MSFE ILLINOIS
MASTER OF SCIENCE IN FINANCIAL ENGINEERING

OPTIMAL EXECUTION
CME GROUP

IE597 PRACTICUM PROJECT #5 - SPRING 2023

CME Group

INTRODUCTION

TEAM MEMBERS

HARIHARAN MANICKAM

PRABHULING MASOODI

ANANYA SINGH

LIANG ZHANG

SPONSORS - CME GROUP

- **Mr. INDERDEEP SINGH**: Executive Director for Architecture and Product Management
- **Mr. MOHANDAS AYIKARA**: Director Data Science
- **Mr. GEORGE DEAMONT**: Manager of Data Science
- **Mr. PIYUSH KALKUTE**: Senior Data Scientist

MOTIVATION

- Optimal Execution of a large order
- Manage
  I. Volume
  II. Slippage
  III. Price Impact
- Participation rate is a key constraint
- Strengthen understanding of market dynamics

OVERVIEW

Problem Statement

- **Liquidate** a *large order** in a **2-hour** time window at the **Best** price.
- Dataset - **10-year bond (ZN) futures** 2 months of **nanosecond-second** High Frequency orderbook data for **20** levels(bid and ask)

Solution

- **Machine Learning model** - predicts next second price change.
  - Parameters of the model: Net-buy, Multilevel Order Flow Imbalance (MLOFI), Mean Reversion Lag
  - Output: Change in mid-price (ΔP)
- **Algorithm** - uses the model to decide the trade at that second
  - Direct constraint - *Participation rate**; Indirect constrains - *Replenishment rate*
- **Test Harness**
  - Simulates a **reactionless market** for the algorithm to trade
  - Gradually **replenishes** orders the algorithm makes

*A ***large order*** is defined as an order that is typically 10 to 20 times the size of the average volume on the top of the book.

*A ***participation rate*** is used to restrict the strategy from liquidating everything right away in a reactionless market, often set between 5%~20%

DATASET

Total Data

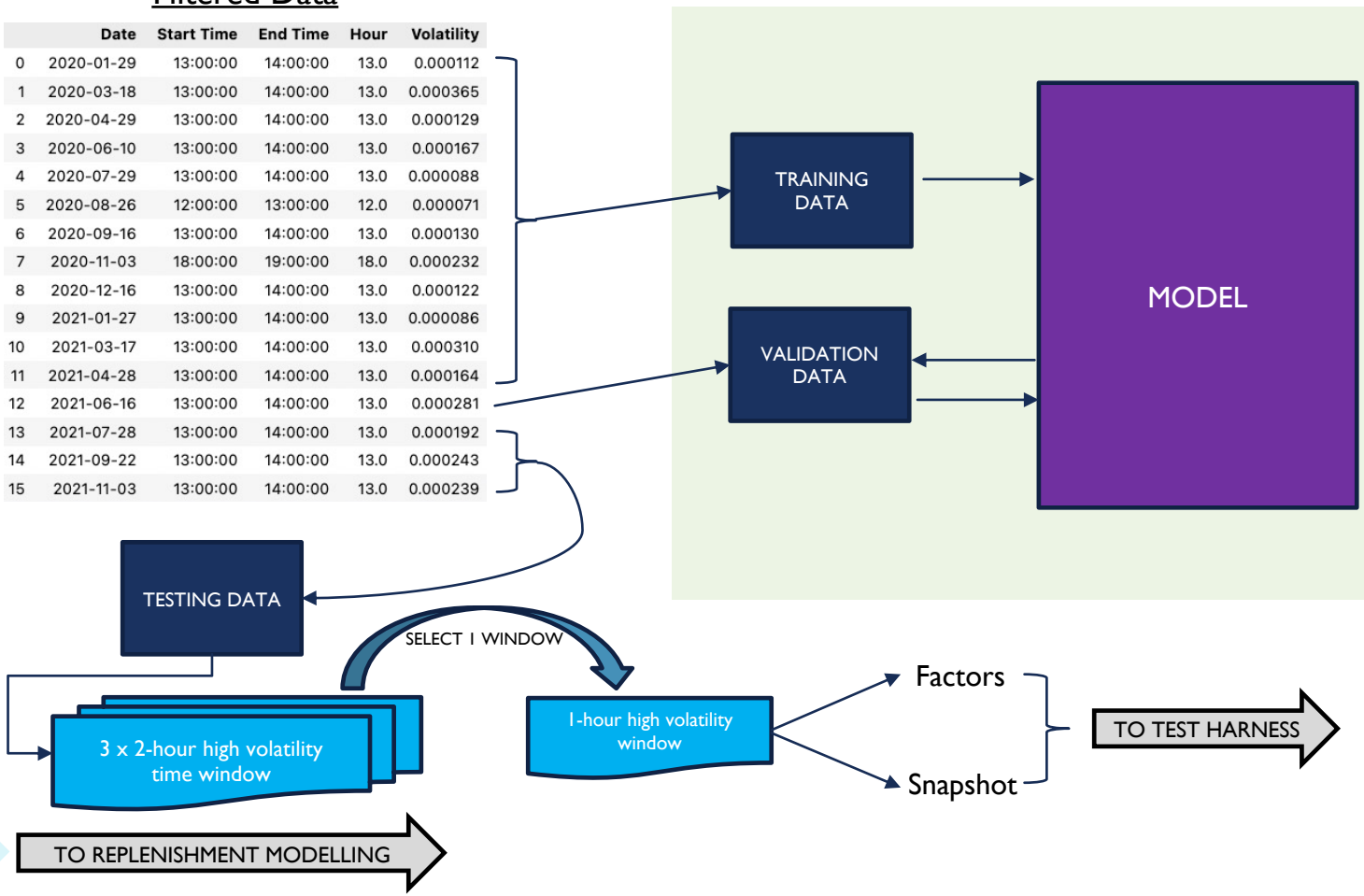
	Date	Start Time	End Time	# of seconds
0	2020-01-28	17:00:00	23:59:59	25200
1	2020-01-29	00:00:00	16:00:00	57601
2	2020-03-17	17:00:00	23:59:57	25198
3	2020-03-18	00:00:00	16:00:00	57601
4	2020-04-29	00:00:00	16:00:00	57601
5	2020-06-09	17:00:00	23:59:59	25200
6	2020-06-10	00:00:00	16:00:00	57601
7	2020-07-28	17:00:00	23:59:57	25198
8	2020-07-29	00:00:00	16:00:00	57601
9	2020-08-25	17:00:00	23:59:59	25200
10	2020-08-26	00:00:00	16:00:00	57601
11	2020-09-16	00:00:00	16:00:00	57601
12	2020-11-03	17:00:00	23:59:59	25200
13	2020-11-04	00:00:00	16:00:00	57601
14	2020-12-16	00:00:00	16:00:00	57601
15	2021-01-26	17:00:00	23:59:59	25200
16	2021-01-27	00:00:00	16:00:00	57601
17	2021-03-16	17:00:00	23:59:59	25200
18	2021-03-17	00:00:00	16:00:00	57601
19	2021-04-27	17:00:00	23:59:59	25200
20	2021-04-28	00:00:00	16:00:00	57601
21	2021-06-16	00:00:00	16:00:00	57601
22	2021-07-27	17:00:00	23:59:59	25200
23	2021-07-28	00:00:00	16:00:00	57601
24	2021-09-21	17:00:00	23:59:54	25195
25	2021-09-22	00:00:01	16:16:34	58594
26	2021-11-03	00:00:00	16:00:00	57601

Filter

Unfiltered

Filtered Data

	Date	Start Time	End Time	Hour	Volatility
0	2020-01-29	13:00:00	14:00:00	13.0	0.000112
1	2020-03-18	13:00:00	14:00:00	13.0	0.000365
2	2020-04-29	13:00:00	14:00:00	13.0	0.000129
3	2020-06-10	13:00:00	14:00:00	13.0	0.000167
4	2020-07-29	13:00:00	14:00:00	13.0	0.000088
5	2020-08-26	12:00:00	13:00:00	12.0	0.000071
6	2020-09-16	13:00:00	14:00:00	13.0	0.000130
7	2020-11-03	18:00:00	19:00:00	18.0	0.000232
8	2020-12-16	13:00:00	14:00:00	13.0	0.000122
9	2021-01-27	13:00:00	14:00:00	13.0	0.000086
10	2021-03-17	13:00:00	14:00:00	13.0	0.000310
11	2021-04-28	13:00:00	14:00:00	13.0	0.000164
12	2021-06-16	13:00:00	14:00:00	13.0	0.000281
13	2021-07-28	13:00:00	14:00:00	13.0	0.000192
14	2021-09-22	13:00:00	14:00:00	13.0	0.000243
15	2021-11-03	13:00:00	14:00:00	13.0	0.000239



MACHINE LEARNING MODEL

TRAINING

1
TRAINING DATA

PARAMETERS

Net Buy

MLOFI

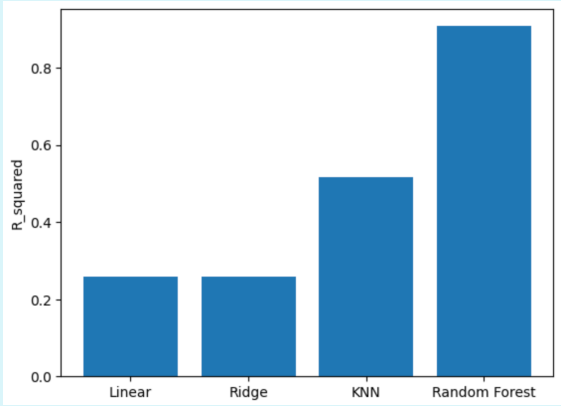
Mean Reversion Lag

2
ML MODEL

SELECTING

VALIDATION DATA

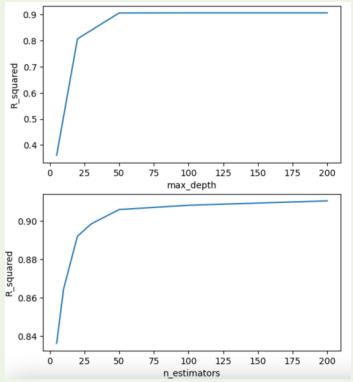
3
PERFORMANCE



BEST MODEL

BOOSTING

RANDOM FOREST MODEL

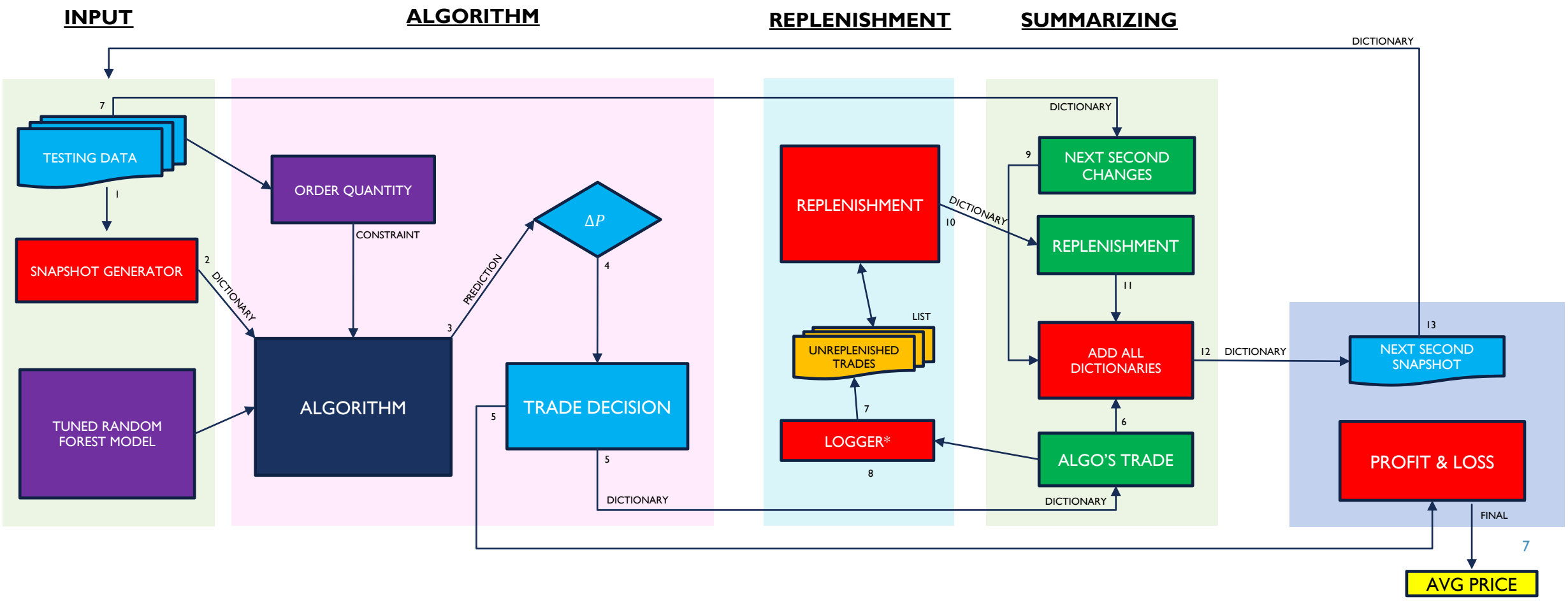


OPTIMIZED

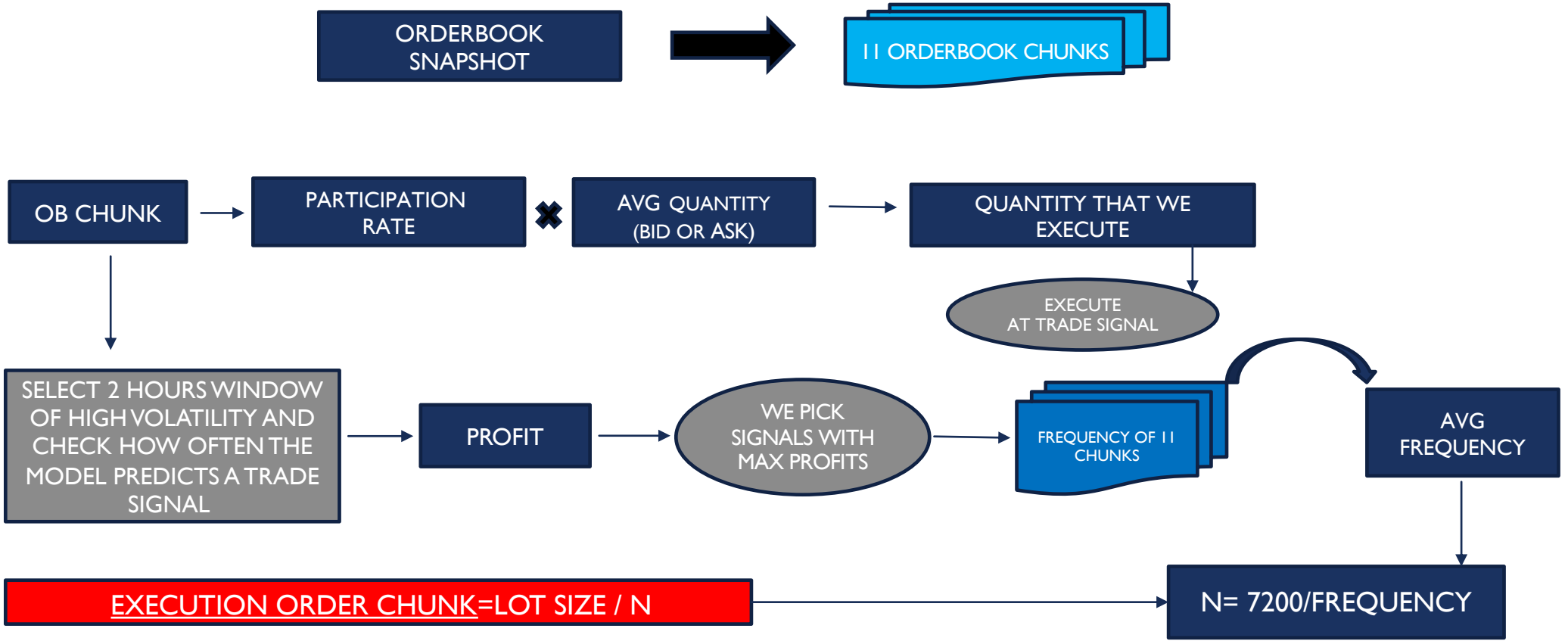
TUNED RANDOM FOREST MODEL

TO TEST HARNESS

TEST HARNESS



EXECUTION ORDER SIZE

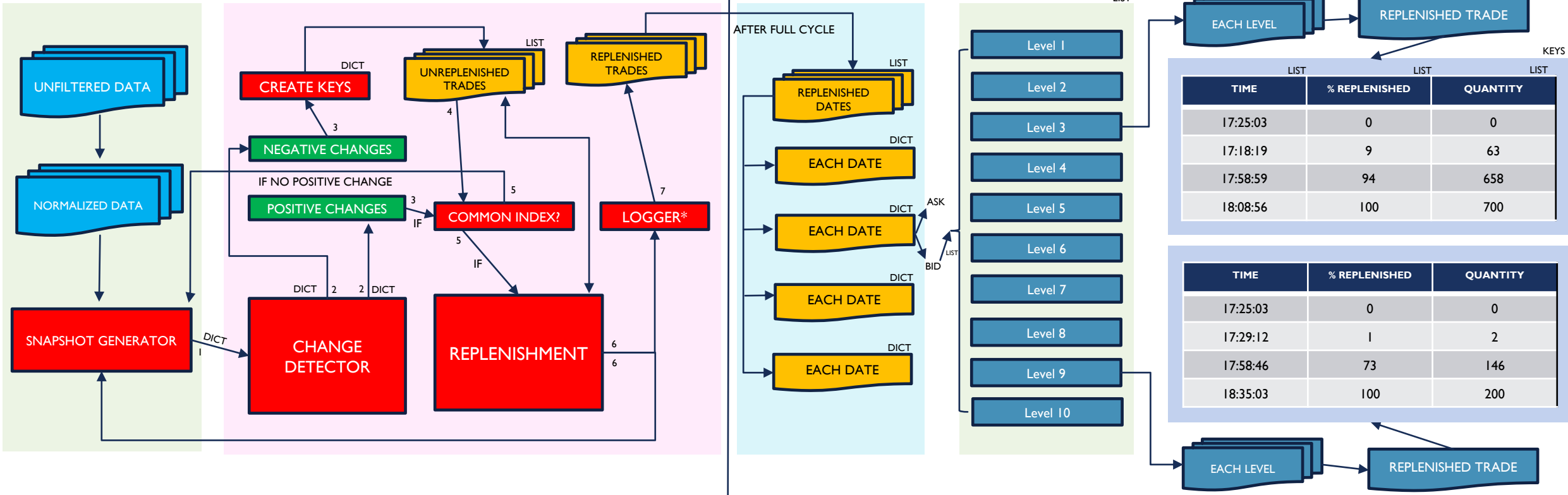


HISTORICAL REPLENISHMENT RATES

INPUT

MODELLING REPLENISHMENT

DATA STRUCTURE



RESULTS

Historical rates

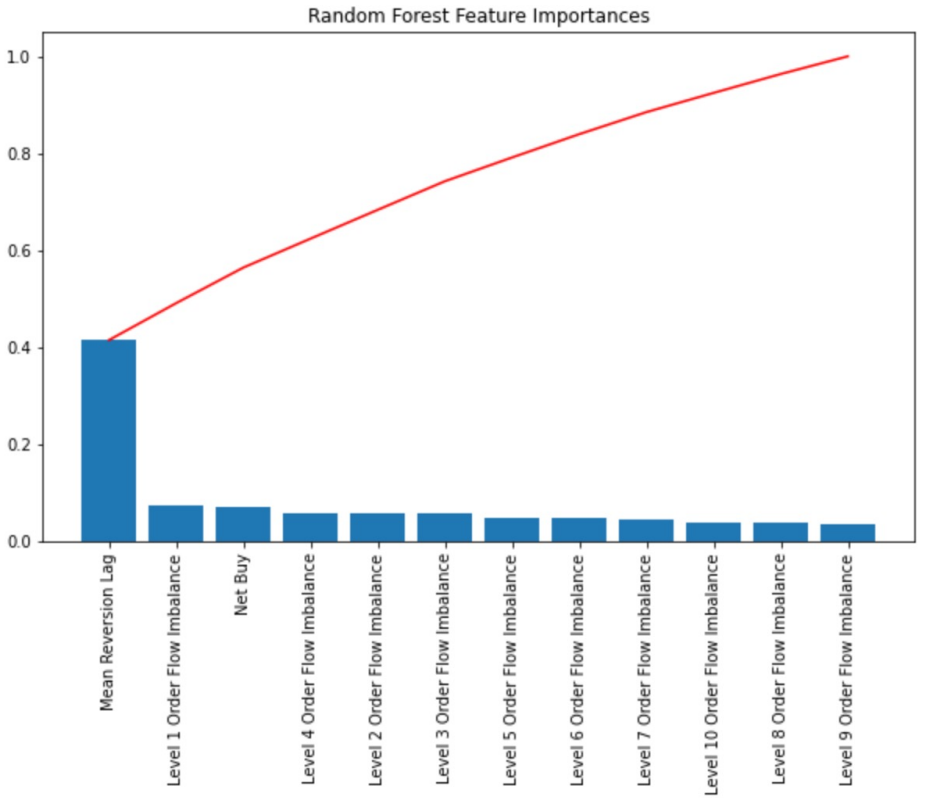
Level	Replenishment Quantity
1	0.385 ~ 0.4
2	0.296 ~ 0.3
3	0.198 ~ 0.2
4	0.293 ~ 0.3
5	0.093 ~ 0.1
6	0.154 ~ 0.2
7	0.131 ~ 0.1
8	0.054 ~ 0.1
9	0.026 ~ 0.1
10	0.012 ~ 0.1

Testing different models

Model	Revenue from Trade($)
Random Forest	6,53,250.00
Linear Regression	6,52,968.75
Ridge Regression	6,44,489.26

Lot quantity: 5,000

Feature Importance



KEY TAKEAWAYS

- Machine Learning techniques
- Limit Order Book Dynamics
- High Frequency Data Processing