High-Frequency Trading Project

Pitch presentation

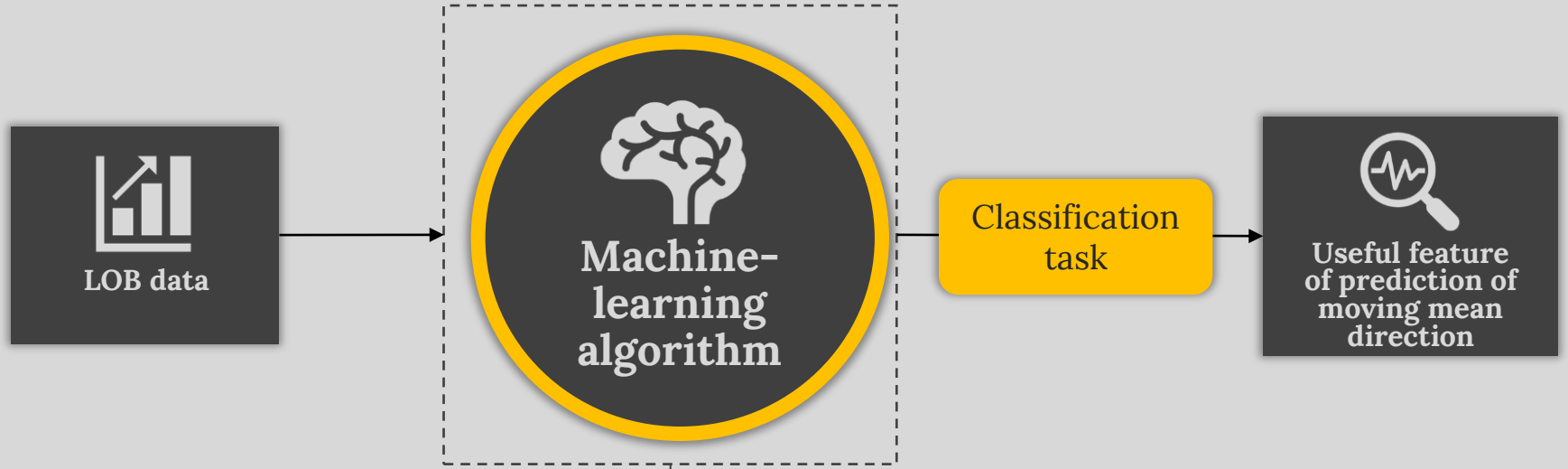
Vyacheslav Blinov

Nikita Bogomazov

Roman Vetrin

Project idea

We propose a solution for a feature generation which can be used in trade engines



A Virtual Machine solution with following features:

- High-speed data storage
- End-to-end preprocessing
- Possibility to adapt to changing trend

Competitive edge

Our solution will have advantages against of most players in the industry

Most of publication do not utilize full information from the LOB microstructure

Open-Close-High-Low data generate only **4** features

Over 20 features for **5** levels

Many trading solutions do not utilize low-latency data storages

High-frequency trading occurs in sub-second time

Our solution utilize high-speed storage for sub-millisecond data access

Business Understanding

We formulated the business and data mining objectives

Business objective

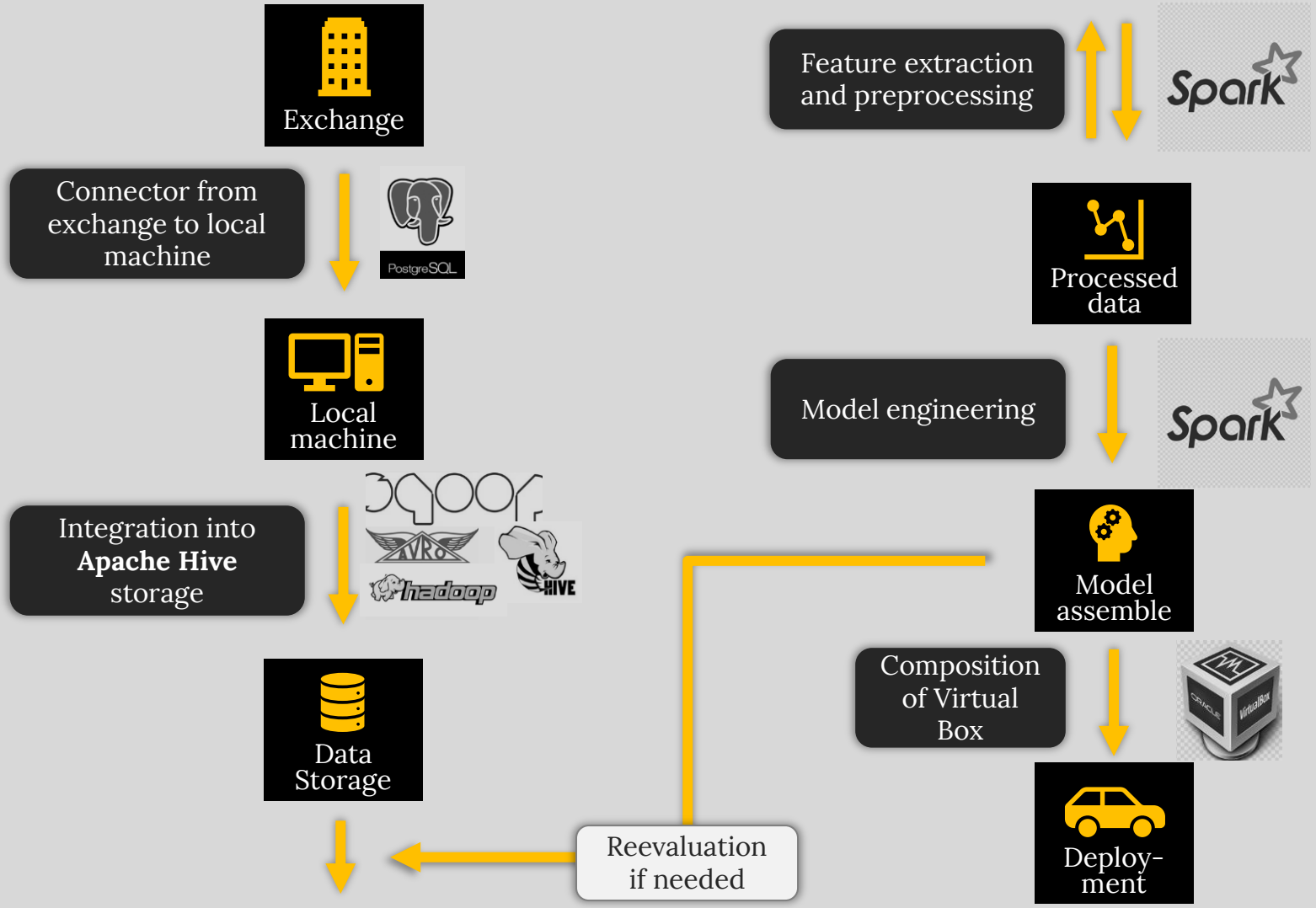
Model	Model capable to generate signal on future trend direction
Storage	Create fast infrastructure
Backtest	Evaluate on the backtest engine

Data mining objectives

Model	Predict moving mean BTCUSDT pair one second ahead
>1s	Ability to preprocess and make predictions faster than 1 second
PnL	Achieve positive PnL during the backtesting stage

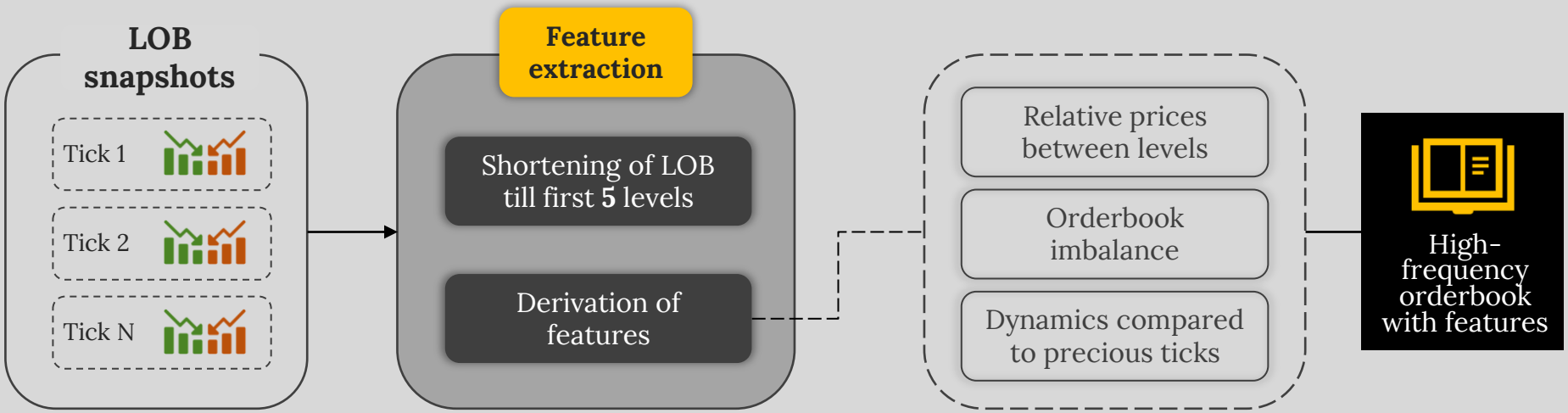
Project Plan

We developed the following stages of the project



Feature generation

We propose a solution for a feature generation which can be used in trade engines

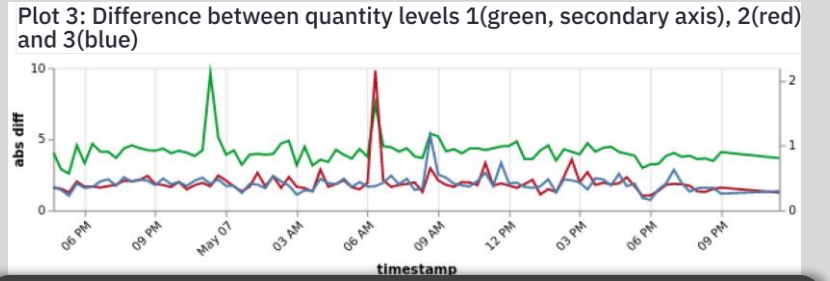


Exploratory Data Analysis

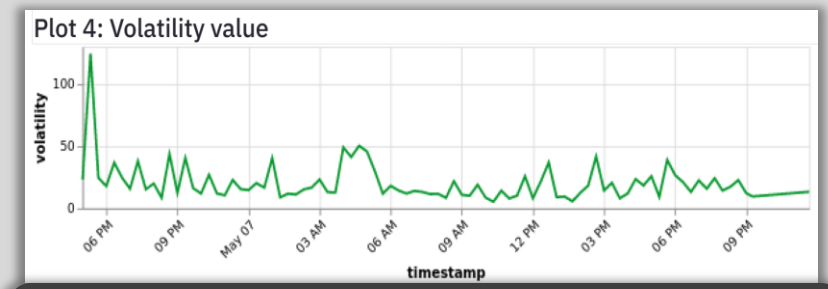
We revealed several insights from the data:

As we may observe, the price changes which appx. 5% margin within the day which is quite high comparing to less volatile assets.

Few sudden changes in candlestick graph corresponds to jumps in price difference between bids and asks.



Some pikes in quantity difference correspond with future highly volatile displayed on the Plot 1.



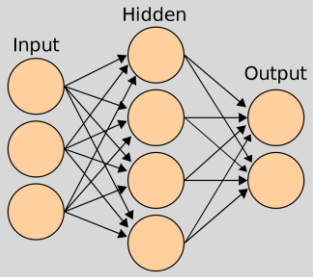
it is feasable to divide the trend into the two parts as the levels seems overall constant throught the day.

Model

The best model among evaluated is Multilayer Perceptron Classification model

Data Storage

High-frequency orderbook with features



Neural network with
12 – 512 – 128 - 2 layers
40 iterations

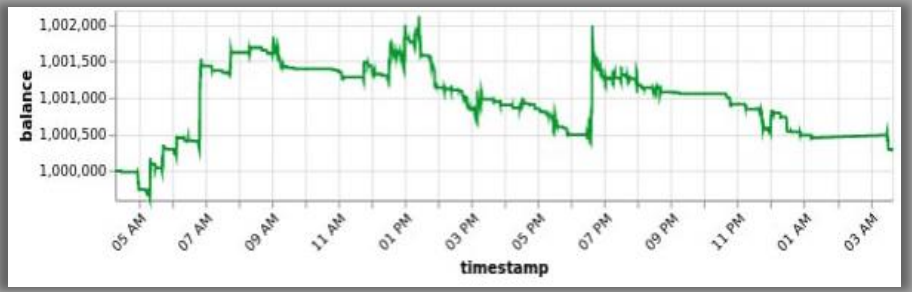
0 1

Signal on upward or downward trend

Backtest settings

0.01%	Commission
0s	Execution time
Asset price corresponds to calculated mid price	

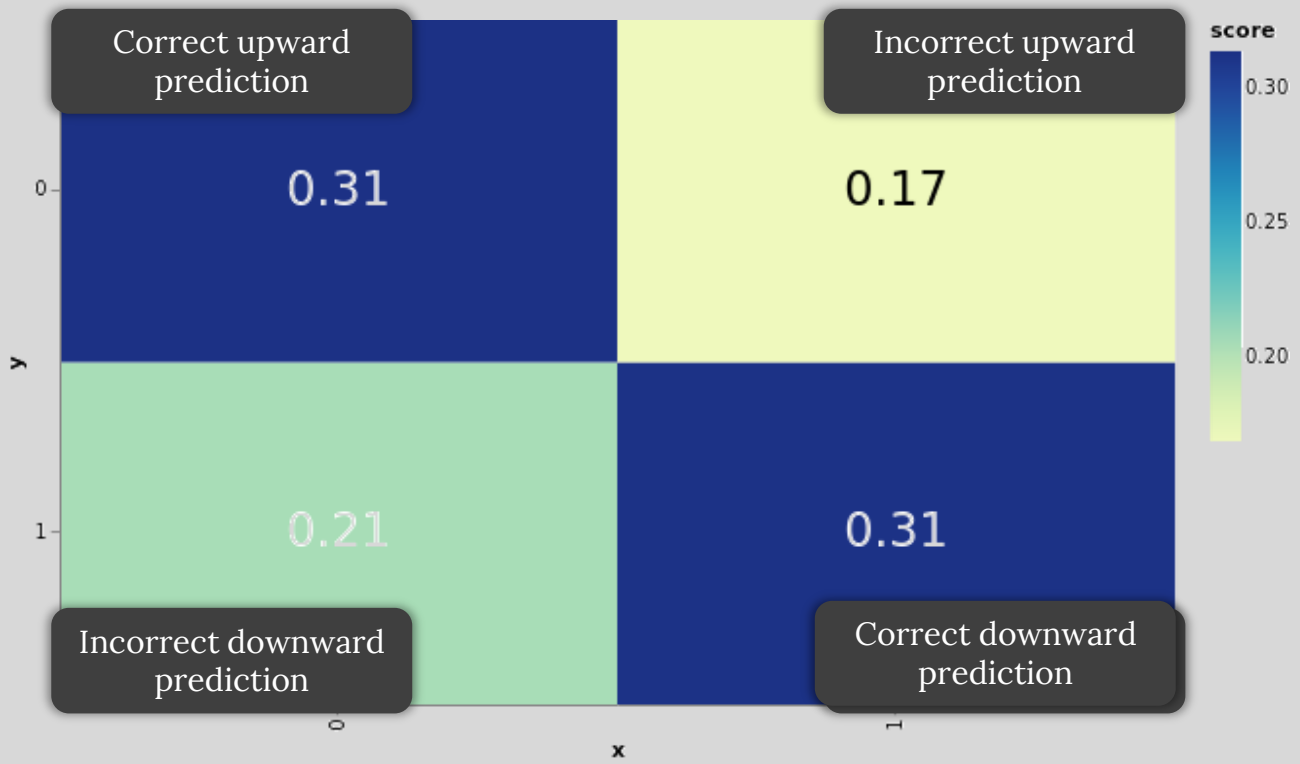
Backtest results



Backtest figures

1m	Start capital
0.14%	Daily return
51%	Expected annual return

Confusion matrix of a model



Next steps

It is recommended to fund a dedicated server which can be placed in the location of Binance exchange and where the virtual machine will be placed.

In this way the latency loss will be minimized and such system will produce more accurate data.

It is expected for the model to become outdated fast given that the cryptocurrency trends tends to change very quickly over time.

Thus, an engineer should be assigned to the task of monitoring performance and retraining of the model using the previously precessed data.

Project participation

Stage	Vyacheslav Blinov	Nikita Bogomazov	Roman Vetrin
Business Understanding	10%	10%	80%
Data Understanding	50%	25%	25%
Data Preparation	80%	20%	0%
Modelling	10%	80%	10%
Evaluation	0%	30%	70%


Thank you for attention
GitHub
GitHub link