

## **Project 1:**

Data type: time series

Number of features: 624 (the first column includes time)

Number of Samples: 393 (the first row includes feature IDs)

Targets for prediction: 11118, 47179, 37997, 77781

Dear all,

I am excited to present an engaging project that involves predictive modeling in the maritime shipping industry using a comprehensive dataset spanning from 1991 to the present day. The primary objective is to forecast future values just after the last recorded month. In the preprocessing phase, each row signifies information from a specific month (lag 0), and the target value for the subsequent month is considered (lead 1). The lag parameter reflects the delay in input information, while the lead indicates the ahead prediction period.

Your task is to train a regression model on the training set and evaluate its performance on a designated test set. Due to the temporal nature of time series data, we aim to assess the model's accuracy over the last three years available in the dataset. To achieve this, consider training the model on all samples from the beginning up to  $n-36$ , and then test it on one sample ahead ( $n$  being the total number of samples excluding the last one without a label). Subsequently, shift down by one step, adding one sample to the train set and testing on one sample ahead.

For reporting and analysis, create an Excel file comprising the last 36 months' data. The first column should represent time, the second column the true target values, the third column the predicted values, and the fourth column the calculated accuracy using the formula:

$$\text{Accuracy} = 100 * (1 - \text{abs}((\text{actual} - \text{prediction}) / \text{actual}))$$

Ultimately, calculate the average accuracy across the 36 predictions.

This project offers a hands-on opportunity to apply regression modeling techniques to real-world data, emphasizing the challenges and nuances of forecasting in the dynamic maritime shipping industry. I encourage you to explore different regression models, fine-tune parameters, and critically evaluate the model's performance.

Best of luck, and I look forward to your insightful analyses.

Warm regards,

Mehdi.