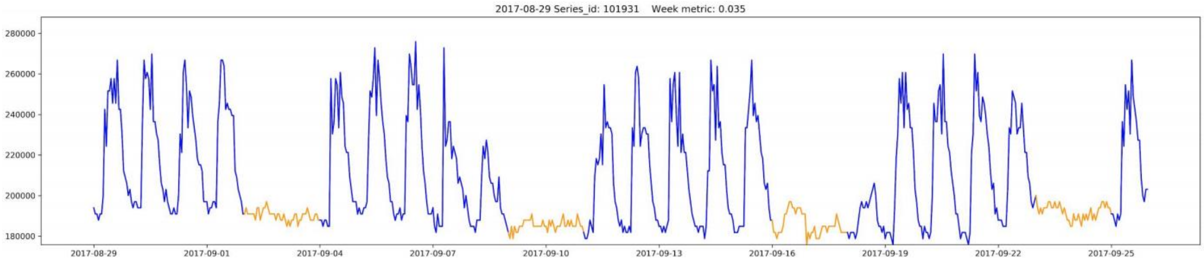
DrivenData Competition
Power Laws: Cold Start Energy Forecasting

Competition

he competition was based on data provided by Schneider Electric which described the electricity consumption in commercial buildings.

The goal was to develop models that, basing on historical data, would be able to predict the future consumption on different detail levels (hourly consumption the following day, daily consumption next week, etc.).

Competition data examples:



Challenges

Specification of the competition

Distinguishing element of the competition was significant limitation regarding the training data set:

- historical training data was limited to at most 4 weeks and the test data comprised a maximum of 14 days (hence the name of the competition - Cold Start)
- the building metadata was of poor quality, in particular the location of buildings (eg. the country, continent) and calendar of working days were unknown.

Aggregation of forecasts

The next challenge was varying ways of forecasts' aggregation; for certain buildings the consumption should have been foreseen for subsequent 24h, while for others the next 7 days were required, in daily aggregation.

Training data

Additional problems were caused by the distribution of test data which was different than that of the training data. It made reliable local model validation very difficult which was the key part of the competition due to the small size of the number of available applications.

Process

01 Exploratory data analysis using Jupyter notebooks

02 Data cleaning consisting of the removal of incorrect readings
03 Proposition of input attributes - due to the volume of the data, and visualization tools
and supplementation of the missing ones
this stage required effective implementation using the Pandas library
04 Proposition of methods for model validation
05 Machine-learning experiments
06 Selection of the best attributes and models for the final solution

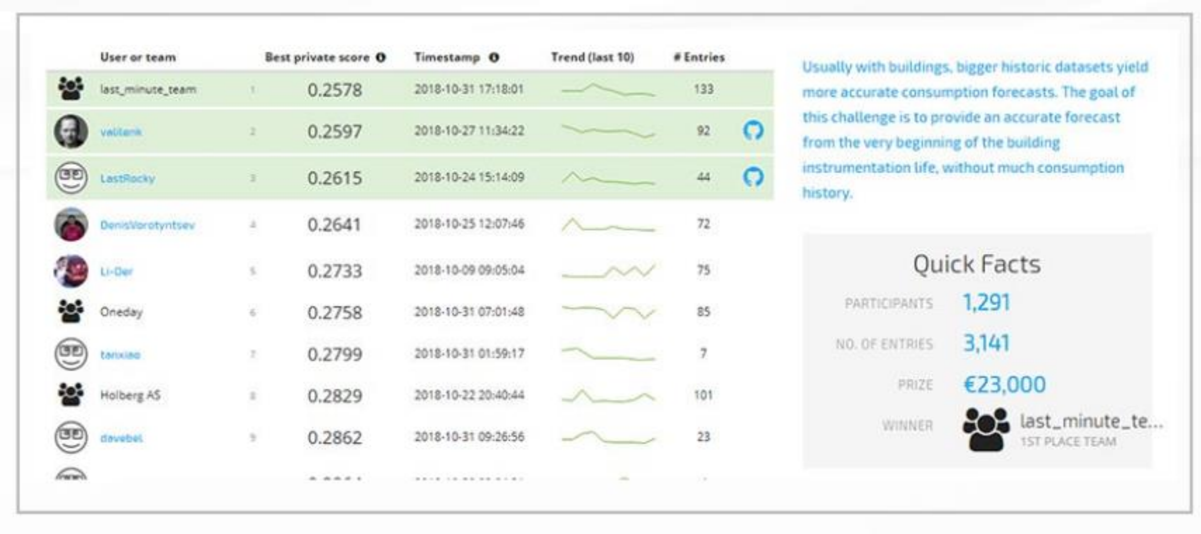
Solution

The final model was based on **neural networks** implemented using the Keras library. The best application that achieved the smallest error was averaging the results generated by our model and model of the other team member (Guillermo Barbadillo).

KEY ELEMENTS:

- treating different aggregation methods as separate problems and creating separate dedicated models for these
- maximal use of training data - one time series was used to create a dozen or so data points
- extended attributes that included, among others, selected data regarding power consumption from the past
- a custom model of neural networks that on the one hand significantly reduced the network size and made it easier to train, and at the same time allowed to capture the daily and weekly rhythms in the forecasts.

Summary



	User or team	Best private score	Timestamp	Trend (last 10)	# Entries
1	last_minute_team	0.2578	2018-10-31 17:18:01		133
2	valilenk	0.2597	2018-10-27 11:34:22		92
3	LastRocky	0.2615	2018-10-24 15:14:09		44
4	DenisVorotyntsev	0.2641	2018-10-25 12:07:46		72
5	Li-Der	0.2733	2018-10-09 09:05:04		75
6	Oneday	0.2758	2018-10-31 07:01:48		85
7	tanxiao	0.2799	2018-10-31 01:59:17		7
8	Holberg AS	0.2829	2018-10-22 20:40:44		101
9	davebel	0.2862	2018-10-31 09:26:56		23

Usually with buildings, bigger historic datasets yield more accurate consumption forecasts. The goal of this challenge is to provide an accurate forecast from the very beginning of the building instrumentation life, without much consumption history.

Quick Facts

PARTICIPANTS	1,291
NO. OF ENTRIES	3,141
PRIZE	€23,000
WINNER	last_minute_te... 1ST PLACE TEAM

The obtained model achieved the result of 0.2578
(which roughly corresponds to the average forecast error of around 25%).
That was sufficient for the first place in the competition.

Used technologies

Keras TensorFlow Microsoft LightGBM scikit learn pandas NumPy jupyter

FintureAI
ul. Skierniewicka 10A
01-230 Warszawa
hello@fintureai.com