Influenza Outbreak Forecast in New York City with Weather Data
Zhengxing Li, Ziying Tao, Guangjun Xue, Yichen Zhang

Predicting
- Influenza is one of the most significant diseases in humans.
- Better epidemic predictions would set up more appropriate public health prevention and intervention strategies in temperate cities.
- Epidemics occur mainly during the season months with abnormal changing of temperature, precipitation, UV radiation, and wind speed.
- Random Forest, Linear Regression, Gradient Boosting, K-Nearest Neighbors Algorithm.
- Based on the future weather data to predict the Influenza.

Data
Datasource:
- Daily weather data from NOAA[2011-2018]
- Weekly influenza infections data from Centers for Disease Control and Prevention[2011-2018]

Data preprocessing:
- Data Cleaning
- Aggregate daily weather data into weekly data
- Merge weather data and influenza data

Features
The raw data includes features: weekly highest temperature, weekly lowest temperature, weekly average temperature, weekly sea level pressure, weekly max wind speed, weekly minimum wind speed, weekly average wind speed, weekly average uv time, weekly average precipitation.

Models
4. KNN:
\[ P(y = j|x) = \frac{1}{K} \sum_{i=1}^{K} \delta(y^{(i)} = j) \]
Given a point \( x \), we will choose \( K \) nearest points from dataset and have a vote among them to decide the prediction.

5. DNN:
Our network consists of one input layer and two hidden layer and the output layer. We use relu as activation function. And epoch=20000.

Discussion
- Overall, all the models predict the outbreak of influenza successfully.
- Gradient Boosting has the best result on testset.

Future
We will use universal weather data, like finding different meteorological stations' data, and using the longer span of year, or do more literature research to find a better way, to do test. Besides, we will also improve our model and try several new models to see the consequence which is more reasonable.

Reference