Ocean Protocol ::
ETH Prediction Challenge

Challenge description

Challenge

Contestants are requested to submit the following:

- Csv with the predictions for ETH price over a 24-hour period, with first prediction on from Mon Oct 17, 2022 at 1:00am UTC, and a prediction every hour on the hour after that, for 24 predictions total.
- A report of 10-30 slides describing your approach, and insights you gained in making the predictions. Think of this as a "conference presentation" for your work.
- Optional: add at least one slide presenting any challenges faced during the bounty and/or recommendations you may have.

Evaluation Criteria

- 80% of the competitor's score will be objectively measured, based on Normalized Mean Square Error(NMSE) of the predictions (1). The README shows the precise calculations.
- 20% of the competitor's score will be based on the slides (2), subjectively measured by a panel of judges. The criteria will be like those of a presentation in a technical conference.

Input dataset

You can download the dataset from here:
https://cexa.oceanprotocol.io/ohlc?exchange=binance&pair=ETH/USDT&period=1h
This data feed returns ETH/USDT ohlc history: {open, high, low, close) over time, drawn from Binance. It's over the most recent 500 hours: every hour, on the hour.

ETH close price hourly plot:

The last 3 rows from our dataset:

	open	high	low	close	volume
dt1					
2022-10-16 14:00:00	1,284.65	1,285.60	1,280.19	1,283.12	8,268.21
2022-10-16 15:00:00	1,283.13	1,287.79	1,282.97	1,286.35	8,622.73
2022-10-16 16:00:00	1,286.36	1,287.71	1,285.48	1,286.44	2,257.71

We converted the time to UTC+0 timezone as it is needed in the competition.
It is important to note that we have the holdout period – it is the number of hours to the first prediction.
In our case we have the latest data at 16:00 so we didn't see 8 data values to the first prediction point.

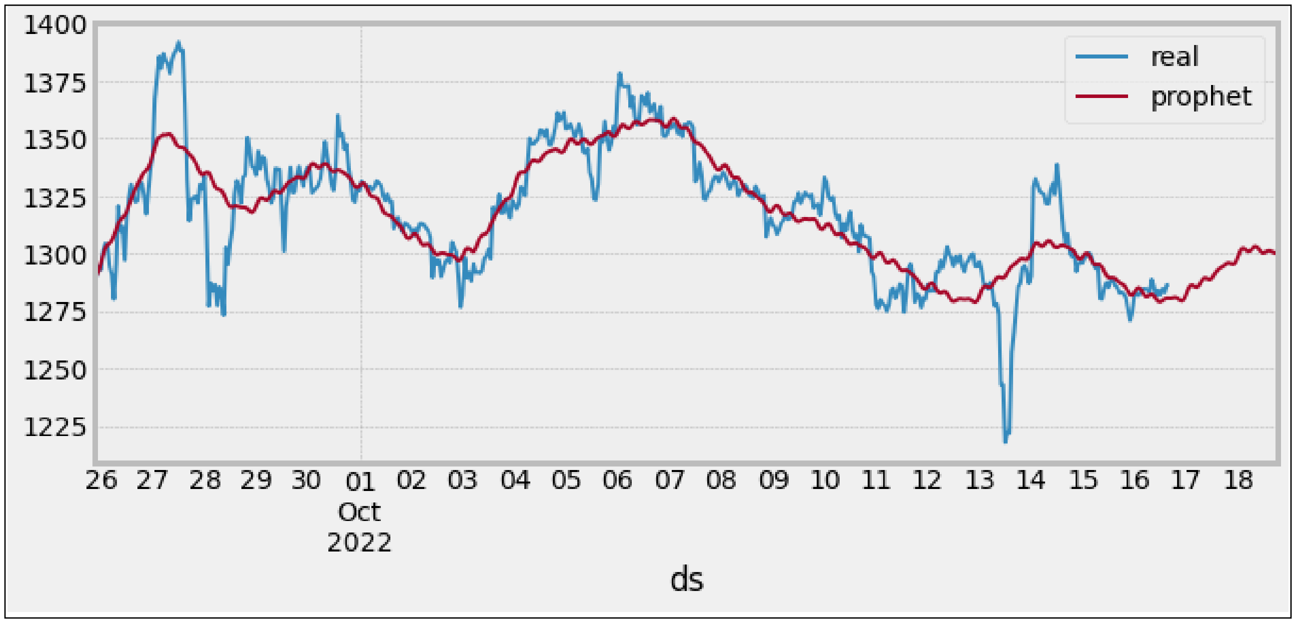
*More details you can find in the following script: *1_get_predictions.ipynb*

ocean

Final predictions

We have used the prophet model for getting the final predictions (https://facebook.github.io/prophet/). The parameters of the model is the following: m = Prophet(weekly_seasonality=True, daily_seasonality=True).

ETH close price hourly predictions:

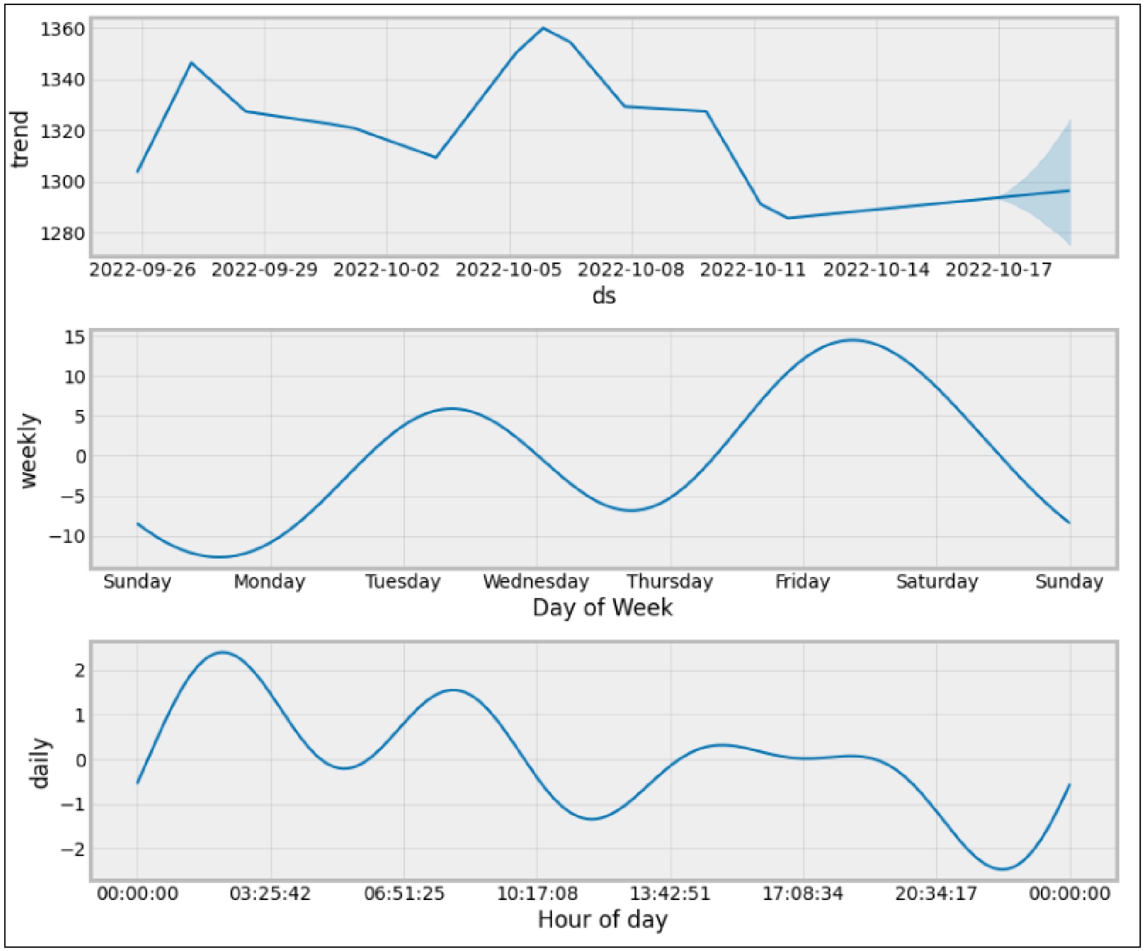


*More details you can find in the following script: *1_get_predictions.ipynb*

Model description

By running the command *m.plot_components()* we can see the following description of the model. We have the trend, weekly and daily components of the model.

Prophet model components:



*More details you can find in the following script: *1_get_predictions.ipynb*

ocean

Models testing

We have tried to train 3 models:
- last available value prediction
- exponential smoothing
- prophet model

We have evaluated them by different metrics:
- nmse – normalized mean squared error (main competition metric)
- mse – mean squared error
- mae – mean absolute error
- mape – mean absolute percentage error

We have used one day for evaluation set (14,15 or 16 of October), the data before that day was used for training but with one more parameter *hours_skip* – it is the number of last values to skip in the training set (holdout period – we will explore the dependencies of that parameter on the quality on the next slide).

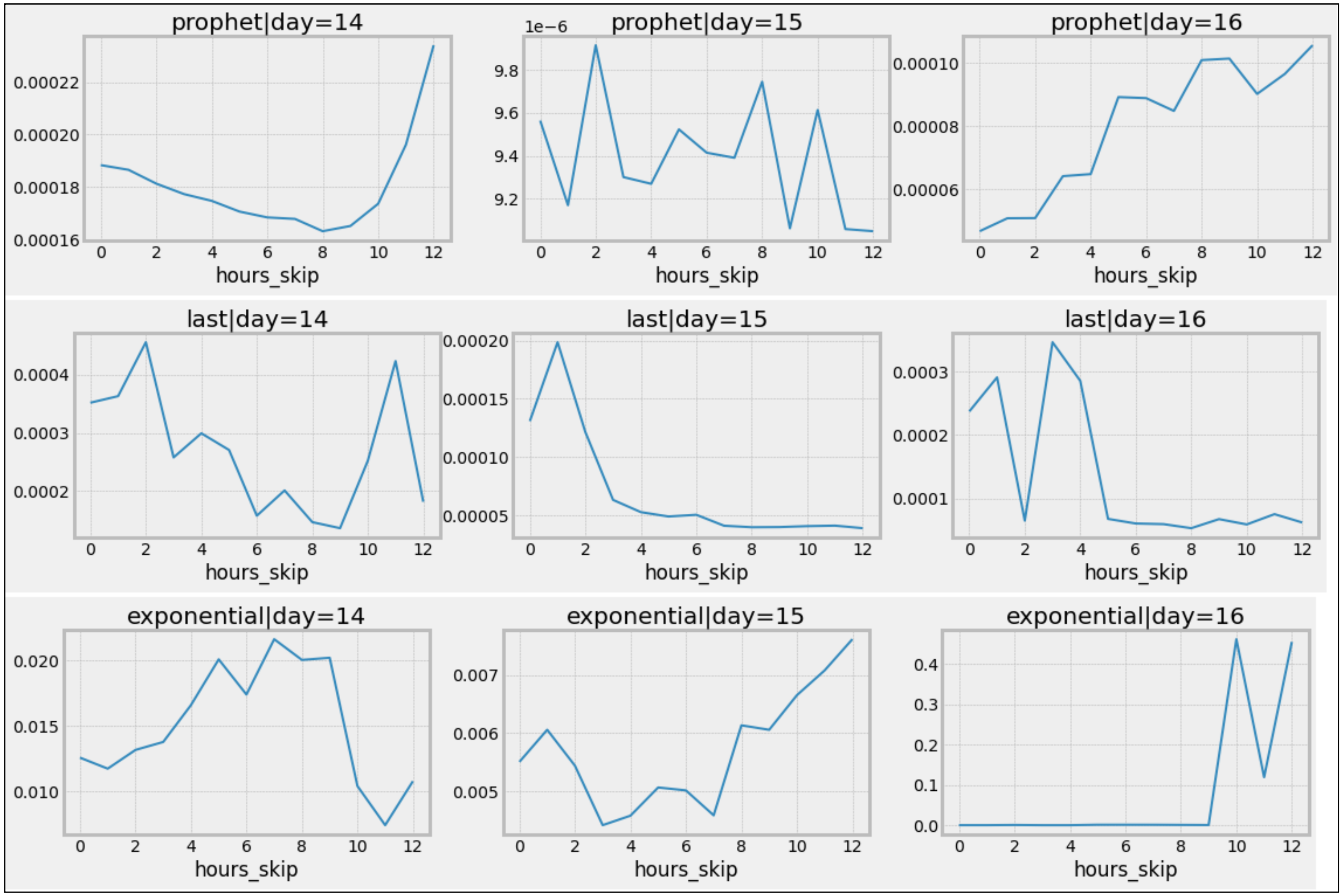
Prophet model is the best model for the most number of validation results:

	nmse			mse			mae			mape		
day	**14**	**15**	**16**	**14**	**15**	**16**	**14**	**15**	**16**	**14**	**15**	**16**
method												
exponential	0.007419	0.004421	0.000128	12,851.449538	7,326.537462	212.199923	109.578713	85.345469	9.183035	0.083042	0.066305	0.007068
last	0.000136	0.000039	0.000054	236.441487	65.011750	89.075313	12.863750	6.647500	5.422917	0.009830	0.005156	0.004164
prophet	0.000163	0.000009	0.000047	282.473698	14.992973	77.421920	15.036990	2.346065	5.122902	0.011374	0.001828	0.003936

*More details you can find in the following script: *3_models_testing.ipynb*

The dependencies of hours_skip parameter on the NMSE metric

As we can see on the picture below based on validations results that there is now a high drop in quality if we will use not fresh data (let's say 5-10 hours to the first prediction). In our case we didn't have a chance to have a fresh data and we have executed the forecasting scripts in 8 hours before the deadline.



*More details you can find in the following script: *3_models_testing.ipynb*

Recommendations & Ideas for the future work

Results:

1. We have tested 3 approaches on the validation sets: last value prediction, exponential smoothing and prophet model. And after that explorations we decided to use prophet model to generate the final predictions.
2. We have shown that there is now a high drop in quality if we will use not fresh data – if we will have 5-10 hours of holdout period to the first prediction.

The ideas how to improve the results:

1. Try to collect more data of ETH prices, maybe 500 points is not enough to get accurate predictions..
2. Based on extended dataset try to train Machine Learning models (Boosting, Random Forests..) that can use simple models predictions (prophet, exponential smoothing…) as features.
3. Based on extended dataset try to train multistep Recurrent Neural Networks (LSTM, GRU) with different architectures.