

Politechnika Warszawska

Wydział Mechatroniki

INTELIGENTNE OBLICZENIA – ROWERY

Predykcja natężenia ruchu

Przemysław Płoński, Maja Rusak

Grupa ARIPM-221

Warszawa 2020

1. Opis projektu

Celem projektu jest przewidzenie natężenia ruchu (liczby przemieszczających się rowerów) na podstawie danych historycznych oraz dodatkowych źródeł zewnętrznych.

Projekt opiera się na danych z Warszawskich rowerów miejskich. Dane są pliki JSON, zapisywane co 10

minut, zawierające dane stacji i listy rowerów znajdujących się na każdej stacji. Do projektu użyte były również dane meteorologiczne ze strony: <https://danepubliczne.imgw.pl>. Projekt został wykonany w języku Python w Jupyter Notebook.

2. Opis programu

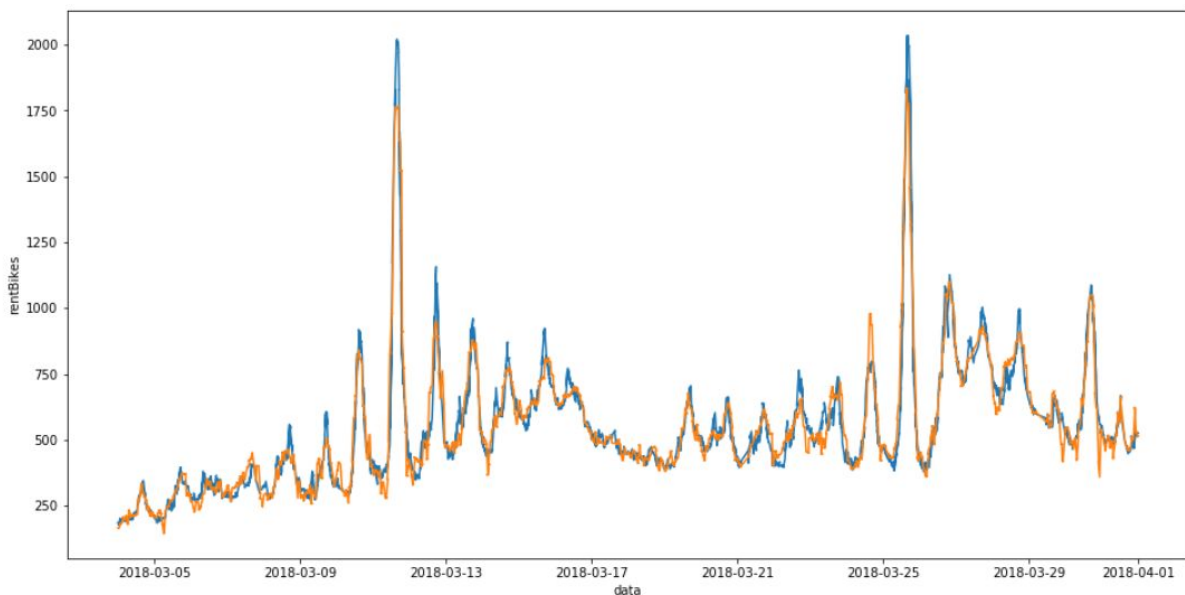
- a) Wczytanie odpowiednich bibliotek, używanych w programie
- b) Definicje funkcji

W tej części zdefiniowaliśmy 6 funkcji, które zostały potem użyte w programie:

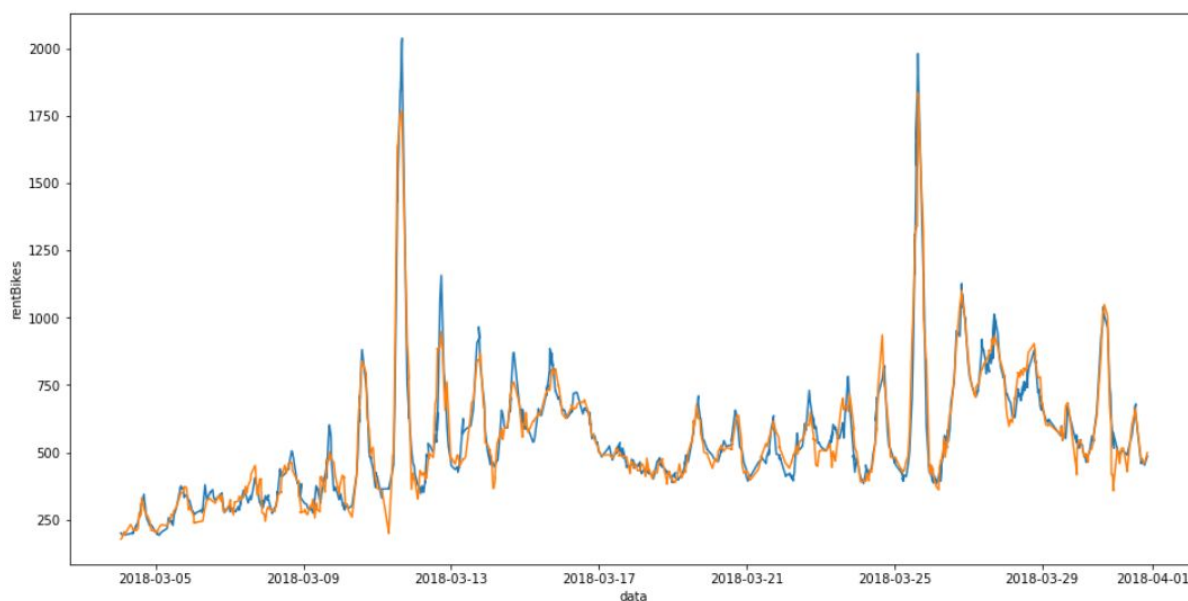
- renameColumns - użyta do zmieniania nazw kolumn przy przetwarzaniu danych meteorologicznych
 - removeNotWarsaw - usunięciu danych meteorologicznych nie będących z Warszawy
 - splitDate - użyte do przetworzenia daty i godziny w danych meteorologicznych
 - addWeekDay - dodanie numeru dnia tygodnia do danych
 - whichDayPart - podział godzin na odpowiednie części dnia (w naszym przypadku podzieliliśmy je od godziny 2 do 22 co dwie godziny)
 - prepareWeatherDF - używane do przetworzenia danych meteorologicznych
- c) Wczytanie danych meteorologicznych - stwierdziliśmy, że będą nam potrzebne 3 dane meteorologiczne, które mają związek z wypożyczaniem rowerów: temperatura, opady deszczu i wiatr
 - d) Wczytanie danych rowerów miejskich - do naszego projektu wybraliśmy następujące dane:
 - data - zawiera rok, miesiąc, dzień, godzinę i minutę czytania danych
 - rok
 - miesiąc
 - dzień
 - godzina
 - minuta
 - wypożyczone rowery - które pobraliśmy z przetworzenia danych dostępnych w pliku JSON odejmując liczbę wszystkich rowerów, do dostępnych
 - e) Przetworzenie danych meteorologicznych oraz danych z Warszawskich rowerów miejskich
 - Na początku do zawierających danych dodaliśmy dwie kolumny, dzień tygodnia oraz porę dnia
 - Następnie dodaliśmy kolumnę "prevRent", w której zapisywana była średnia ilość wypożyczonych rowerów tydzień wcześniej. Próbowaliśmy jeszcze wziąć zamiast tej zmiennej średnią z dnia poprzedniego, z całego dnia lub z danej pory dnia, ale tak uzyskaliśmy najlepsze wyniki
 - Kolejno dodaliśmy również kolumny z temperaturą, opadami oraz wiatrem. Ze względu na inny czas pomiaru danych meteorologicznych i danych z Warszawskich

rowerów miejskich, musieliśmy dodawać uśrednioną wartość temperatur, opadów oraz wiatru dla interesujących nas godzin

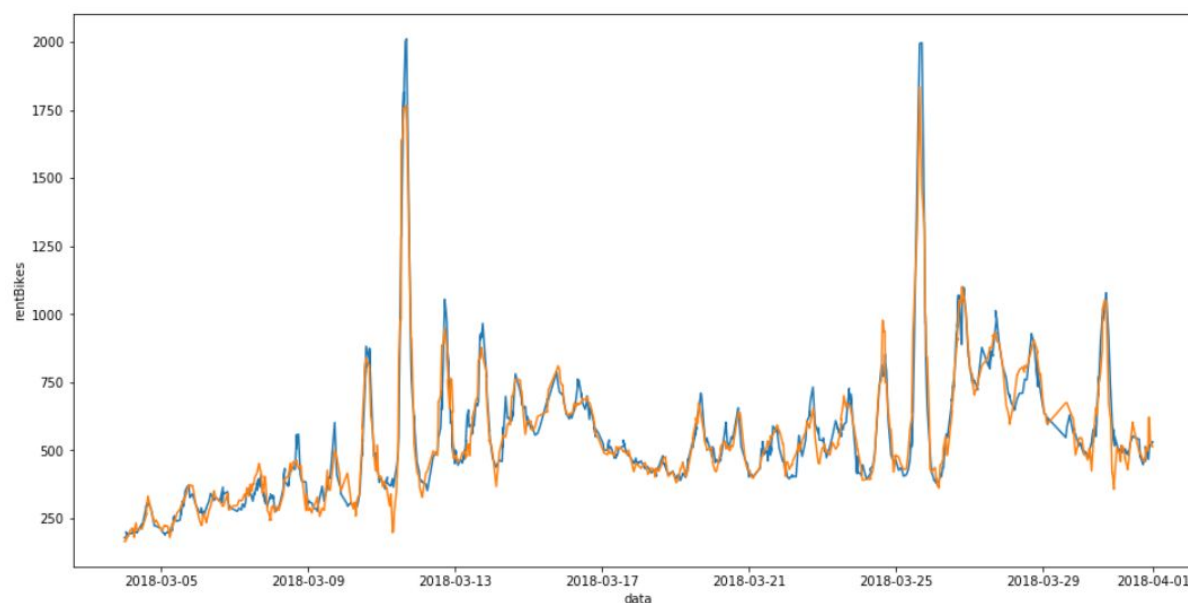
- f) Wykresy punktowe pokazujące zależności danych meteorologicznych od wynajętych rowerów - aby sprawdzić czy zależności między wynajmowanymi rowerami i danymi meteorologicznymi mają sens, wykonaliśmy odpowiednie wykresy punktowe. Jedyne niepoprawności jakie zauważyliśmy to zależność wypożyczeń rowerów od opadów deszczu, gdzie przy opadach około 2mm zostało wypożyczonych ponad 600 rowerów. Uznaliśmy ten wynik za mało prawdopodobny i w późniejszych krokach usunęliśmy dane, które na to wskazywały.
- g) Usunięcie brakujących wartości oraz odstających wartości, zauważonych podczas analizy wykresu wynajęcia rowerów od deszczu.
- h) Tworzenie modelu oraz podział zbioru na treningowy, walidacyjny i testowy - zbiór danych został podzielony na zbiór treningowy, walidacyjny i testowy w proporcjach 60/20/20. Jako dane wejściowe zostały użyte dane dotyczące dnia tygodnia, pory dnia, temperatury, opadów deszczu, godziny i średniej ilości wypożyczonych rowerów tydzień wcześniej. Do utworzenia modelu zastosowaliśmy również regresję wielomianową używając funkcji: `PolynomialFeatures`.
- i) Predykcja:
 - R^2 dla zbioru treningowego: 0.932
 - R^2 dla zbioru walidacyjnego: 0.925
 - R^2 dla zbioru testowego: 0.923
- j) Wykresy zależności wynajętych rowerów od przewidywania wynajęcia rowerów od daty



Wykres 1: Wykres dla zbioru treningowego



Wykres 2: Wykres dla zbioru walidacyjnego



Wykres 3: Wykres dla zbioru testowego

Wnioski:

Korzystając z danych meteorologicznych i historycznych informacji o wypożyczonych rowerach miejskich w Warszawie, udało nam się otrzymać model, który przewiduje natężenie ruchu na satysfakcjonującym poziomie. Przewidywanie jest wykonane z dokładnością do 2 godzin, co wydaje się rozsądnym przedziałem czasowym.