Hybrid Model: Deep learning GRU neural network and K-nearest neighbors for Wind Power Forecasting

Team : datateam-UCM **Code:** https://github.com/ManuelAngel99/KDD_CUP_2022

Fernando Sebastián Huerta*
ferseb39@gmail.com
BCAM-Basque Center for Applied Mathematics Bilbao
Madrid, Spain

Manuel Angel Suarez*
manuesua@ucm.es
Universidad Complutense de Madrid, UCM
Madrid, Spain

Daniel Velez Serrano*
danielvelezserrano@mat.ucm.es
Universidad Complutense de Madrid, UCM
Madrid, Spain

Alejandro Carrrasco Sánchez*
acarrasco2008@gmail.com
Unspecified affiliation
Madrid, Spain

Eugenio Neira Bustamante*
eug.neira@gmail.com
Unspecified affiliation
Madrid, Spain

Introduction

Proposed problem

- Spatial Dynamic Wind Power Forecasting data-set

Variable Name	Specification
TurbID	Wind Turbine ID
Day	Date of the record
Tmstamp	Created time of the record
Wspd (m/s)	The wind speed recorded by the anemometer
Wdir (°)	Angle between wind direction and turbine nacelle
Etmp (℃)	Temperature of the surrounding environment
Itmp (℃)	Temperature inside the turbine nacelle
Ndir (°)	Nacelle direction, i.e., the yaw angle of the nacelle
Pab1/2/3 (°)	Pitch angle of blade 1/2/3
Prtv (kW)	Reactive power
Patv (kW)	**Active power (target variable)**

- With more than eight months of 10-minute data for all of the above variables, a **2-day prediction** of the target variable Patv was to be calculated for **288 10-minute intervals**

Introduction

Data exploration

- Data from a **real** wind turbine park with large amount of “**invalid” values (up to 30%**)

- **Large Patv - Wind speed correlation** -
  - Wind predictions are crucial, but competition rules impeded utilizing **external data sources**
  - Large Patv variability due to stochastic distribution of wind speed

Figure 1: Correlation matrix for the different dataset variables

Introduction

Data exploration

- Data from a **real** wind turbine park with large amount of “**invalid” values (up to 30%**)

- **Large Patv - Wind speed correlation** -
    - Wind predictions are crucial, but competition rules impeded utilizing **external data sources**
    - Large Patv variability due to stochastic distribution of wind speed

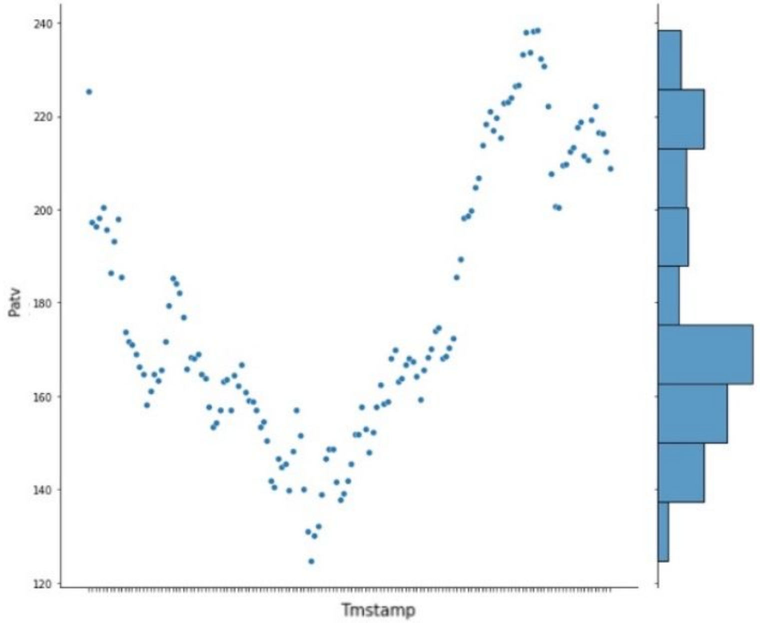


Figure 2: Median values Patv values grouped by Timestep

Introduction

Main competition challenges

- Fitting models to predict **short and long predictions**: model performance varies greatly.
- Evaluated on a **per turbine basis**, but creating per-turbine models (LGB models) discarded due to limitations on execution time and **model size** introduced by the competition organizers.
- A large number of NaN and invalid data points ➜ challenged when training models with low bias, over-fit easily.
- **Offline model validation transfer to online execution was poor,** due to :
    - Large dependence on stochastic wind speed conditions
    - Online evaluation of a few samples
    - A large number of invalid data points were removed when evaluating

Detailed method

Recurrent Neural Network model

- Model inspired by "**PaddlePaddle** WPF Baseline GRU", **transferred to Pytorch**.
- **Single** model for 134 turbines:
  - More robust ➜ better score
  - Simpler ➜ faster execution time
- **Higher dropout** rate and larger batch size

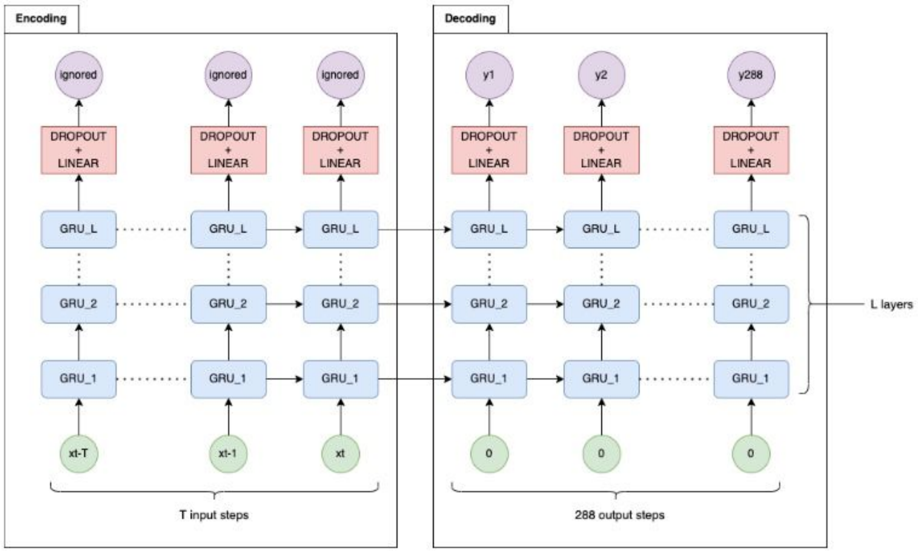
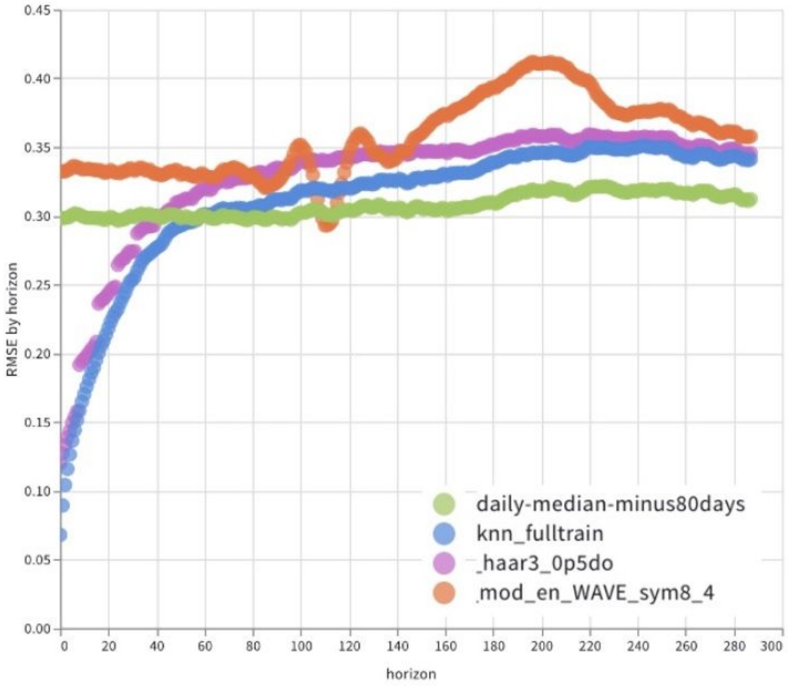


Figure 3: Diagram of the GRU model.

Detailed method

Traditional models: K-nearest neighbors



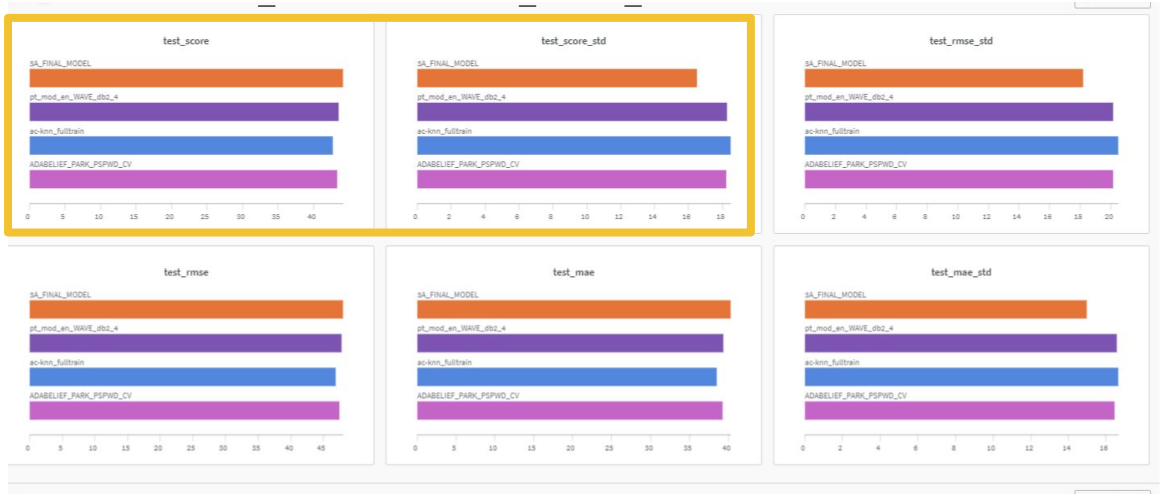
- Large percentage of invalid data points:
  - models over-fit easily,
  - high error for longer-term predictions.
- Explored **simpler models for long-term predictions**: mean and median models per timestamp and TurbID, ARIMA models, different Wavelets transform models, Fourier transform models, exponential smoothing models and KNN models.
- Best traditional model performer: **KNN for time series**

Offline leaderboard

Weights & Biases

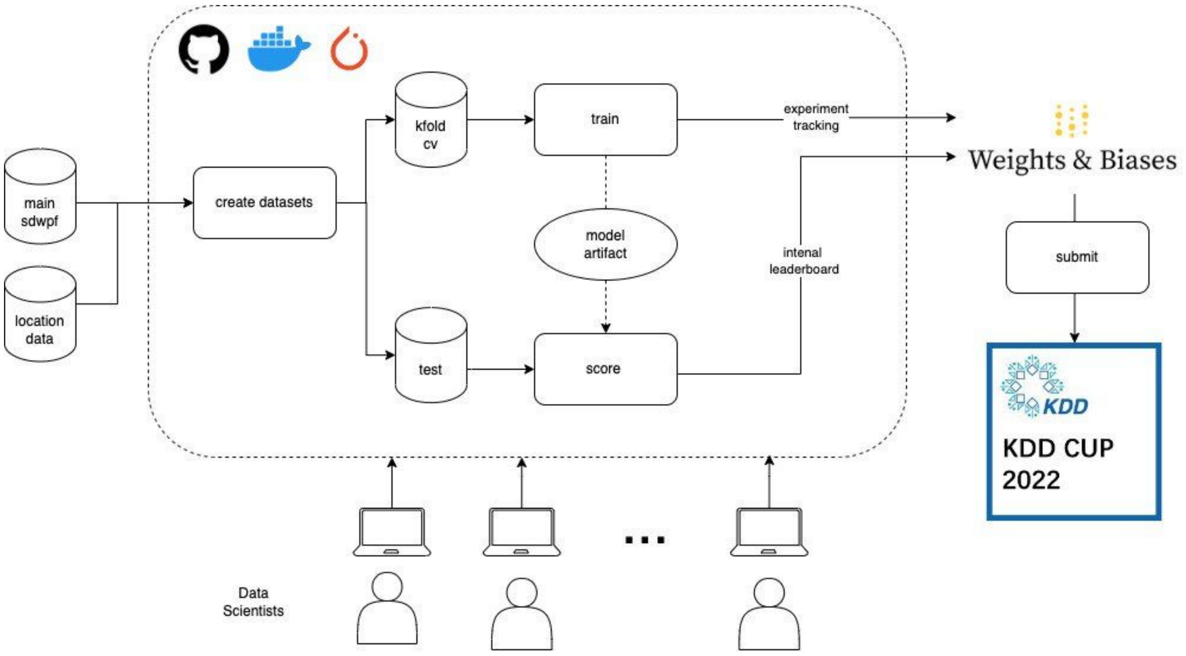
- **Automatic** model comparison between team members
- 2,048 test samples with Improved evaluation code: 2,048 samples scored in 15 minutes.
- Importance of **STD of 2048** predictions to evaluate model variance and model transfer to online test set.

Offline Test_score   Test_score_std



Environment

Team workflow

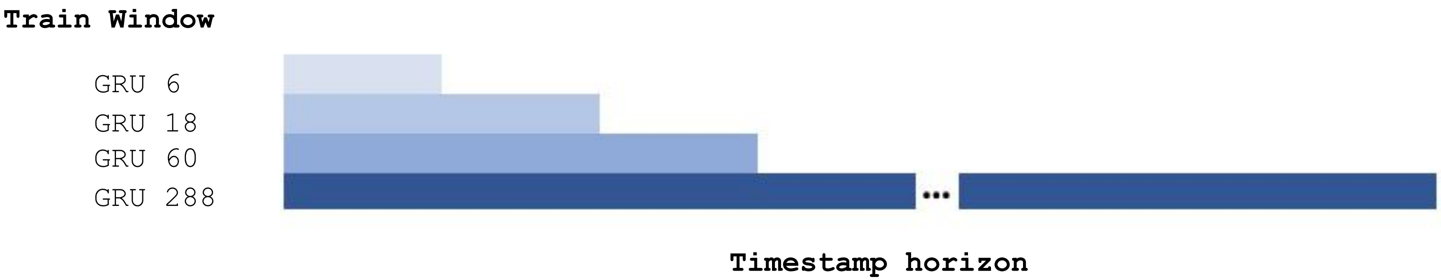


- **Docker** (working linux & windows platform).
- **Poetry** packaging and dependency management easy
- **Weights and biases API**
- **Pytorch** framework

Modeling strategy - GRU Model

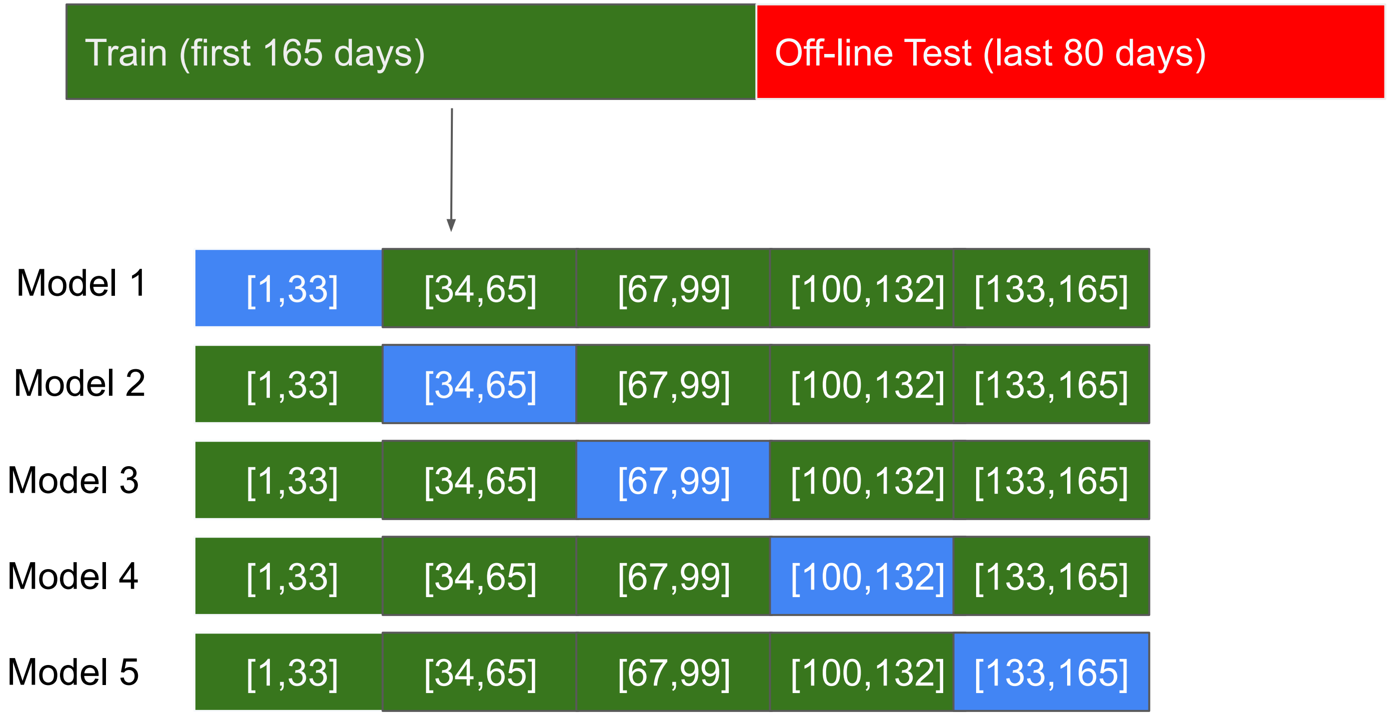
GRU model for different horizons

- Given that the prediction horizon is **ambitious** (long).
- Four models were trained for different horizons (288, 60, 18, and 6) in cross-validation.
- Named GRU288, GRU60, GRU18 and GRU6.



Modeling strategy - GRU Model

Cross-validation: one model for each fold (5 folds)



W&B API

Metrics for 5 GRU folds

Modeling strategy - KNN Model

One model using full data+test_x

Each knn prediction (test_x) searches through the full train data + the 14 previous days provided with test_x:

1) The last **6** data points (1 hour) are compared to historical sequences of length **6.**
2) By measuring the distance (euclidean distance) the closest **2,000** neighbors are chosen.
3) A prediction of length **288** is produced by computing a weighted average for each neighbor's distance.

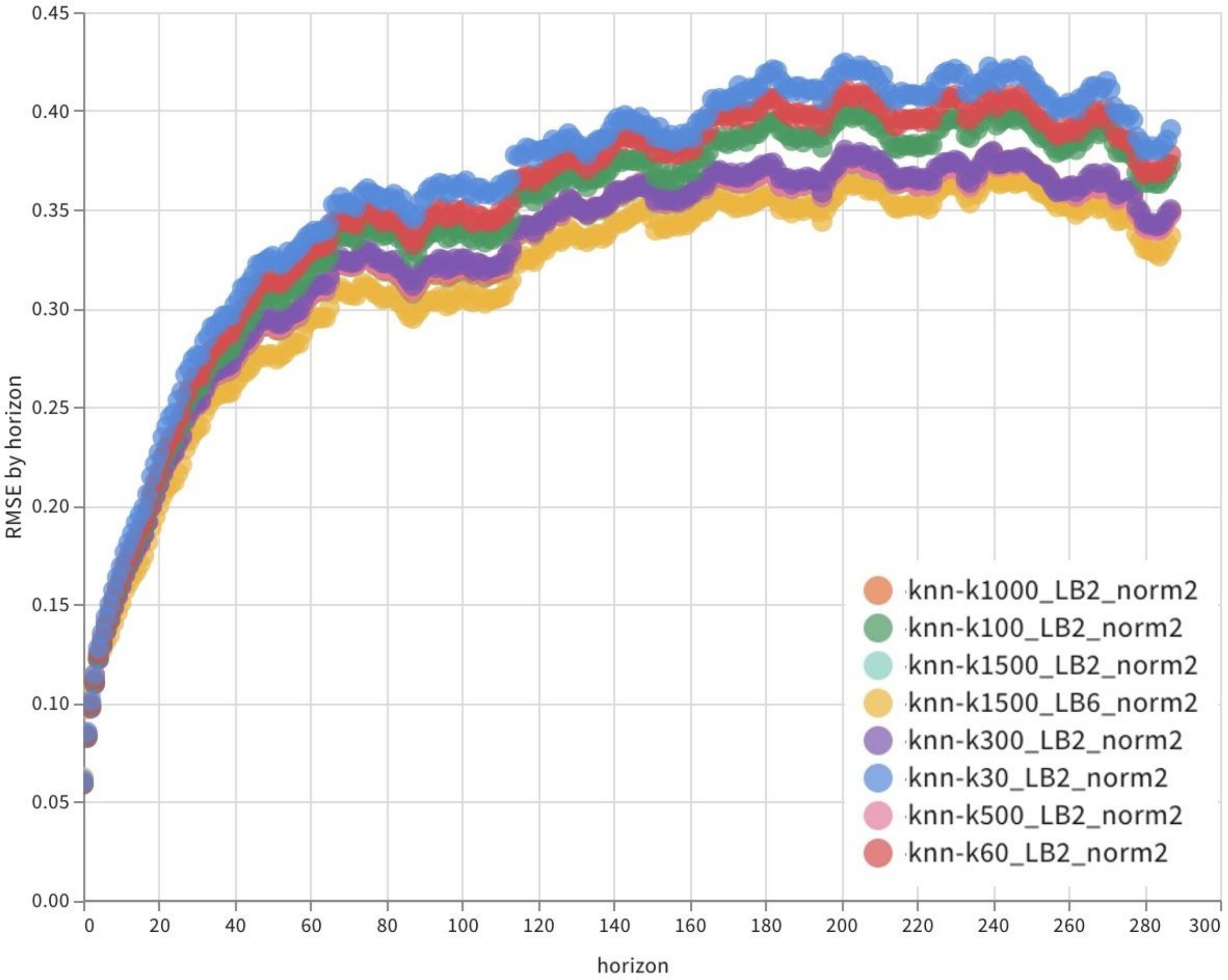
$$\texttt{Weights} = \frac{1}{1 + \text{distance}_{\text{neighbor}}}$$

No cross-validation. Hyperparameters chosen by off-line testing with w&B set-up

KNN Models

Hyperparameter tuning using W&B

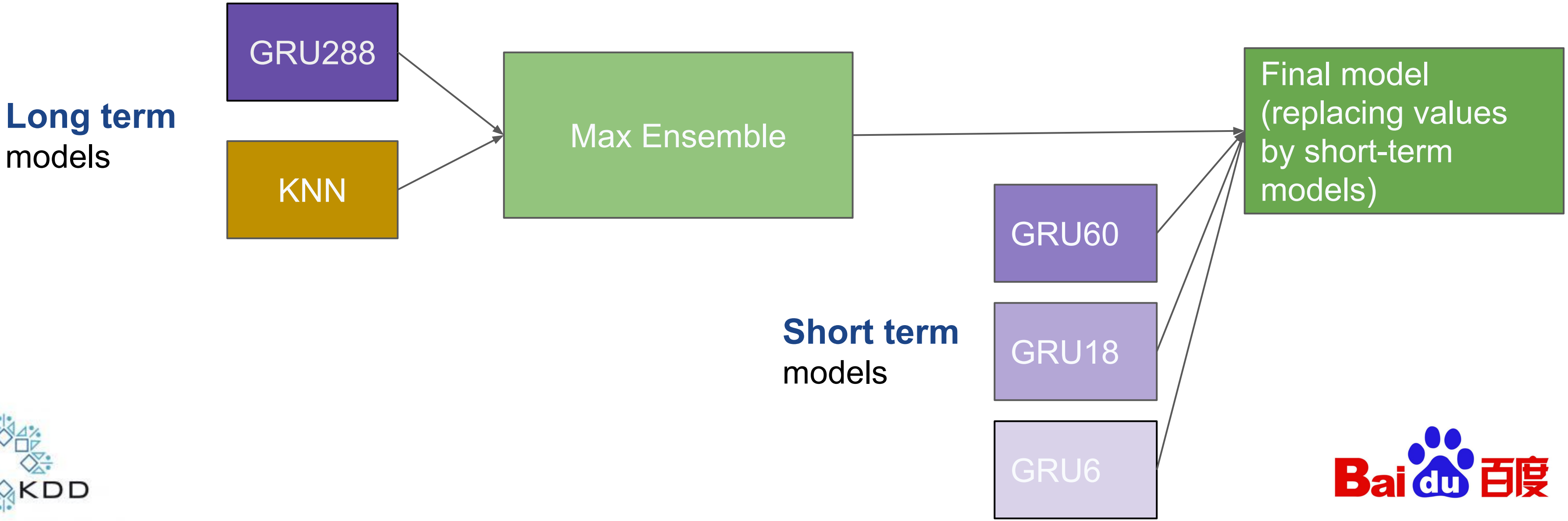
Example of average error by horizon graphs introduced in W&B ➜ enabled easy model ensembling by prediction horizon



Ensemble - final model

GRU + KNN

For each GRU we applied **MAX** k-fold instead of **MEAN** kfold.

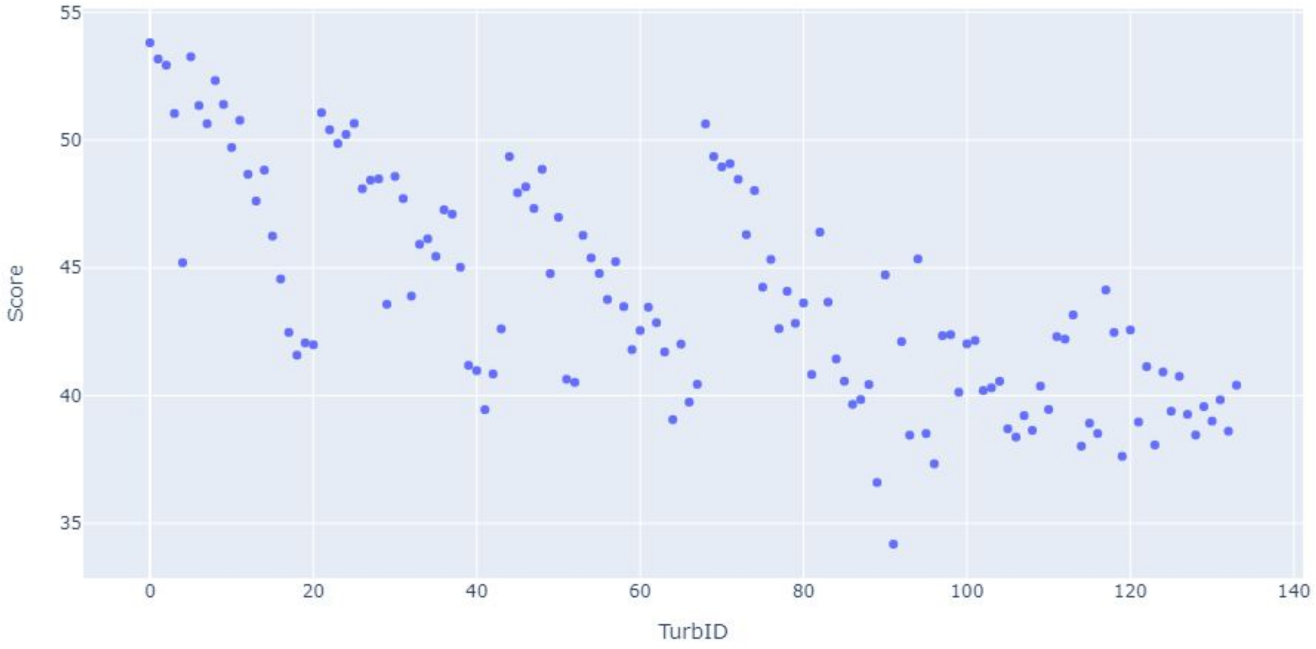


Ensemble - final model

Predictions - 5 models and hybrid model

Ensemble - final model

Score by turbID



Conclusions

Final remarks

- Real turbine park data, large number of invlaid data points, challenging prediction **without** weather forecasts.

- Poor online-offline transfer :
    - **Hard** to have a robust offline measurement. Some models with better offline score worsen when evaluated online.
    - Created an evaluation environment for online validation and comparison of different model metrics with 2048 scores with W&B.

- Final ensemble of various models chosen through off-line evaluation environment: GRU trained with different time horizons trained in 5 folds and KNN model.

- Online test period corresponds to periods with higher Patv (higher scores). By performing the **maximum** instead of the mean, the outliers in the 2,048 decreased error **deviation** - few samples in test (142).

KDD

UNIVERSIDAD
COMPLUTENSE
MADRID

Baidu 百度

飞桨 AI Studio