Решение пробного задания по предсказанию серий.

Код: <https://github.com/ikamensh/time_series>

Данные: 250 активов, цены которых меняются каждый час. Цены не меняются по субботам, и данные за субботы отсутствуют. Всего 45 недель данных по всем активам.

БЛА БЛА БЛА нужно online learning… :/

Функция потерь: Absolute mean error (выбрана потому, что эта самая ошибка составит убытки при неправильных решениях. Именно ее необходимо минимизировать; выбор квадратной или другой функции потерь даст худшую на практике модель).

Модели, которые я попробовал: LSTM, CNN + LSTM, catboost (представляя последовательность за неделю как 120 features). Хорошие результаты дала LSTM c 2мя слоями:

our model allows for the profit of: 143098.0 (only last 8 weeks)

maximum possible profit is: 201218.0 (only last 8 weeks)

Если бы это была оплачиваемая работа и код бы реально решал какие активы купить, его можно было бы улучшить следующим образом:

1. Дополнить основную модель специальной моделью для каждого актива
2. Попробовать классические модели предсказания последовательностей, такие как:

GARCH, ARIMA, Markov chains и другие

1. Использовать Grid Search для оптимизации гиперпараметров
2. …
3. PROFIT!

Код: <https://github.com/ikamensh/time_series>

Запускается так:

1. read\_data.py
2. train\_rnn.py
3. forecast.py

необходимы библиотеки:

* keras
* tensorflow
* sklearn
* pickle
* …