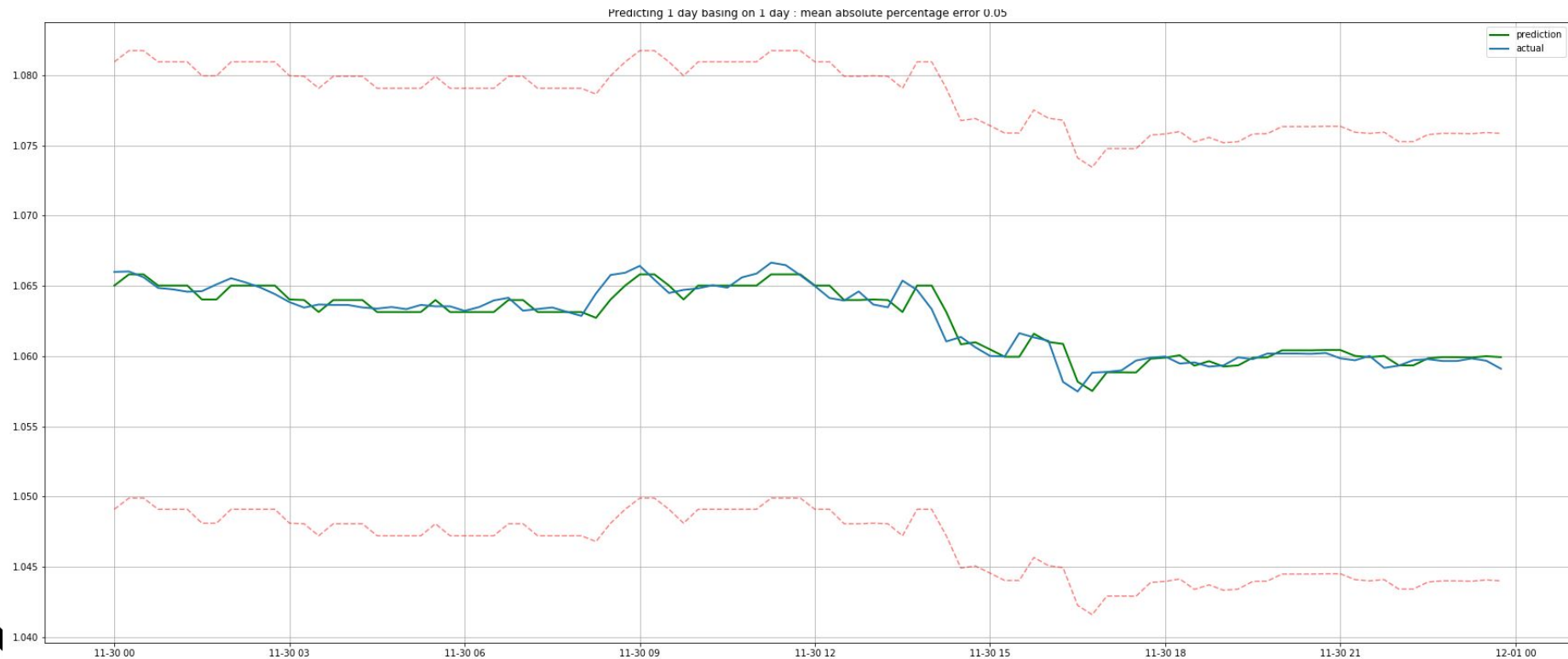
Przewidywanie 6 godzin → 2 dni



Przewidywanie 1 dzień → 6 godzin



Przewidywanie 1 dzień → 1 dzień



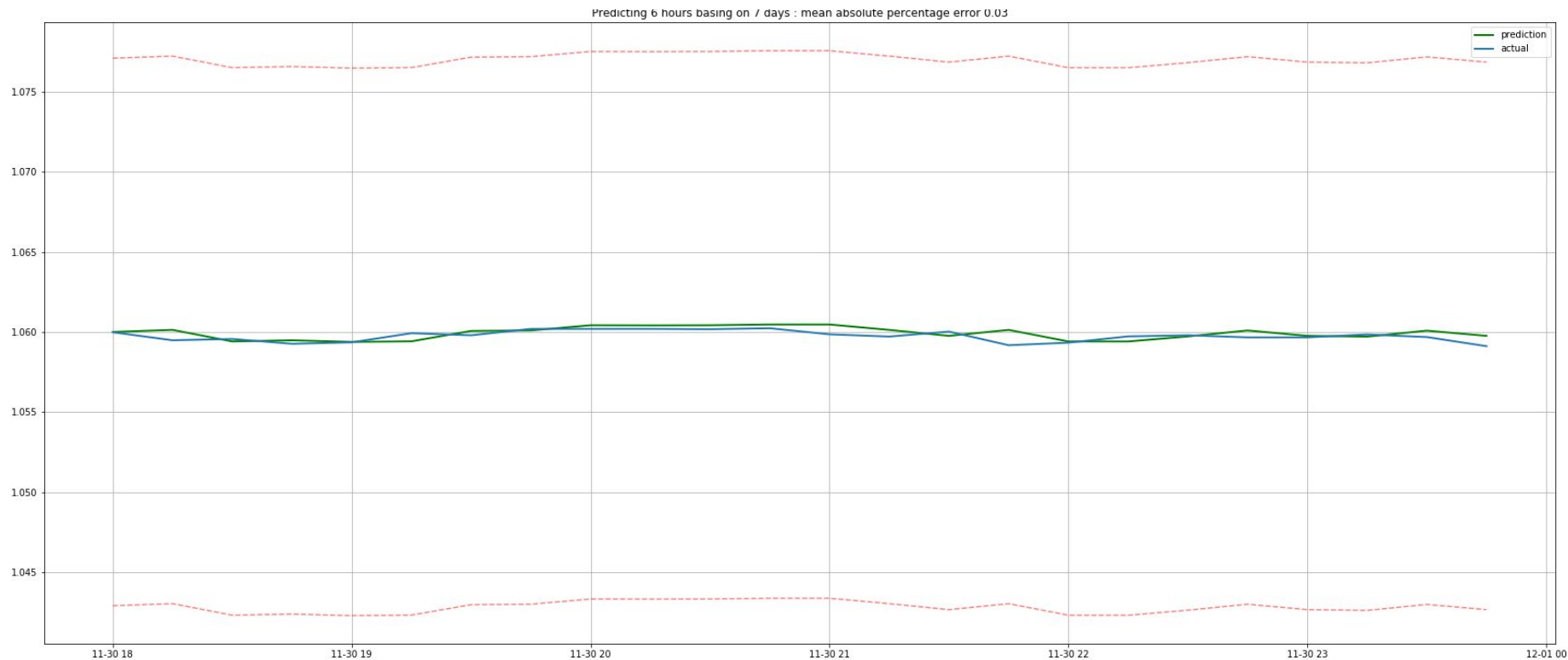
Przewidywanie 1 dzień \rightarrow 7 dni



Przewidywanie 1 dzień → 12 dni



Przewidywanie 7 dni → 6 godzin



Przewidywanie 7 dni → 1 dzień

Predicting 1 day basing on 7 days : mean absolute percentage error 0.05



Przewidywanie 7 dni → 7 dni



Przewidywanie 7 dni → 12 dni

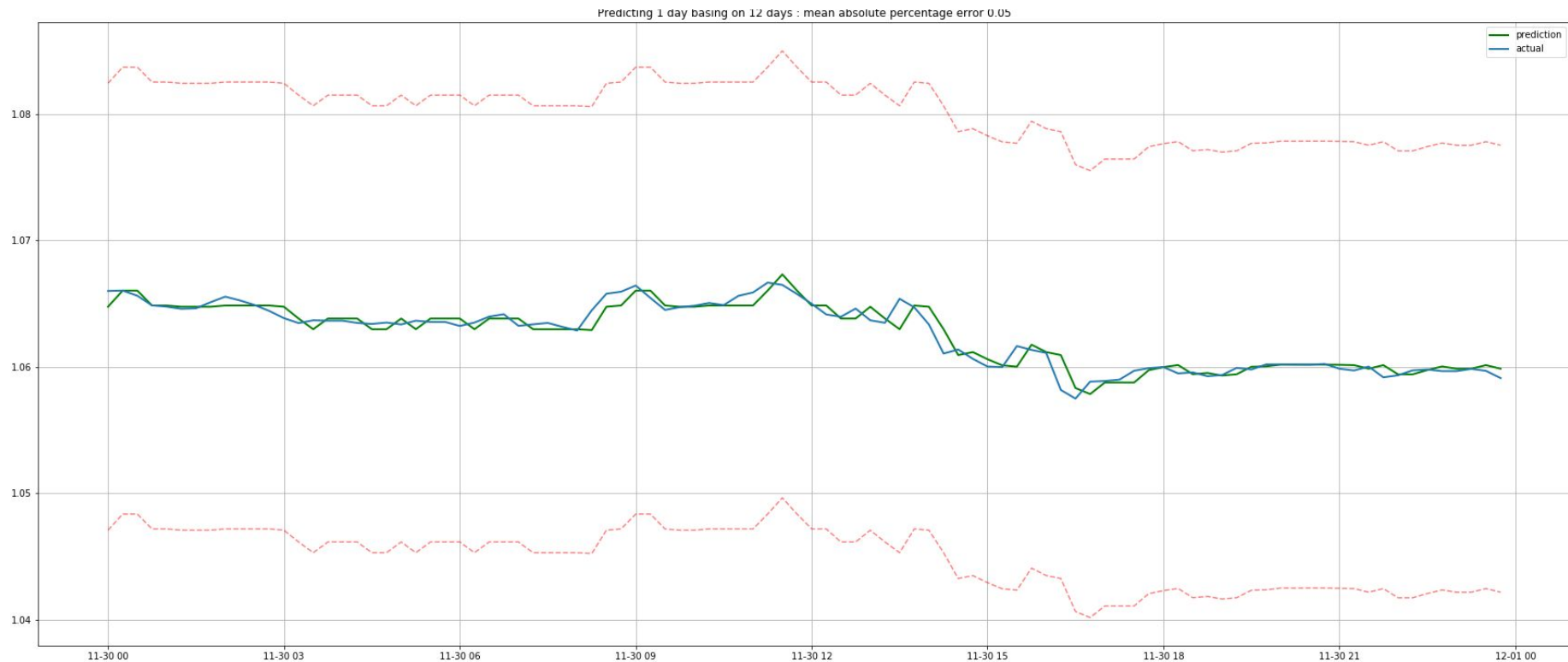
Predicting 12 days basing on 7 days : mean absolute percentage error 0.13



Przewidywanie 12 dni → 6 godzin



Przewidywanie 12 dni → 1 dzień



Przewidywanie 12 dni → 7 dni

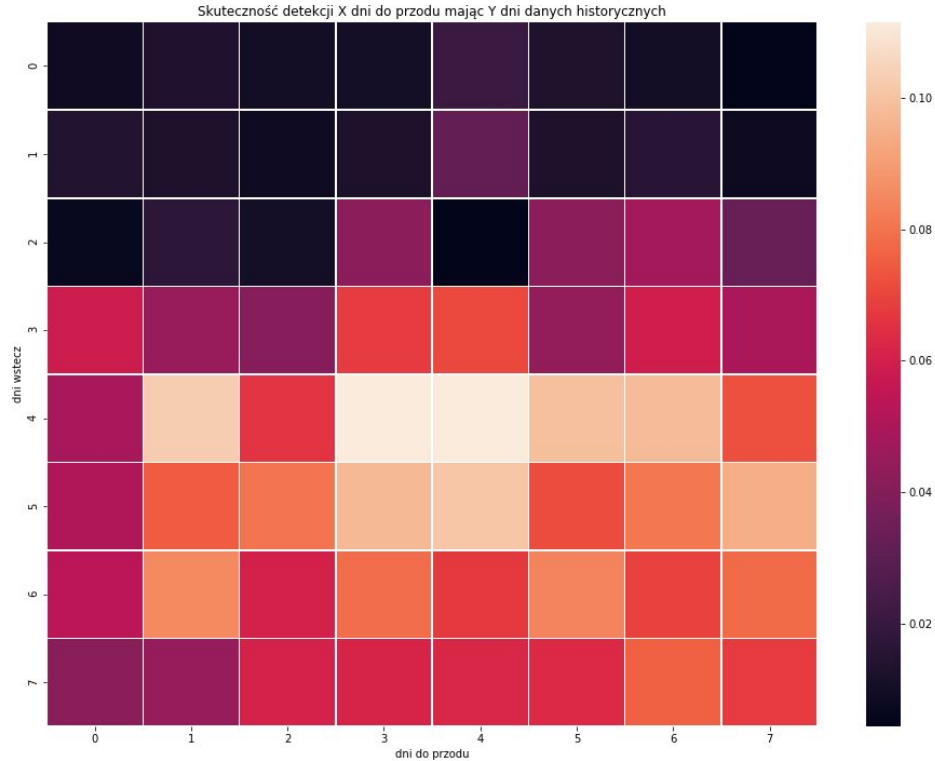


Przewidywanie 12 dni → 12 dni

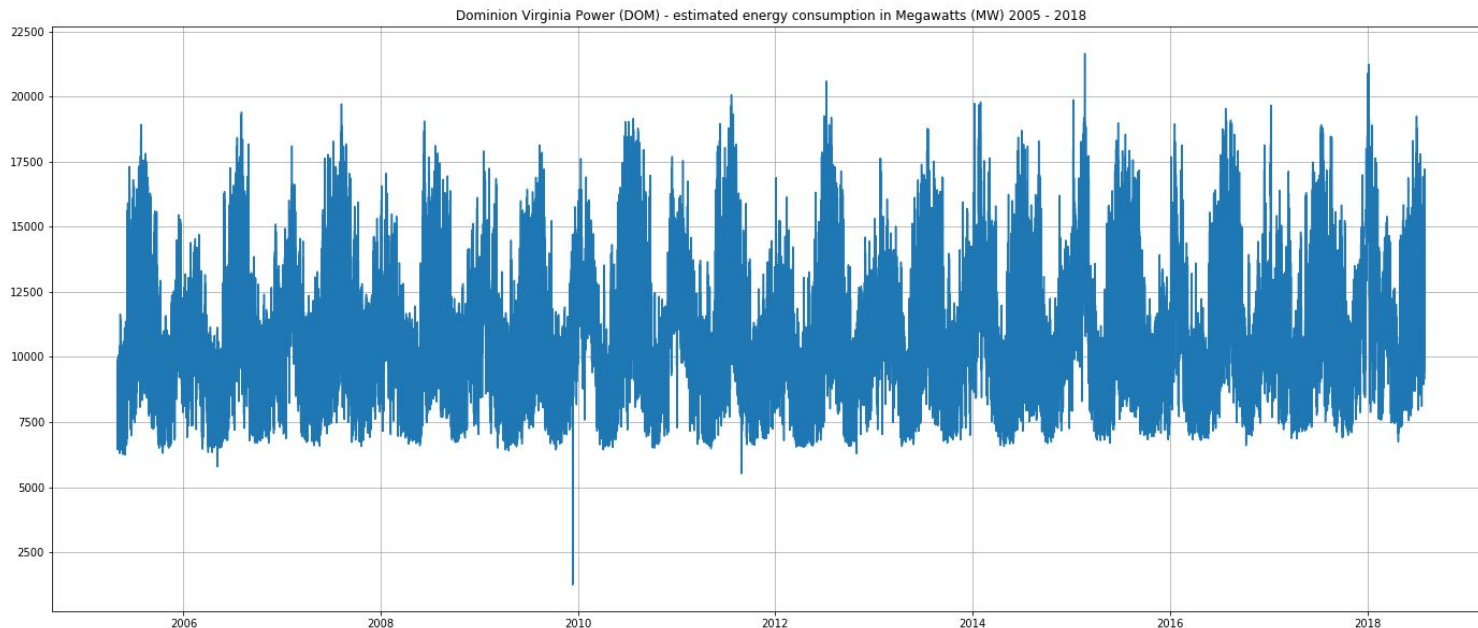


heatmapa

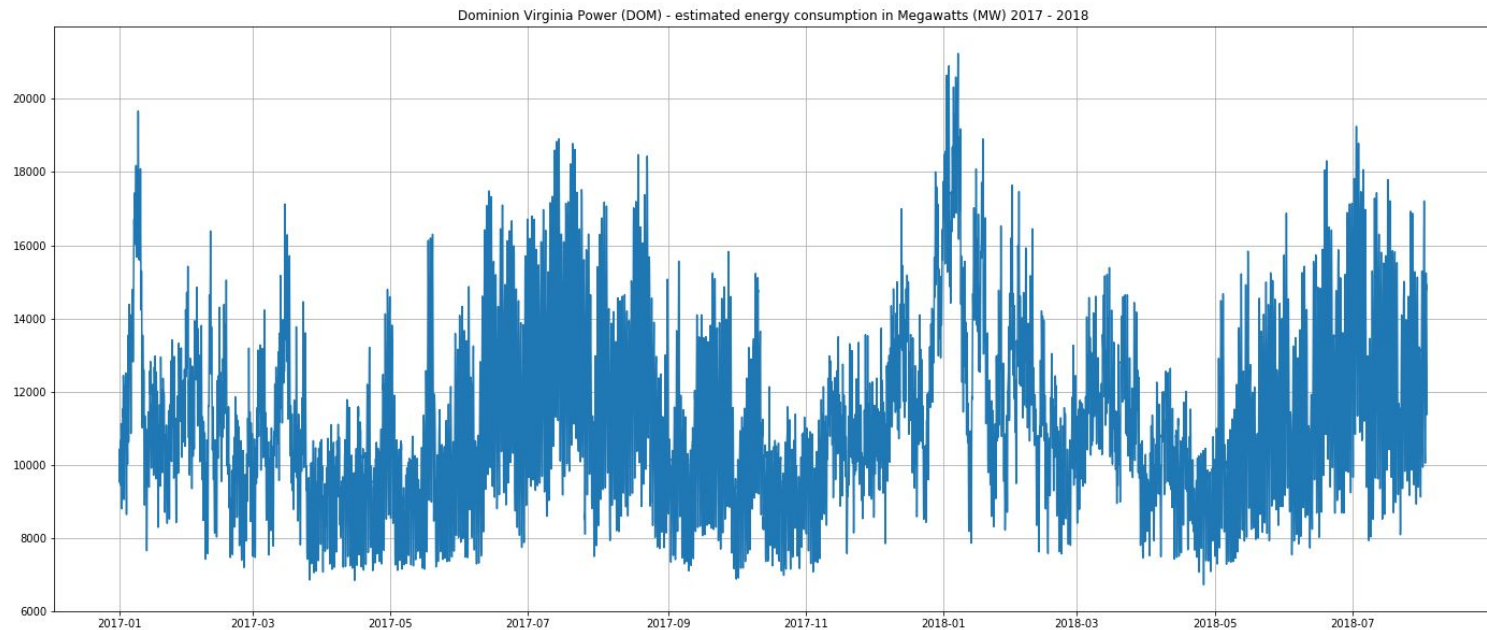
0 → 6 godzin
1 → 12 godzin
2 → 1 dzień
3 → 2 dni
4 → 3 dni
5 → 4 dni
6 → 5 dni
7 → 6 dni



Szacunkowe zapotrzebowanie na energię elektryczną DOM - 2005-2018



DOM - 2017-2018

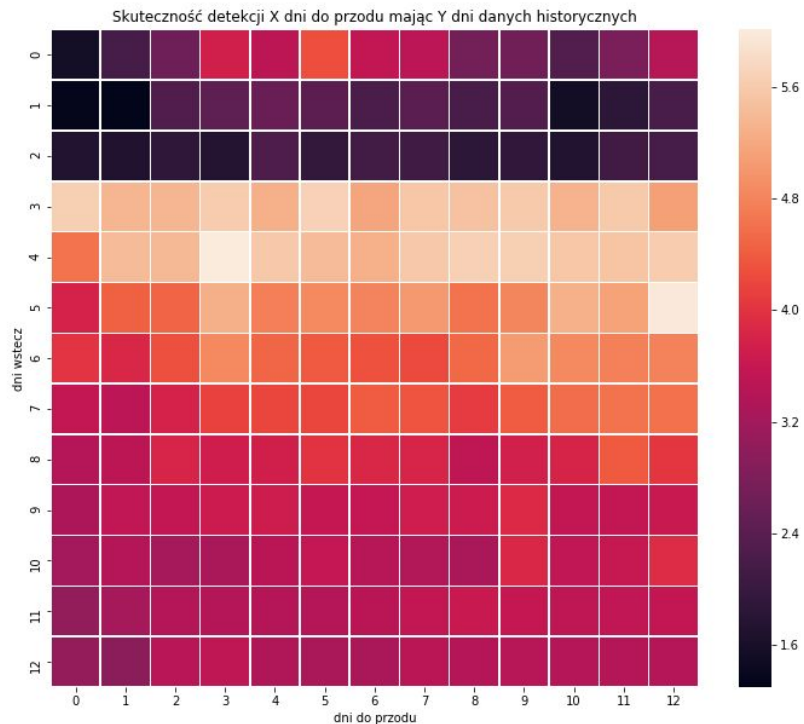


heatmapa

0 → 1 dzień

•
•
•
•
•
•

12 → 13 dni



Bibliografia

1. https://nbviewer.jupyter.org/github/Yorko/mlcourse_open/blob/master/jupyter_english/topic09_time_series/topic9_part1_time_series_python.ipynb
2. <https://www.kaggle.com/thebrownviking20/everything-you-can-do-with-a-time-series>
3. <https://xgboost.readthedocs.io/en/latest/>
4. <https://github.com/rychuhardy/eksploracja-danych-timeseries/tree/master/final>



Zastosowanie algorytmów Gradient Boosted Decision Trees do prognozowania szeregów czasowych

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

Bartłomiej Bukowski
Ryszard Sikora

