

Bi-Weekly Report:

Meeting 2:

- To-Do
 - Search for literature
 - get to know your data set
 - Write import script for data set and prepare data filtering (cities, pollutants, ...)
- Tools
 - workspace needed. [In-process]
 - HTW Cloud to share the PDF.
- Analytical goals
 - 1st goal is to predict 1 hour with input data of 6 hours.
 - 2nd could be weeks.
- Preparation of the project plan.
 - May
 - Data and Project understanding.
 - literature research
 - June
 - Selection of models as per the data.
 - July
 - Selection of autoML level. (mostly 3)
 - Aug
 - Finalizing the project.
 - along with paper work
 - Sept
 - Finishing the paper work.

Meeting 3: (06-06-2023)

PROF

Work done:

- Understanding of the data. ([refer dataUnderstanding.ipynb](#))
 - knowing the data set.
 - main db.
 - cities db.
 - chemicals db.
 - merging the dbs to as per convince.
- Graphs ([refer graphs](#)).
 - Plotting bar graph the data for each chemical in a single
 - 6 graphs = 6 chemicals as **y** and cities as **x**
 - Correlation Matrix of all columns.
- Discovering facts about data. ([refer dataUnderstanding.ipynb >> Facts about data](#))
 - The number of cities with Stickstoffmonoxid [nitric oxide] data available is: 61
 - The number of cities with Stickstoffdioxid [nitrogen dioxide] data available is: 61

- The number of cities with Schwefeldioxid [sulfur dioxide] data available is: 7
- The number of cities with PM2.5 data available is: 46
- The number of cities with PM10 data available is: 59
- The number of cities with Ozon data available is: 29
- The data is available from 2019-12-31 to 2022-10-01

Meeting 4: (20-06-2023)

- Understanding of the data. ([refer dataUnderstanding.ipynb](#))
 - Plotting the PM10 daily average for all city.
 - Collected data where the wert value is greater than 35.(Table below)

Chemical	City	Wert count(for 2 years 10 months)
PM10	Bottrop-Welheim	34
PM10	Duisburg-Walsum	37
PM10	Essen-Vogelheim	67
PM10	Lünen-Niederaden	16
PM10	Netphen Rothaargebirge	05
PM10	Ratingen-Tiefenbroich	23
PM10	Simmerath (Eifel)	11
PM10	Solingen-Wald	10
PM10	Aachen-Burtscheid	14

Literature review:

- Hybrid Artificial Neural Network Models for Effective Prediction and Mitigation of Urban Roadside NO2 Pollution
 - Most roadside pollutant variables, e.g., oxides of nitrogen, were found to be significant in predicting NO2. [Page no. 3524]
 - The network was trained using the Levenberg-Marquardt backpropagation algorithm, which adjusts the values of weights between interconnecting neurons based on some error function of the model and target values to minimise the overall error. [Page no. 3527]
 - The algorithm is based on a non-linear optimization method called the gradient descent. [Page no. 3527]
 - The three feature selection methods:
 1. Stepwise regression is a linear search strategy that is based on two known predictor selection techniques, namely, forward selection and backward elimination methods.
 2. PCA operates by transforming the original input space via singular value decomposition into a set of orthogonal vectors, called principal components (PCs).
 3. CART is a machine learning technique capable of building regression and categorisation models based on several input data. [Page no. 3528]
 - The Performance Indicators:

1. the root mean squared error (RMSE)
 2. The coefficient of determination (r^2)
 3. The coefficient of determination (r^2) [Page no. 3528]
- The second most common predictors selected were temperature, hour of the day, the particulate matters, and CO. This finding seems to account for the close association between the said predictors and NO₂ in the atmosphere
 - AutoML to Date and Beyond: Challenges and Opportunities
 - The authors selected the papers published from January 2001 to February 2019.[Page no. 4]
 - Only Papers dealing with forecasting of outdoor air pollutants were considered. .[Page no. 4]
 - There is a growing number of published articles since 2001 that cite the use of air pollution forecasting tools based on ANNs, with almost 50% of the identified papers published since 2015 alone.[Page no. 7]
 - The results reveal that airborne particulate matter with an aerodynamic diameter smaller than 10 μm (PM₁₀) and 2.5 μm (PM_{2.5}) are the most examined variables among the papers identified.[Page no. 7]
 - hourly step used by most of the Papers reviewed which matches our data
 - The data covering a period of a year or more has been highly recommended. This is to ensure that seasonal factors which have been identified to strongly influence air pollution levels are taken into account. .[Page no. 8]
 - Meteorological variables especially wind speed, wind direction, relative humidity and atmospheric turbulence have been found to have a massive influence on the dispersion and concentration of several air pollutants. [Page no. 10]
 - The paper has given a road map for using ANN to predict weather forecast.[Page no. 13 onwards]