

Analiza stężenia pyłu PM2.5 w Pekinie, lata 2010-2014

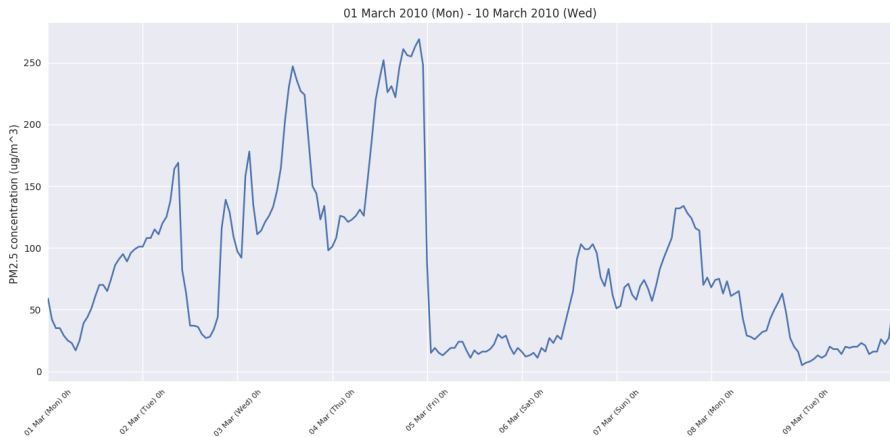
Jakub Pokrywka

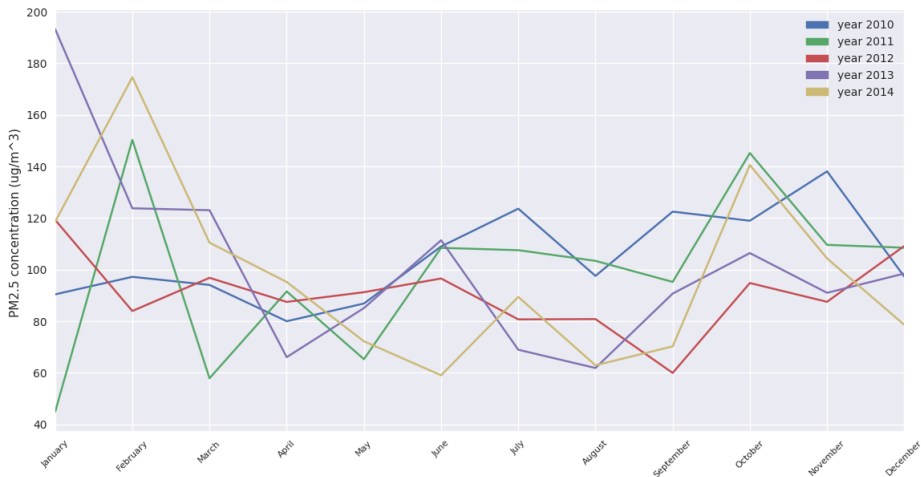
jakubpokrywka@gmail.com

20 lutego 2018

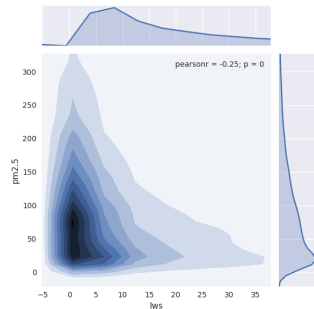
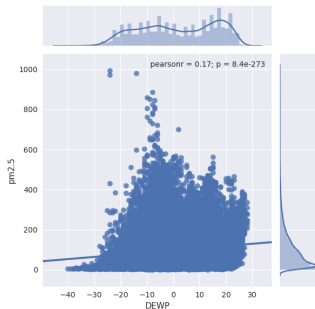
Dane przedstawiają codzienny zapis stężenia cząsteczki PM2.5 oraz danych atmosferycznych w Pekinie, w latach 2010-2014. Po analizie zbioru jako szeregu czasowego nasuwają się następujące wnioski:

- Brak wyraźnej okresowości wysokości stężenia w zależności od godziny,
- występuje pewna okresowość wysokości stężenia w zależności od miesiąca. Najwyższe stężenia zaobserwować można w okresie październik-luty. Prawdopodobnie jest to wynikiem sezonu grzewczego. Podejrzenie niska wartość na wykresie dla stycznia 2011 wynika z braku danych w tym miesiącu (90% braków).
- nie ma trendu zmiany średniego stężenia na przestrzeni lat.





Widoczny jest natomiast duży wpływ warunków atmosferycznych na wartość stężenia PM2.5. Występuje korelacja dodatnia między PM2.5, a punktem rosy (DEWP- wsp. korelacji 0.17), co wiąże się ze zmianą wilgotności. Wpływ prędkości wiatru (lws) jest istotny- nim większy, tym mniejsze stężenie. (wsp. korelacji -0.25). Kierunek wiatru ma pozorny wpływ na wartość stężenia (wynika to z nierównomiernego natężenia wiatru z różnych stron).



W celu przewidzenia stężenia PM2.5 stworzono model regresji liniowej oraz model neuronowy. Dla obu modeli najlepszy okazał się zestaw cech: miesiąc, godzina, punkt rosy, temperatura, ciśnienie, prędkość wiatru, czas opadu śniegu, czasu opadu deszczu, kierunek wiatru (zakodowywany jako wektor binarny).

- Naiwny model (zawsze średnia) daje MSE równe 8257.
- Model regresji liniowej daje MSE równe 6051.
- Model neuronowy (5 warstw ukrytych) daje MSE 4791 z odchyleniem standardowym 267 (walidacja krzyżowa z podziałem 0.1).

Ze względu na wielkość zbioru, model neuronowy daje zdecydowanie lepszy wynik niż prosta regresja liniowa, zatem w tym przypadku jest rekomendowany.