

# Final Project Milestone: Daily Water Demand Forecasting

Project Category: General Machine Learning

Haocheng Fan, hfan11

October 27, 2023

## Motivation

My hometown, Shaoxing, is often called the "Eastern Venice" of China. But nowadays, we face big challenges in managing our water. With more people and changing weather patterns, it's getting harder to make sure we have enough water for everyone. As we use more technology and our population grows, our water needs become more complex. It's important for us to find solutions to these problems. We're looking into how machine learning can help us predict daily water needs so we can use our water wisely.

## Preliminary Experiments

### 1. Data Collection:

Due to the difficulties of data collection in China, we decide to find the similar region in population and climate parameter. Finally We sourced the dataset from IEEE dataport. It is the subtropical humid region in Pakistan, with the similar average temperature and latitude as Shaoxing. This includes the daily data for water consumption, date, average temperature, average dew point, humidity, average precipitation. This enables us to do the daily water consumption forecasting.

### 2. Data Pre-processing:

- Construct the Correlation matrix to see the reasonable variables for the models:

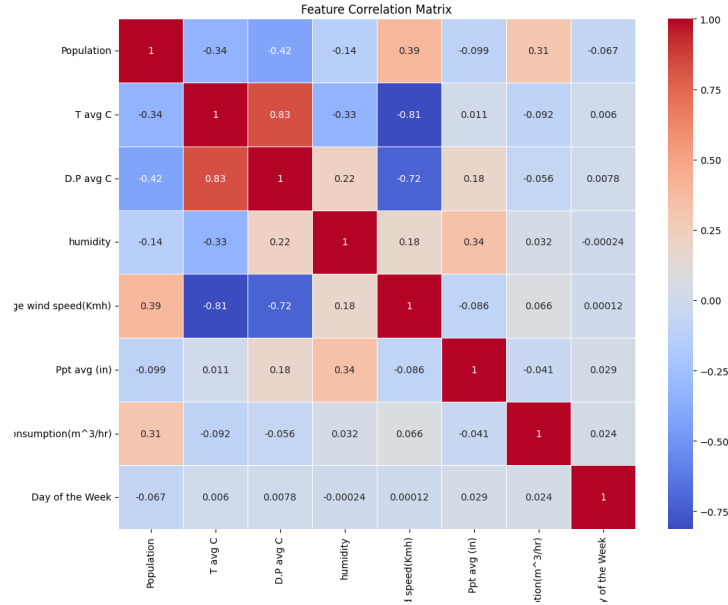


Figure 1: Correlation Matrix

- Select reasonable variables for the models: Figure 1 shows the correlation heatmap. It can be referred that the population have the maximum impact as they have the maximum positive correlation with water consumption. Humidity, wind speed and day of the week also have the positive impact although the correlation is small. Therefore, the major parameter is population and the others are used to improve accuracy.
- Analysis the water consumption related to the population. To clean the outlier data, we use the IQR method. Then we get the cleaned data to process to the training Model.

### 3. Model Preparation and Training:

Set up the three machine learning model algorithms: Linear Regression (Baseline), Random Forest, and Support Vector Regression Model. We use 80 % as train set and 20% as validation set.

### 4. Model Evaluation:

- Linear Regression Model:

<b>Metrics</b>	<b>LR</b>
$R^2$	[0.482]
MSE	[1010268]
MAE	[815.85]

Table 1: Metrics for Linear Regression (LR)

- Random Forest Model:

<b>Metrics</b>	<b>RF</b>
$R^2$	[0.357]
MSE	[1228537]
MAE	[901.55]

Table 2: Metrics for Random Forest (RF)

- Support Vector Regression Model:

<b>Metrics</b>	<b>SVR</b>
$R^2$	[0.379]
MSE	[1169332]
MAE	[881.64]

Table 3: Metrics for Support Vector Regression Model (SVRM)

## 5. Error Analysis and Future Work:

Our results aren't as good as we hoped. When we look at our graphs, small changes in population can cause big changes in water use. This makes it hard to train our model. We didn't have enough time to clean our data as much as we wanted.

In future iterations of this project, we aim to investigate data decomposition methods, such as PCA and SVD, to better discern underlying patterns and potentially reduce the influence of noise in our dataset. If you have any better ideas, please let me know.

## References

Q, Shuang & Zhao, RT. (2021). Water Demand Prediction Using Machine Learning Methods: A Case Study of the Beijing–Tianjin–Hebei Region in China. *Water*, 13(3), 310. <https://doi.org/10.3390/w13030310>.

Kavya, M., Mathew, Aneesh, Shekar, Padala Raja Sarwesh, P. (2023). Short term water demand forecast modelling using artificial intelligence for smart water management. *Sustainable Cities and Society*, 95, 104610. <https://doi.org/10.1016/j.scs.2023.104610>.