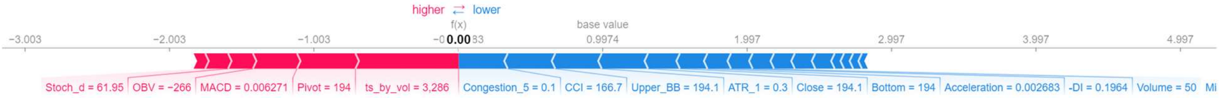
Blockhouse QR Test

Task 1 – Feature Extraction:

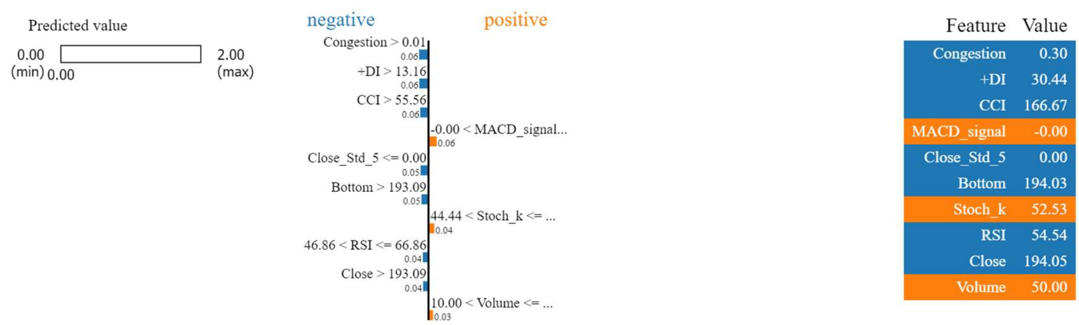
Looking at the data provided and the avoiding the use of technical indicators I came forward with 5 features that were extracted from the dataset.

1- Gap Theory: Even though the data does not account for pre-market moves and is not able to showcase the moment in the early phases of the trading day, the momentum of the stock in pre-market can be calculated through the difference in aftermarket prices being a great indicator of the momentum of the stock.
2- Acceleration: The acceleration of the stock can be calculated by first of differentiation of the moving average of the stock. I have chosen a moving average of 14 which accounts for the momentum of the stock in the last 14 transactions.
3- Congestion: The congestion of the stock calculated by the skew in the ask and bid price and creates a rolling window of 5 transactions capturing the skew in the week.
4- Execution time/Volume: This feature accounts for the effect of the volume of the order upon execution time for a transaction and how it affects the trading efficiency.
5- Pivots Top Bottom: The simple mathematical calculation of the High, Low and Close of a stock showcases the changes that can impact the trading charts. This also enables us to look out for the trend changes in the stock.
6- Open and Close (Std_5): This feature evaluates the standard deviation of the open and close prices of the stock over a period of 5 transactions.
I have added a sixth feature as the gap theory doesn't hold for this dataset.

I have implemented SHAP and lime to showcase the PPO model details and characteristics.



SHAP output



Lime output

Task 2 – Slippage Calculation:

After going through the data provided, I found these 2 papers:

a) Optimal execution strategies in limit order books with general shape functions - Aure´ lien Alfonsiy, Antje Fruthz and Alexander Schied
https://arxiv.org/abs/0708.1756

b) Quantifying Long-Term Market Impact - Campbell R. Harvey, Anthony Ledford, Emidio Sciulli, Philipp Ustinov, and Stefan Zohren
https://www.man.com/maninstitute/quantifying-long-term-market-impact

Summary:

a) This paper presents an optimal execution strategy in limit order books that with general shape. The general shape enables the model to build a model that provides a nonlinear price impact of the market orders. First, they model the dynamics of a limit order book by repeated market orders of a large trader like the Obizhaeva and Wang (2005) paper. They showcase 2 models, 1st with recovery of the LOB volume being exponential and 2nd with recovery of the extra spread being exponential. The cost minimization process is incorporated with the integral of the arrival price along the spread. The second problem that is tackled is the size of the child orders to reduce the impact of the order. For model 1, the unique solution of the equation is $F^{-1}\left(X_0 - N\xi_0^{(1)}(1 - e^{-\rho\tau})\right) = \frac{h_1(\xi_0^{(1)})}{1 - e^{-\rho\tau}}$, , where the intermediate orders are given by $\xi_1^{(1)} = \cdots = \xi_{N-1}^{(1)} = \xi_0^{(1)}(1 - e^{-\rho\tau})$, . For model 2, the unique solution becomes $F^{-1}\left(X_0 - N[\xi_0^{(2)} - F(e^{-\rho\tau}F^{-1}(\xi_0^{(2)}))]\right) = h_2(F^{-1}(\xi_0^{(2)}))$, with $\xi_1^{(2)} = \cdots = \xi_{N-1}^{(2)} = \xi_0^{(2)} - F(e^{-\rho\tau}F^{-1}(\xi_0^{(2)}))$, being the intermediate orders. There are different cases explained in the paper discussing the impact and change in the model from function being one to one function, function being one to one only if it's strictly increasing to asymptotic behavior of the optimal strategy when the number of trades tends to infinity. The paper also devles into block shaped LOBs and how the permanent impact of the trades does not need to considered to achieve a closed form solution.

b) The paper focused on implementing their Expected Future Flow Shortfall (EFFS) model to determine the market impact of metaorders. When trades are autocorrelated with one another their implied impact get complex to evaluate, this is where the EFFS model excels. The hidden slippage cost is factored in as the impact of various metaorder executions impart temporary as well as permanent changes. It leverages on Implementation Shortfall metric and the model is created with $(A_{i+1} – A_i)$ x $E[\sum Q_{i+j}]$. The metaorder is divided into multiple child orders with of various sizes instead of executing them all at once. There are multiple cases discussed in the paper ranging from child orders with relaxation time to stitching metaorders into composite metaorders. As the stock follows the usual random walk motion, a factor using a simple propagator model each trade is added with a 0.1 $\sigma\sqrt{|Q|}$. Based on the square root law the model keeps reducing the impact of the individual trade on the consecutive trades. The final approximated formula incorporating the stitched metaorders turns out to be:

$$EFF\ shortfall \approx \frac{\overline{Q \cdot (A_2 - A_1)}}{\bar{Q}} \cdot \overline{|Q|} \frac{2p-1}{2-2p}.$$

where p is the probability of the next trade being the same sign, A2 and A1 being the different arrival times and the average of the signed and absolute trade size, bar Q and bar |Q|.