Carnegie Mellon University
Mechanical Engineering

Rhine Water Level Prediction using Historic Data

Team 5: Akanksh Shetty, Lakshmi Maddirala, Kanishk Mair, Xinzhe Qi, Yuxin Yao

covestro

Motivation and Objective

Uncertainty over the water level may causes

1. Raw material shortages
2. Difficulty in distribution of finished goods
3. Covestro is not able to use all the water available to maximum extent.

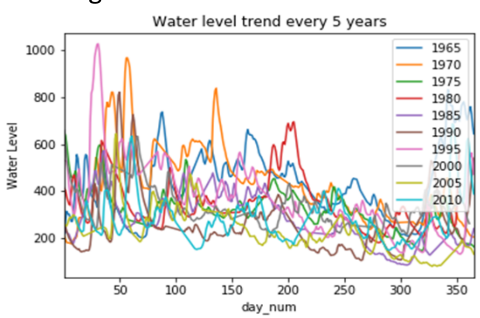
Improved predictions will improve Business Value by:

1. Planned production of finished goods.
2. Avoiding inventory buildup and customer dissatisfaction.
3. Efficient use and less dependency on external sources for water.

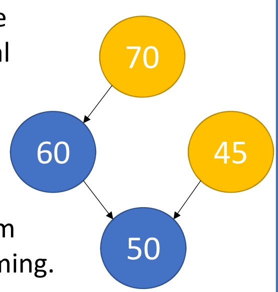
Dataset

The dataset comprised of two files:

1. Weather recorded at nearby weather stations
2. Discharge and water level from the stations

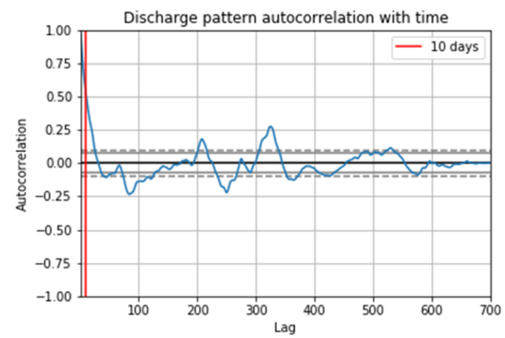


- The models were made for the 2 stations rather than a general model for any station.
- The nodes at Duesseldorf (50) and Cologne (60) get discharge from other nodes.
- Accordingly, the discharge from them were accounted as Incoming.



Time Series Analysis of data

The features and the labels in the dataset are related in time-based manner. So the features for the days to be predicted were modeled with ARIMA.



Based on auto-correlation graphs, 10 days of previous data was used to model the feature set. They are:

River Data

- Discharge
- Incoming water

Weather Data

- Rain level
- Snow level

Regression models

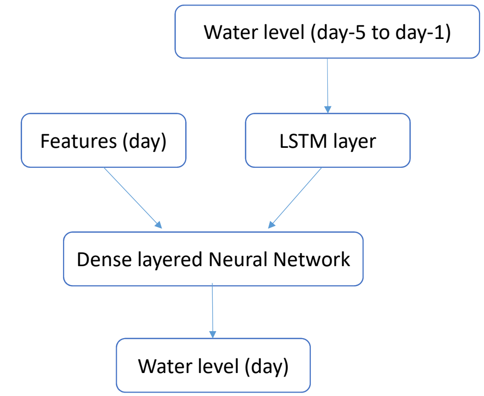
The data was tested on the following regression models. Water level for previous days and features for the prediction day was used. Hyperparameters were chosen based on the number of features.

Regression Model	Days in history		
	5	10	20
Linear	0.420187	0.43	0.444608
Gradient Boosting Regressor	0.32934	0.329385	0.32992
Random Forest Regressor	0.387973	0.389259	0.361833
Bayes Ridge Regression	0.420197	0.435457	0.444754
Passive Aggressive Regressor	0.152189	0.348067	-0.0038
TheilSenRegressor	0.424078	0.409245	0.407229

The best algorithm was Bayes Ridge Regressor, which was able to reach maximum R^2 of 0.445.

Machine Learning model

The following model architecture with 5 days of water level history was modeled.



For the above model with hyperparameters optimized, an R^2 of 0.4219 was achieved.

Conclusions

Improving prediction Accuracy

- From all the results, it can be observed that the model accuracy saturates to around 0.4. Thus there is a need for improved feature data.
- The classification features could not be modeled using ARIMA and were omitted. Accurate modeling of these can boost the predictions.
- The modeled trained only used features for the output day. A more complex model containing previous day data of features might improve performance.

Direct use

- There are multiple weather predicting tools which can be used to get feature data for prediction days.
- Incorporating this instead of the filled values will improve the prediction accuracy.