

# Programowanie zaawansowane - projekt

Magdalena Czumak  
Przemysław Matuszczak  
Bartosz Sowiński  
Norbert Sulżycki

1 lipca 2020

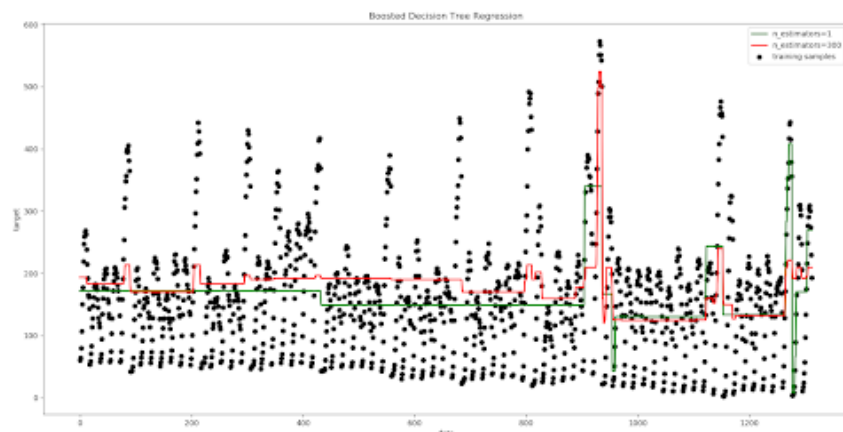
## 1 Problem badawczy

Problemem badawczym postawionym w niniejszej pracy jest prognoza obciążenia parkingu samochodowego umieszczonego w Birmingham, UK (regresja). Dane wykorzystane w projekcie zostały zaczerpnięte ze strony <https://archive.ics.uci.edu/ml/datasets/Parking+Birmingham> (plik: *dataset.csv* zawiera wszystkie dane na temat wszystkich parkingów). Ze wskazanych danych zostały wydzielone dotyczące tylko jednego parkingu *BHMBCCMKT01* jako osobny plik (plik: *dataset – parking.csv*), następnie wydzielone zostały wszystkie dane dotyczące sobót dla tego parkingu (plik: *dataset – soboty.csv*).

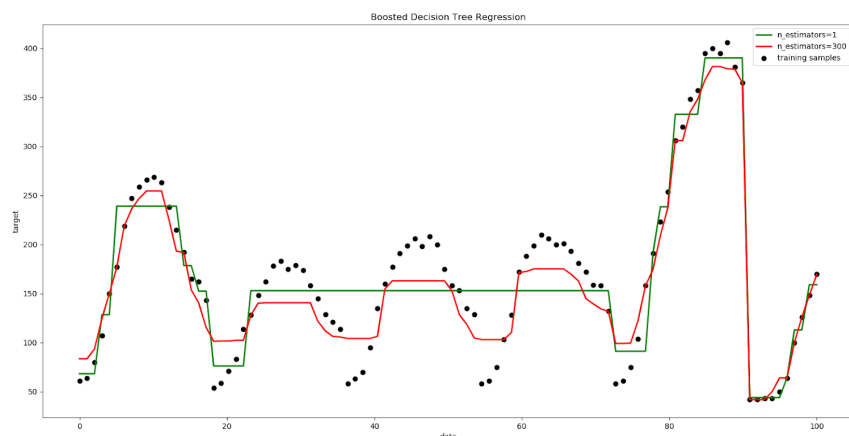
## 2 Model

Modele zostały zbudowane przy wykorzystaniu algorytmu Drzewa Decyzyjnego ([https://scikit-learn.org/stable/auto\\_examples/tree/plot\\_tree\\_regression.html](https://scikit-learn.org/stable/auto_examples/tree/plot_tree_regression.html)). Kod dla obu modeli jest dostępny jako dwa osobne algorytmy (plik: *modelowanie.py* oraz *parking.py*).

Pierwszy wykres sporządzony został na podstawie wszystkich możliwych danych, które były dostępne. Zaobserwować można fakt, że jest on stosunkowo płaski i zdecydowanie mniej zróżnicowany niż posiadane dane (przy obu parametrach *n\_estimators*). Parametr *n\_estimators* świadczy o tym, na ile części dane są dzielone, a następnie algorytm z każdej takiej pojedynczej części oblicza regresję liniową. Oznacza to, że im mniejsza wartość *n\_estimators*, tym na mniej części dane są dzielone.



Kolejnymi zbudowanymi modelami (dla dwóch wartości  $n\_estimators$ ) są widoczne na poniższym wykresie linie zielona i czerwona. Zbudowane one zostały na podstawie tylko i wyłącznie dostępnych danych z sobót.



### 3 Analiza porównawcza

Zbudowanie powyższych modeli przy użyciu Drzewa decyzyjnego nie było trafnym pomysłem. Nie nadaje się ono do rozwiązania postawionego problemu, czyli prognozy przyszłego obciążenia parkingu. Aby możliwa była predykcja, koniecznym byłoby poznanie funkcji będącej modelem danych. W tym przypadku takich modeli jest wiele funkcji (osobno dla każdego przedziału determinowanego przez parametr  $n\_estimator$ ).

Jednakże warto wspomnieć inną cechę wspomnianego parametru. W przypadku mocno zróżnicowanych danych, na które wpływ miały różne szумы, zmniejszając  $n\_estimators$ , zmniejszamy jednocześnie wrażliwość modelu na zakłócenia. Zjawisko to jest zdecydowanie uwidocznione w przypadku pierwszego modelu, gdzie mimo dużego zróżnicowania danych, modele są wypłaszczone.

## 4 Podsumowanie

Podsumowując, zastosowanie Drzewa decyzyjnego nie nadaje się do tworzenia prognozy.

## 5 Podział zadań

- Analiza problemu i przygotowanie danych wejściowych
  - Magdalena Czumak
  - Przemysław Matuszczak
  - Bartosz Sowiński
  - Norbert Sulżycki
- Przygotowanie danych dla modeli
  - Magdalena Czumak
  - Norbert Sulżycki
- Tworzenie kodu Python
  - Magdalena Czumak
  - Przemysław Matuszczak
  - Bartosz Sowiński
  - Norbert Sulżycki
- Analiza danych wynikowych
  - Magdalena Czumak
  - Bartosz Sowiński
- Przygotowanie dokumentu LaTeX
  - Przemysław Matuszczak
  - Norbert Sulżycki