Regression hackathon

Dataset:

This hackathon will analyse and model actual data about a wind generation power station. In this station, we have 10 different sensors located at different (unknown) positions around the wind turbines. These sensors measure the temperature, wind speed and direction at each location. The information contained in the datasets follows this form:

1. Attributes (31 variables):
   - **TLXH80** Temperature measured at sensor X
   - **WSLXH80** Wind speed at sensor X
   - **WDLXH80** Wind direction measured at sensor X

2. Output:
   - **WG** Wind power generation of the station

Description:

The file **TRAIN_DATA.csv** provides the records for approximately 2 months of data in an hourly fashion. The objective here is to provide the best model to predict the WG variable.

The file **TEST_INPUT.csv** provides the input for the test set (no WG variable).

You must provide forecasts for the TEST_INPUT.csv values. The best forecast will be considered the one with the **lowest RMSE** value in the test set.

The **forecast** must be **uploaded** to the task at **Moodle**. Use the following code to create the output file with the predictions and upload the .csv document to Moodle. Modify the team name so as to include the names of your team.

```
prediction.to_csv("TeamName.csv", sep=" ", header=False, index=False)
```

Prepare a one-page document with a comprehensive analysis of the procedure, analysis and models employed and upload it to Moodle as well. Include in your submission the code that you have employed to model the data and produce the final predictions. **Include both of these two documents with the team members' names! Failure in doing so may lead to 0 grades for those excluded.**