Predicting Stock Volatility with High-Frequency Market Data

Shira Rozenthal & Alex Olteanu

Faculty Mentor: Professor Hamm

What is Volatility?

A stock's volatility is a measure of **the magnitude of fluctuations in its price, calculated as the annualized standard deviation of price.**

Volatility is the primary determinant of the likelihood an option on that stock will land in the money and, therefore, a key assumption in any pricing algorithm.

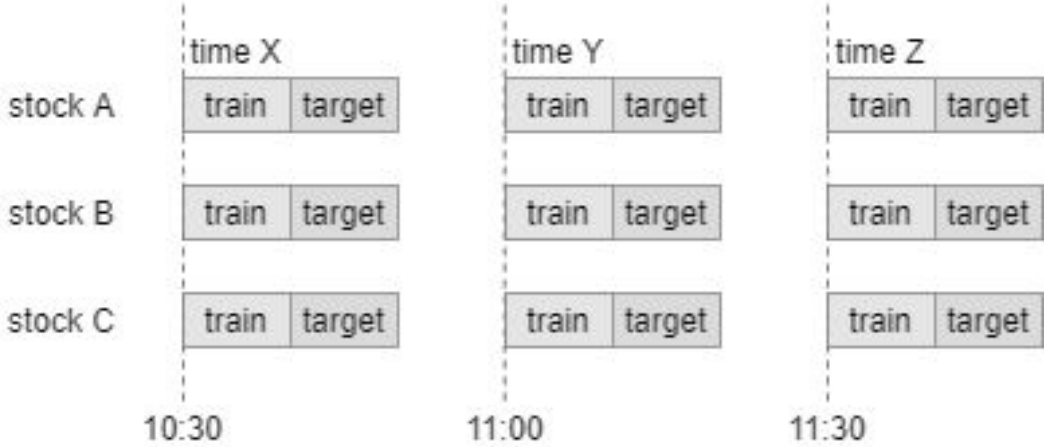
High-Frequency Data

For each stockID/timeID pairing, we have:

Order Book: Effectively representing interest in the market, this file contains the bid/ask price+size and time of a received market order.

Trades: Derived from the file above, this file contains the ticker, price, size, and time of those orders that actually executed.

Target: the realized volatility computed in the 10-minute window that follows the feature data.


time X
time Y
time Z
stock A
stock B
stock C
train
target
10:30
11:00
11:30

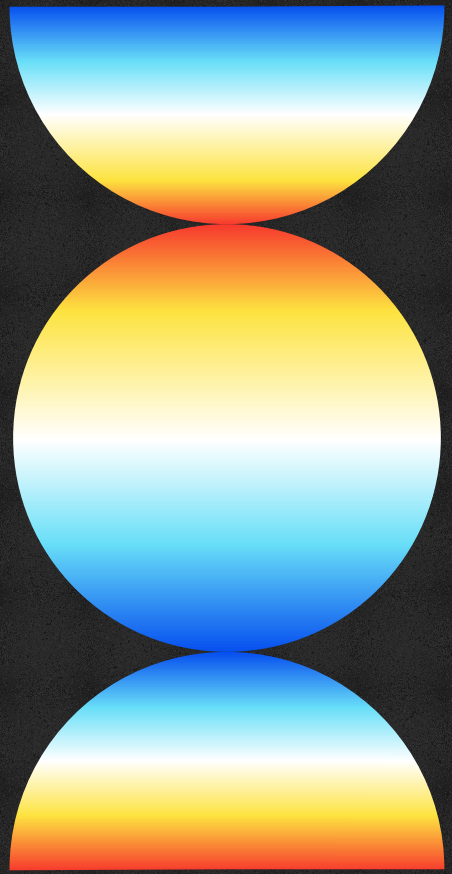
Project Overview

Working with HFT Data

- EDA / Access Functions
- Reverse-Engineering timeID

Feature Engineering & Modeling

- K-NN Market Conditions
- Running LightGBM and Optuna



Demo