Predictive Analysis of the Air Quality of Shunyi District in Beijing, China based on PM2.5 Levels

Sean Torres and Anusia Edward

Background, Purpose, and Objective

Background: In the past two decades, there has been a rapid increase in the number of deaths around the world as a result of air pollution, especially in China (Zheng et al., 2021). China's air quality is ranked as one of the lowest among countries worldwide based on its Environmental Sustainability Index value. One of the most affected locations in China is the Shunyi District of Beijing (Wang et al., 2018).

Purpose: The purpose of this study is to address the growing issue of poor air quality and its negative effects on health outcomes by analyzing the Shunyi District of Beijing, China.

Objective: This study's objective, therefore, is to use secondary data to develop a predictive model aimed at predicting air quality in terms of $PM_{2.5}$ levels based on units of time (month, year, hour), corresponding pollutants such as PM_{10}, SO_2 (ug/m^3), NO_2 (ug/m^3), CO (ug/m3), O_3 (ug/m^3), temperature (^oC), pressure (hPa), dew point temperature (^oC), and wind speed (m/s).

Explored Solutions

- Testing out different models to help predict $PM_{2.5}$ levels in Shunyi District of Beijing, China:
  - MODEL 1: Ordinary Least Squares Regression
    - Easy to interpret
  - MODEL 2: Partial Least Squares Regression
    - Considers associations between variables that help predict the outcome
  - MODEL 3: Random Forest
    - Best to use when dealing with large data
  - MODEL 4: Elastic Net
    - Helps reduce error when it comes to predicting the outcome

- Metrics used to evaluate models:
  - R^2 values + root mean square error (RMSE) values
    - High R2 value and low RMSE value can be used as metrics to identify a good model

Results

- Final model selected:
  - Elastic Net model
- Top 5 variables to help predict $PM_{2.5}$ levels in the atmosphere:
  - Large Particulate Matter (PM_{10})
  - Carbon monoxide (CO)
  - Nitrogen dioxide (NO_2)
  - Sulfur dioxide (SO_2)
  - Wind speed (WSPM)

Future Implications

- Explore other districts of China to verify results obtained
- Further research on the effects of wind speed on air pollution
- Utilize a new model to help predict $PM_{2.5}$ levels
  - MODEL: Autoregressive integrated moving average (ARIMA) model
    - This model takes into account data taken over a specified timeframe. In our study, the specified timeframe is from 2013-2017.

References

Li YR, Han TT, Wang JX, Quan WJ, He D, Jiao RG, Wu J, Guo H, Ma ZQ. (2021) Application of ARIMA Model for Mid- and Long-term Forecasting of Ozone Concentration. 42(7):3118-3126. doi: 10.13227/j.hjkx.202011237. PMID: 34212637.

Wang, L., Zhang, F., Pilot, E., Yu, J., Nie, C., Holdaway, J., Yang, L., Li, Y., Wang, W., Vardoulakis, S., & Krafft, T. (2018). Taking Action on Air Pollution Control in the Beijing-Tianjin-Hebei (BTH) Region: Progress, Challenges and Opportunities. International journal of environmental research and public health, 15(2), 306. https://doi.org/10.3390/ijerph15020306

Zheng, S., Schlink, U., Ho, K. F., Singh, R. P., & Pozzer, A. (2021). Spatial Distribution of PM2.5-Related Premature Mortality in China. GeoHealth, 5(12), e2021GH000532. https://doi.org/10.1029/2021GH000532