HSBC Hackathon

Problem statement, along with a simple model, is available at below link. Participants will need to open this link and create a copy to work on it (File -> Save a copy in Drive..)

Read-only link -

https://colab.research.google.com/drive/1xXu7qgSf1_KeSlpy-sZ4CnXfaDLurEWm#scrollTo=XVNKK2CNrQK0

In the output, below is required from students

1. Minimum #predictions > 10000

2. Low RMS

3. 2 or 3 files with below names [Same folder]

a. Model.ipynb : Notebook file

b. Model.pdf : Model documentation

c. Model.pickle (optional) : Trained model

Background Summary

An order driven market is a financial market where all buyers and sellers display the prices at which they wish to buy or sell a particular security, as well as the amounts of the security desired to be bought or sold. In these markets, participants may submit limit orders or market orders.

In a limit order, you specify how much of the asset you want to buy or sell, and the price you want. If there are matching orders on the book (e.g. someone who wants to sell at the same price, or lower, as the price at which you want to buy), your order will be filled immediately. If not, your order will stay on the book until matching orders arrive (which could be never). It is also possible for a limit order to be only partially filled, if the counterparty wants to trade a smaller amount than you did. In that case the rest of the order remains on the book.

In a market order, you only specify how much of the asset you want to trade. Your order is then filled immediately at the best price currently available on the market. For instance, if you place a market buy order, you will be matched with the current lowest-priced sell order on the book. If that order is not large enough to completely fill yours, the next-lowest sell order will be used to fill some more of yours, and so on.(You are encouraged to go through the 1st suggested reading for a pictorial understanding of order book and price dynamics)

In this competition, we use tick data. Tick data refers to any market data which shows the price and volume of every print. Additionally changes to the state of the order book occur in the form of trades and quotes. A quote event occurs whenever the best bid or the ask price is updated. A trade event takes place when shares are bought or sold.

The aim of this competition is to determine the relationship between recent past order book events and future stock price for 30 seconds time-horizons. Few factors that are explored in the literature to predict price movements:

Order arrival rate

Bid-ask spread

Order book imbalance

Trade volume @ Bid price vs Trade volume @ Ask price

Certain factors, such as current order book imbalance, tend to have good predictive power for very short-time time-horizons (under 10-20 seconds), however other factors might be important for time-horizons of more than a minute.

Equity markets are very fast and it is important to understand that multiple high-frequency events can occur in the same milliseconds. Analysing and understanding the data is critical before applying machine learning models.

This problem is based on real-life problem we work on. Another important point to note - trade event and quote event timestamp will rarely be at the same-time, usually quote event time stamp is before trade event time stamp. Refer to examples in the link on how you could join this data

Suggested Reading

Basic introduction of Limit Order Book

Analyzing an Electronic Limit Order Book

Algorithmic Trading and DMA

HFT - Limit Order Book

Advanced Topics

The Information Content of an Open Limit Order Book

Order Imbalance Based Strategy in High Frequency Trading

Problem Statement

The aim of the problem is to develop a forecasting model to predict a stock's short-term price movement. The use of such prediction models is widely prevalent in algorithmic trading. Algorithmic trading, sometimes referred to as high-frequency trading in specific circumstances, is the use of automated systems to identify true(money making) signals among massive amounts of data that capture the underlying stock dynamics. These models can be leveraged to develop profitable trading strategies(akin to hedge funds) to help investors/traders achieve better returns. Contestants are expected and encouraged to think of empirical models/heuristics in order to better predict the price evolution of the hypothetical stock.

Submission Instructions

Algorithm/model should be developed without changing the order in the submission. Steps mentioned is the order of code execution.

1. Download files if not in local environment
2. Install all required libraries
3. Parameters such as file names for in-sample data & out-sample data
4. Algo-specific parameters
5. Functions to load data & evaluate performance
6. Train the model using in-sample data
    - Trained model should be a pickle or a function with values
7. Predict with the trained model using out_sample data and evaluate the model performance

Important Instructions

1. Cells that begin with "#[DONOTCHANGE]" shouldn't be changed
2. Submission will not be considered if the program fails to run
3. Tick frequency shouldn't be modified for out sample data
4. Only first-prediction will be considered for a mid-price until it changes [More detail below]
5. Model code should be commented

6. Submitted notebook shouldn't ideally require trade_in.csv and quote_in.csv files (insample files)
7. Brief description of the model (Preferably less than 1-page)

Required Output

1. Minimum #predictions > 10000
2. Low RMS
3. 2 or 3 files with below names
    1. Model.ipynb : Notebook file
    2. Model.pdf : Model documentation
    3. Model.pickle (optional) : Trained model

Queries

For any queries regarding the problem statement, you can contact the event coordinator:

Akshat Mittal | +91-7055223255

Or mail us at events@cognizance.org.in