

Kaggle Datathon Machine Learning Competition

Description:

The manufacturing industry heavily relies on energy consumption, and predicting energy load types can help reduce costs and improve efficiency. In this competition, we were given a dataset containing energy usage information from DAEWOO Steel Co. Ltd in Gwangyang, South Korea. The dataset included several types of coils, steel plates, and iron plates, and the energy consumption information was stored in a cloud-based system. The task was to predict the variable Load_Type, which can help manufacturing companies better understand their energy consumption patterns and adjust their operations accordingly.

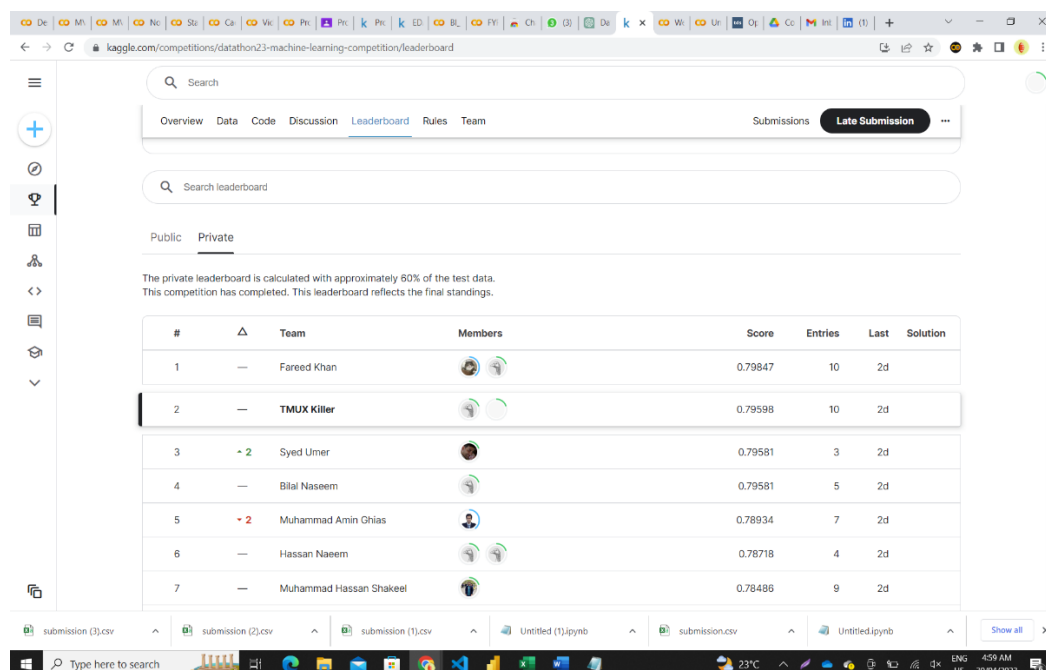
To approach this problem, we started by analyzing the dataset's variables and performing missing value analysis. We discovered a pattern in the missing date values and imputed the date accordingly. We then imputed the other variables using the date as they were dependent on it. Next, we used LazyPredict, a Python library for automatically trying multiple machine learning models and their hyperparameters, to identify the models that could provide the best accuracy for our predictions. We tried the AdaBoost and Random Forest classifiers and found that Random Forest performed better. After hypertuning the model's parameters, we achieved 0.796 accuracy on the Kaggle leaderboard and secured the second position. This competition provided us with a valuable opportunity to apply our machine learning skills to real-world data and demonstrate our ability to develop effective solutions for industrial energy consumption forecasting.








Kaggle Competition Link:

<https://www.kaggle.com/competitions/datathon23-machine-learning-competition/leaderboard>

Kaggle Code Link:

<https://www.kaggle.com/code/alisurani/datathon-notebook>



#	Team	Members	Score	Entries	Last	Solution
1	Fareed Khan		0.79847	10	2d	
2	TMUX Killer		0.79598	10	2d	
3	Syed Umer		0.79581	3	2d	
4	Blal Naseem		0.79581	5	2d	
5	Muhammad Amin Ghias		0.78934	7	2d	
6	Hassan Naeem		0.78718	4	2d	
7	Muhammad Hassan Shakeel		0.78486	9	2d	