SANA JAHAN

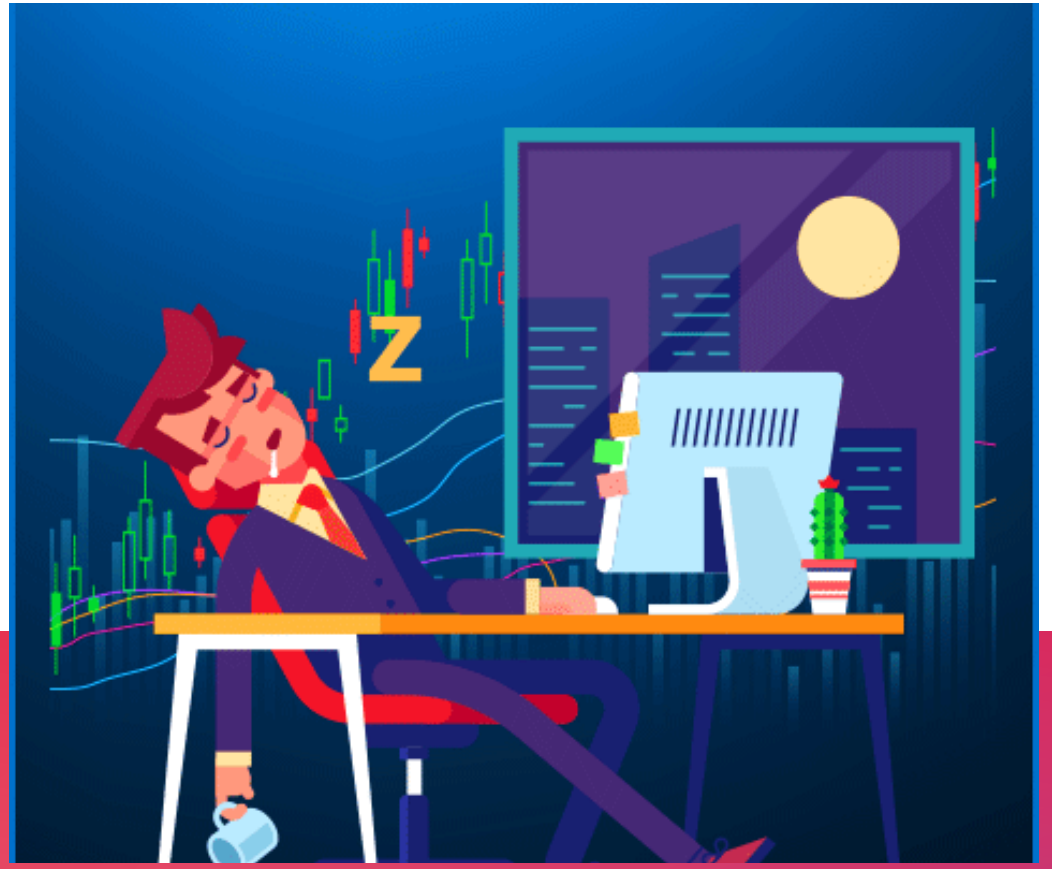
ALLISON BIEBER

DOMINIQUE SCHREINER

DEEP REINFORCEMENT LEARNING IN THE STOCK MARKET

THE AGENTS

Each trained bot will **buy, hold**, and **trade** stocks based on the deep reinforcement algorithms.

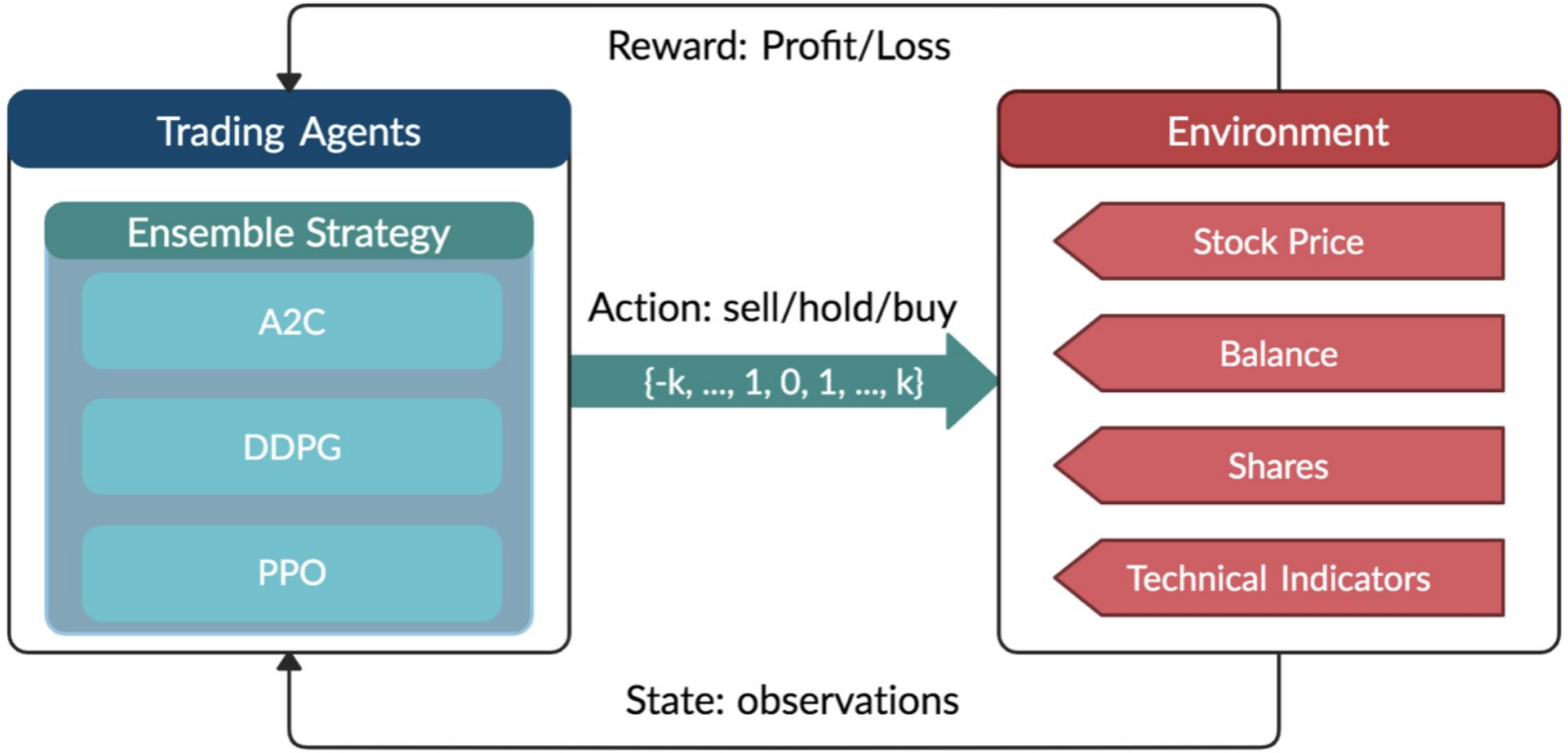


ENSEMBLE STRATETGY

STEP 1 RETAIN AGENTS

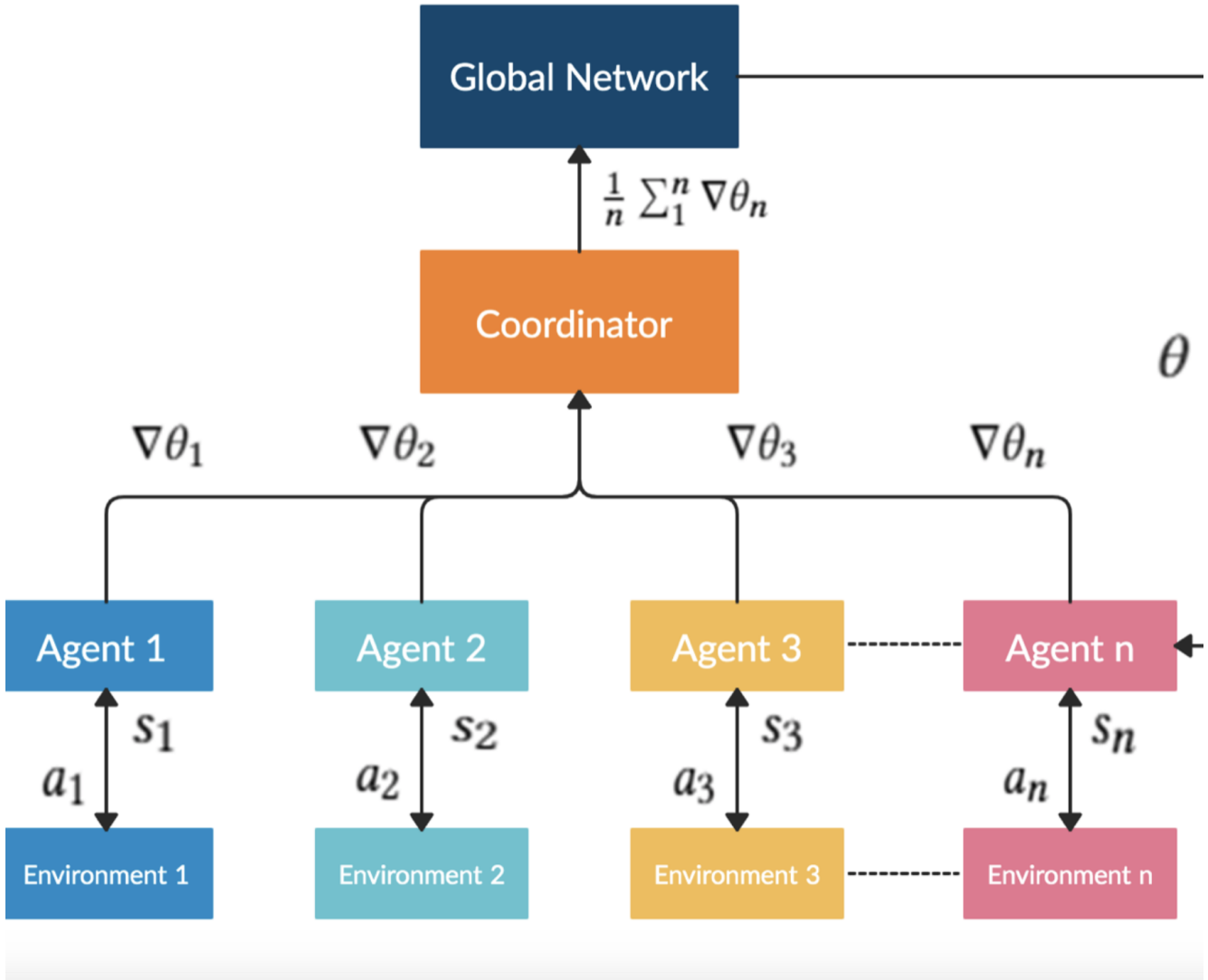
STEP 2 VALIDATE & ADJUST

STEP 3 SELECT MODEL



DEEP REINFORCEMENT LEARNING ALGORITHMS

1. Advantage Actor Critic (A2C)
2. Deep Deterministic Policy Gradient (DDPG)
3. Proximal Policy Optimization (PPO)



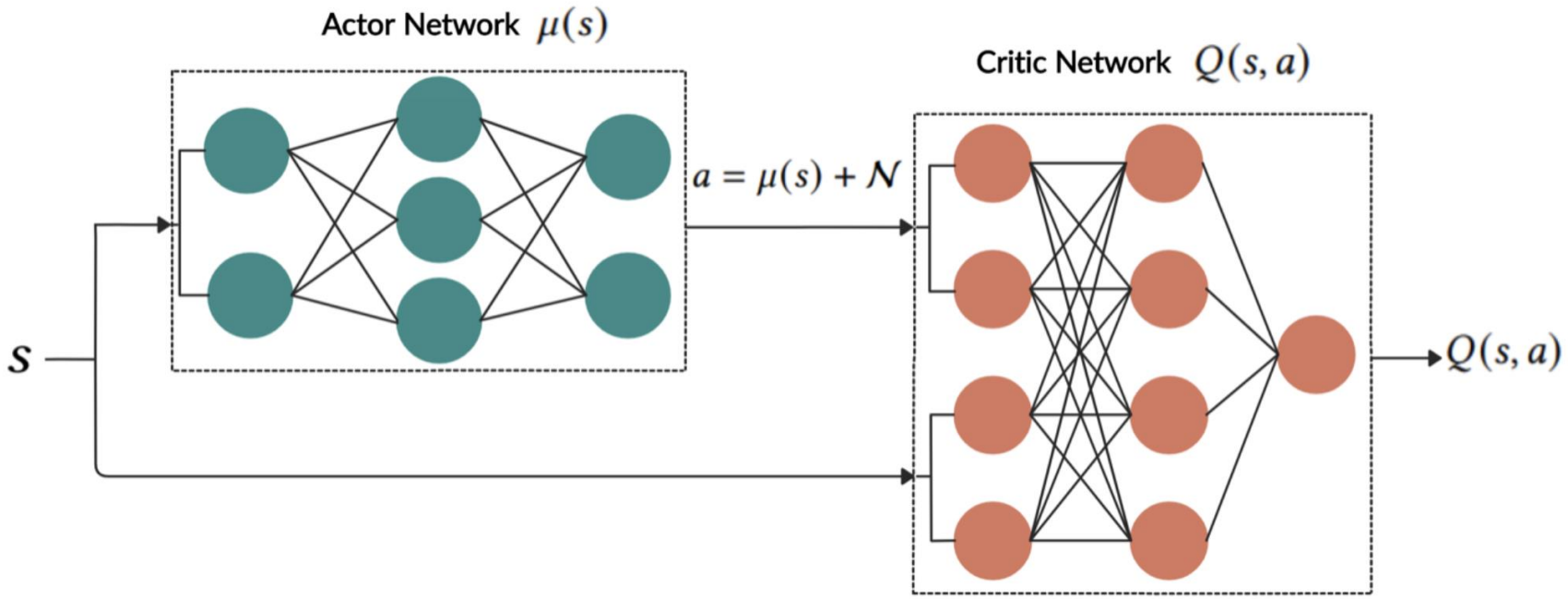
Advantage Actor Critic

- **a typical actor-critic algorithm**
- **uses copies of the same agent**
- **agent works independently**

A2C

DDPG

- combines the frameworks of both Q-learning and policy gradient
- uses neural networks as function approximators

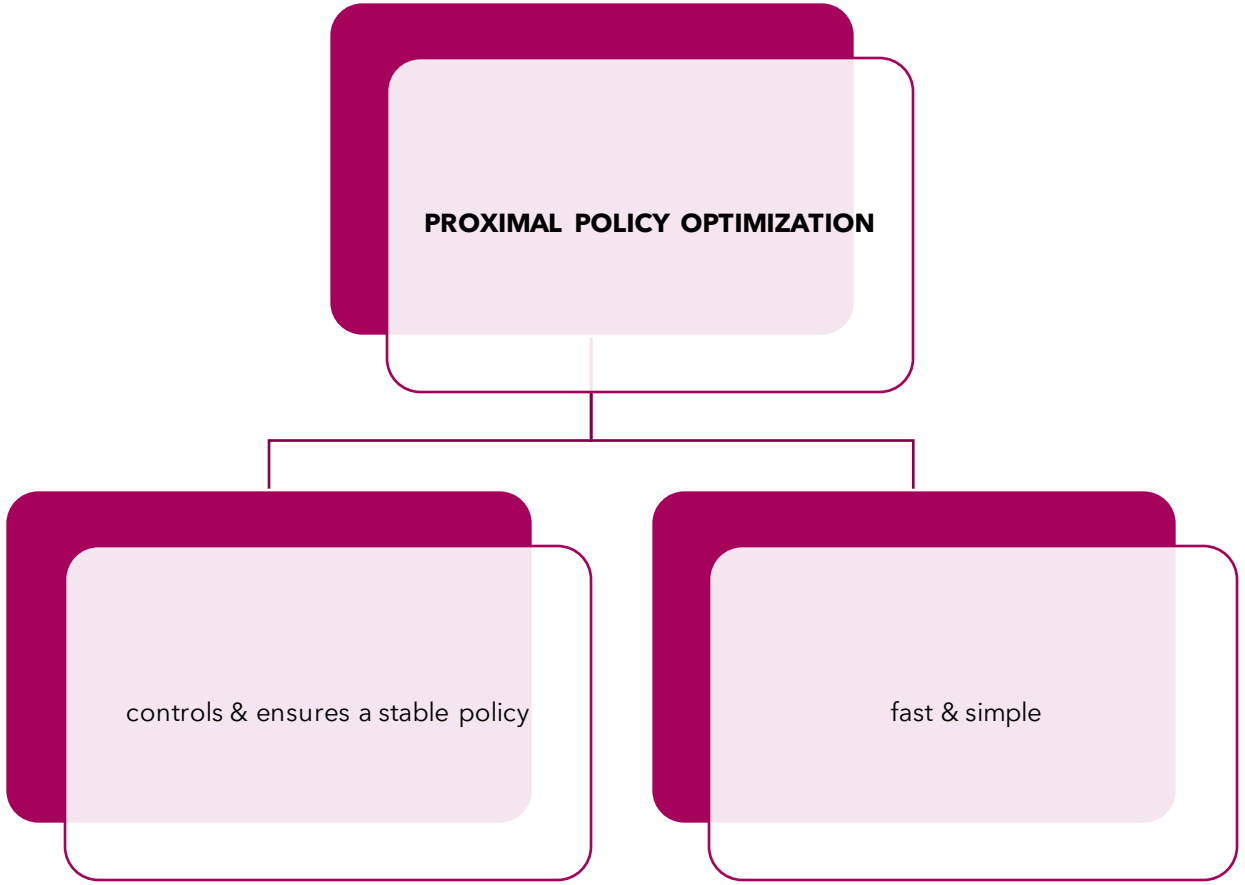


PPO

We explore and use PPO as a component in the ensemble method. PPO is

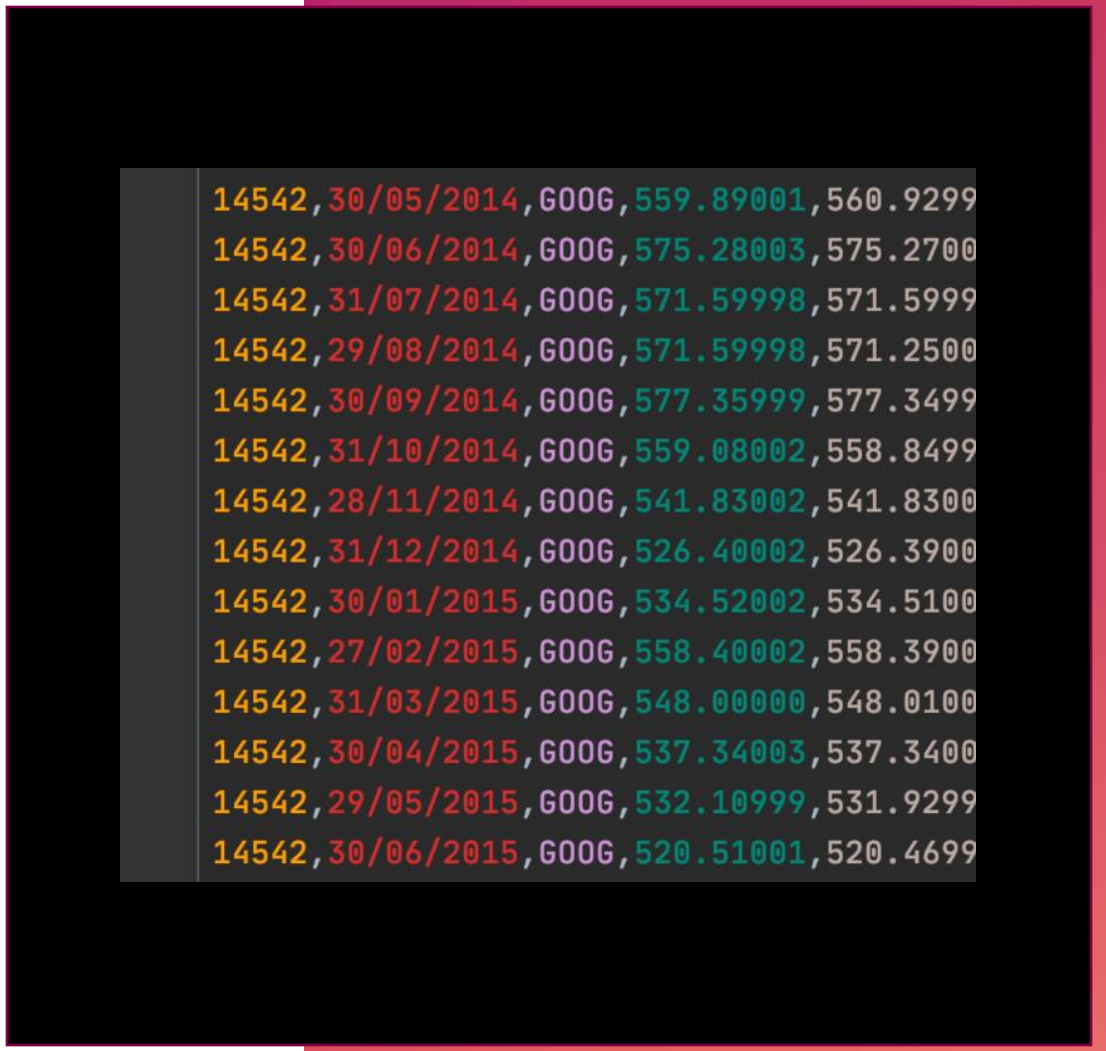
DEEP DETERMINISTIC POLICY GRADIENT

PPO


PROXIMAL POLICY OPTIMIZATION
controls & ensures a stable policy
fast & simple

STOCKS USED

- Dow Jones Industrial Average, top 30 stocks
  - January 1st, 2016
- Wharton Research Data Services
- Stock Data:
  - Open/Close, High/Low prices
  - Date
  - Adjustment Factor (ajexdi), Trading Volume
- Individual stocks were also analyzed
- Bot was run with and without stock indicators



RETURNS

Ran initial trials to validate the results of the source code program

Source code resulted in approximately 70% cumulative returns

Our trials:
- Results in slightly lower overall returns
-Ranged from 31% to 58%

```
==========================================
turbulence_threshold:  96.08032158358223
======Model training from:  20090000 to  20160104
======A2C Training========
Training time (A2C):  1.2218458652496338  minutes
======A2C Validation from:  20160104 to  20160405
A2C Sharpe Ratio:  0.19006740363250735
======PPO Training========
Training time (PPO):  4.645297348499298  minutes
======PPO Validation from:  20160104 to  20160405
PPO Sharpe Ratio:  0.18083600047873166
======DDPG Training========
Training time (DDPG):  0.7524083614349365  minutes
======DDPG Validation from:  20160104 to  20160405
======Trading from:  20160405 to  20160705
previous_total_asset:1057901.3978936314
end_total_asset:1029655.67744456
total_reward:-28245.720449071378
total_cost:  3134.832061731782
total trades:  1379
Sharpe:  -0.0949932917325246
==========================================
```

UPDATED RESULTS

Retained original Dow Jones 30 from 2016

Original data analyzed through May 2020

Updated data set extended the dataset through December 2020

Results: Maintained same total returns range (~30-50%)

RETURNS – SINGLE STOCK

Next step was to test the bot's utility on trading a single stock

A single stock can experience more volatility than a set of 30

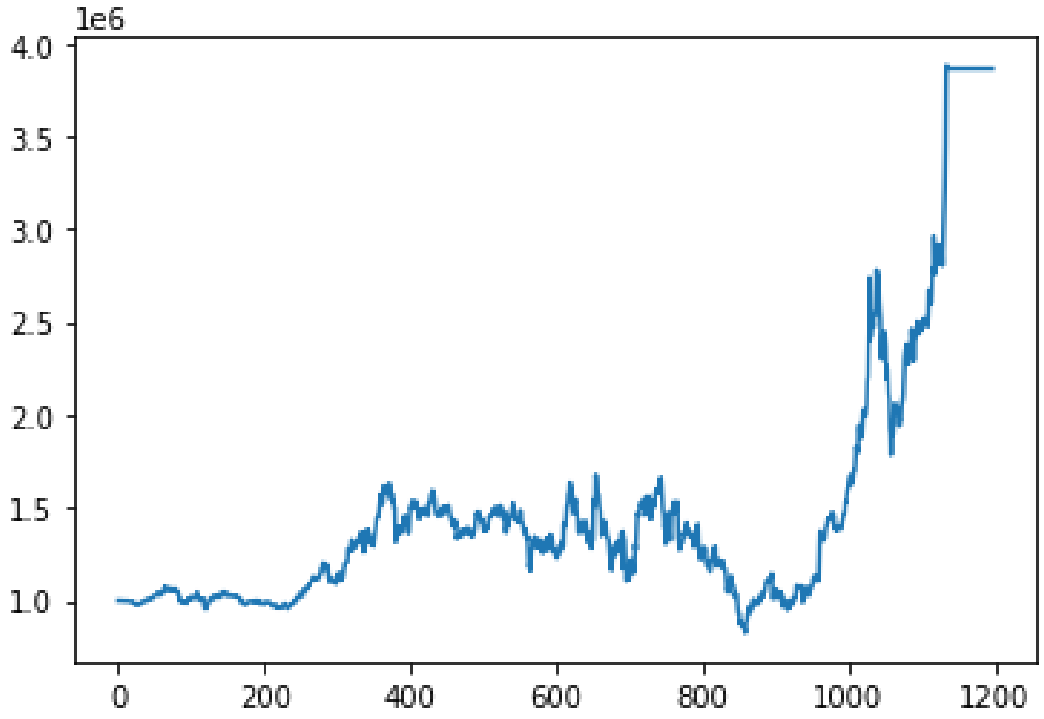
Method – reduced size of state space from a set of 181 to a set of 7

Chose a fairly volatile stock ,TSLA, that undergoes drastic increases

Showed returns of 300%+ which is consistent of the fast-growing environment

Limited succesful case

- Vulnerable to mass market movements
- Other stocks examined had lower average returns due to 2020 crash



RETURNS - INDICATORS

Various indicators used to guide the bot's decision making

RSI, MACD, CCI, ADX

Values of the indicators and the stock values form the bot's state space

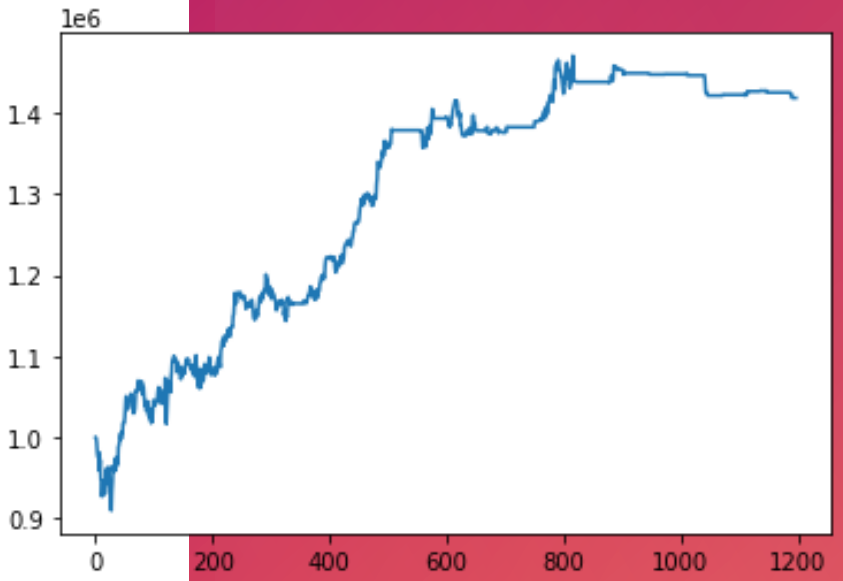
Task - Measure performance when the bot does not have the indicators

Method – Remove the indicators from the trading state space initialization

Results – Results remained in the same range, but on the low end from the trial runs of the bot with the indicators

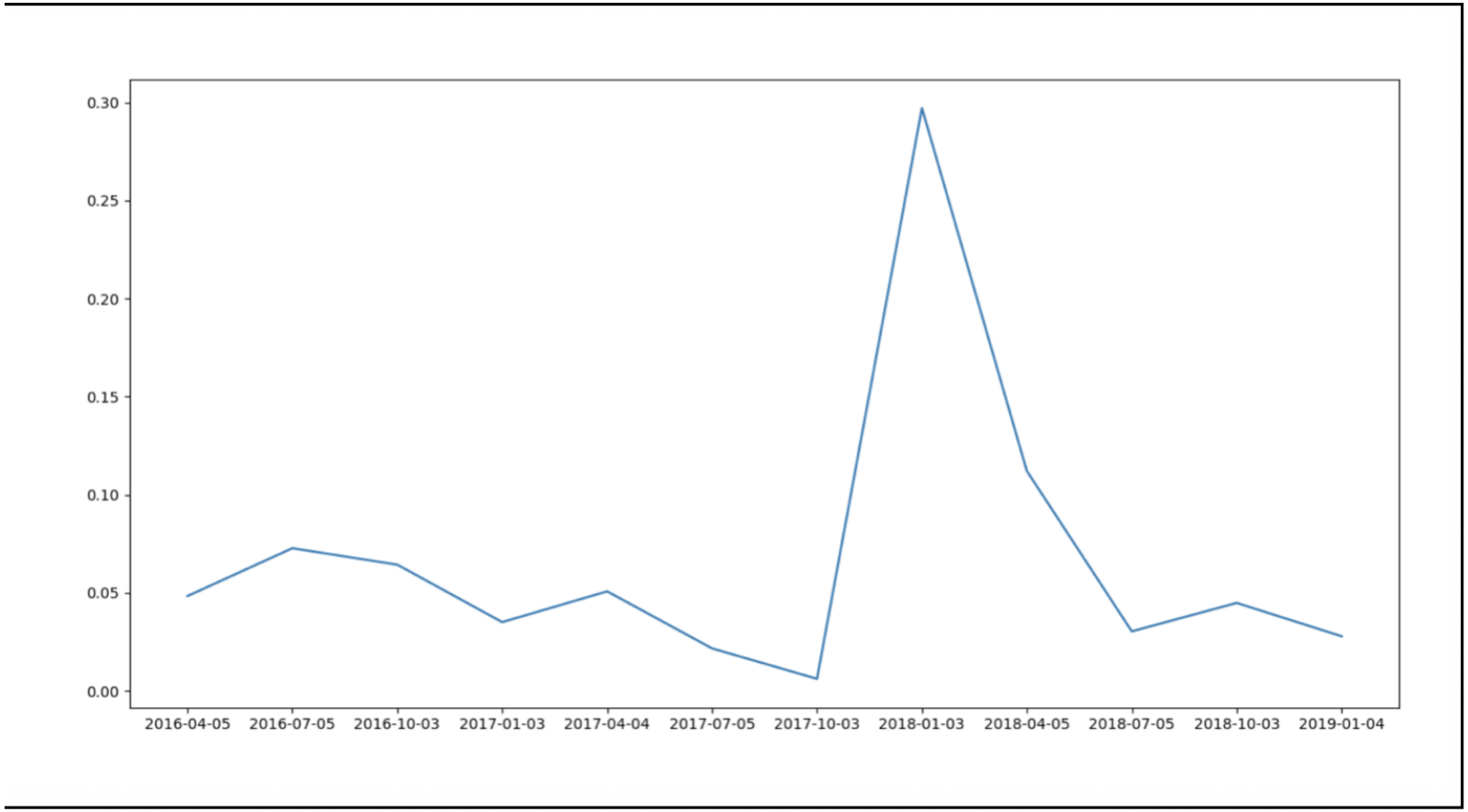
MARKET IMPACT

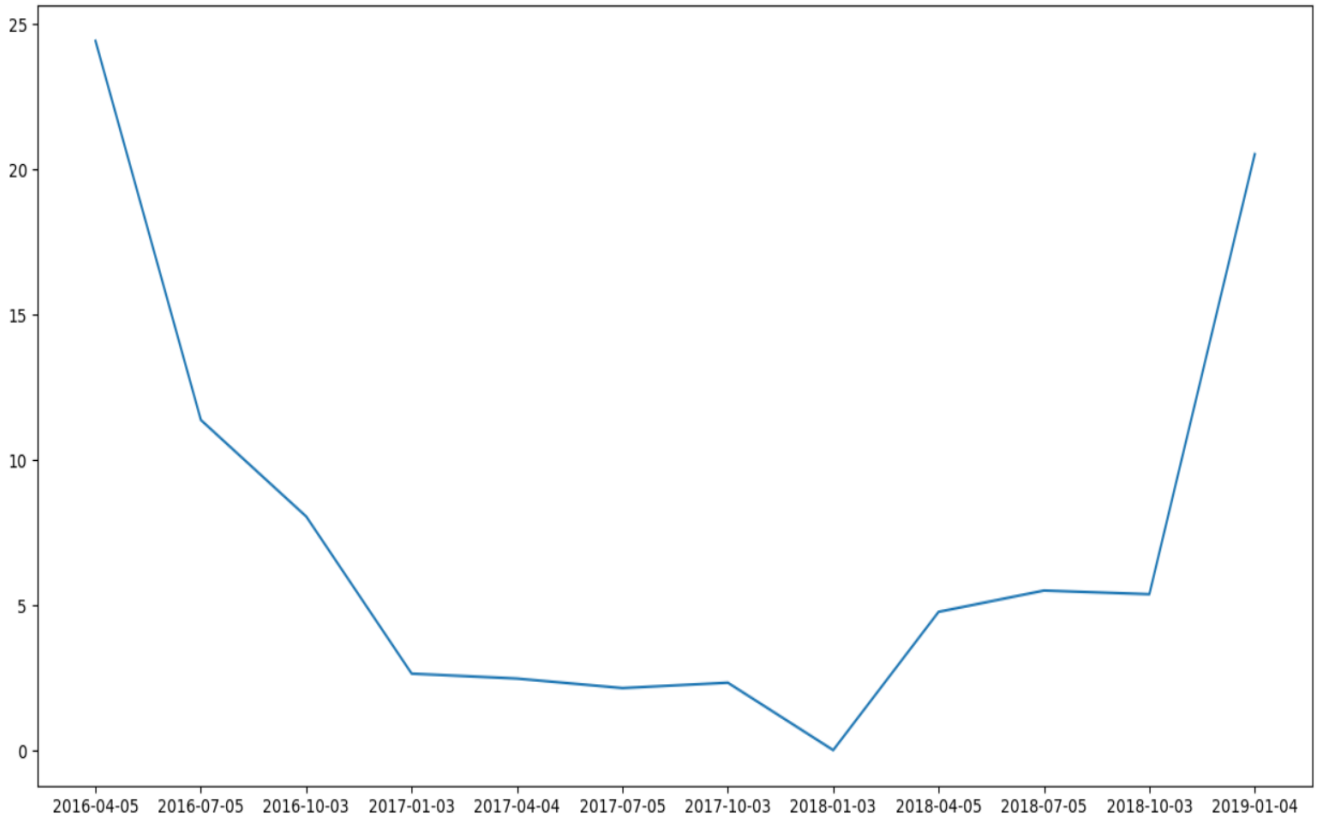
- Bot takes into consideration massive market changes
- Turbulence measures extreme changes
- Determines the risk aversion of the bot
- Able to maintain overall account value through market 2020 market crash



AGENT PERFORMANCE ON CURRENCY EXCHANGE RATE

- **Interestingly, currency of any value did not significantly impact the bot's performance to generate maximum reward**
- **Performance of Canadian currency**





PERFORMANCE ON CURRENCY OF HIGH VARIANCE (SOUTH AFRICAN CURRENCY)

TRUST REGION POLICY OPTIMIZATION ALGORITHM

One of our goals of this project was to add one more algorithm to the ensemble strategy. The algorithm we picked is the Trust Region Policy Optimization algorithm.

Trust Region Policy Optimization algorithm was used to trade several times and the overall results improved.

Overall performance of the agent in buying, selling and holding stocks was still impressive with a return of 20-50%

FUTURE SCOPE AND CONCLUSION

This is a very promising project in the financial domain.

The advantage of this project was the use of DRL and the concept of ensemble strategy

The ensemble strategy provides the flexibility to add new algorithms

We tried to add new features like currency rate to analysis the performance of the agent. Similarly, other features relevant to stocks can be used by the agent.

THANK YOU!
QUESTIONS?