

Financial Context: Reinforcement Learning for Making Prediction on Financial Time Series & Portfolio Optimization Using Python

Serge Yumbi

University of Derby

College of Engineering & Technology

Module: Research Project

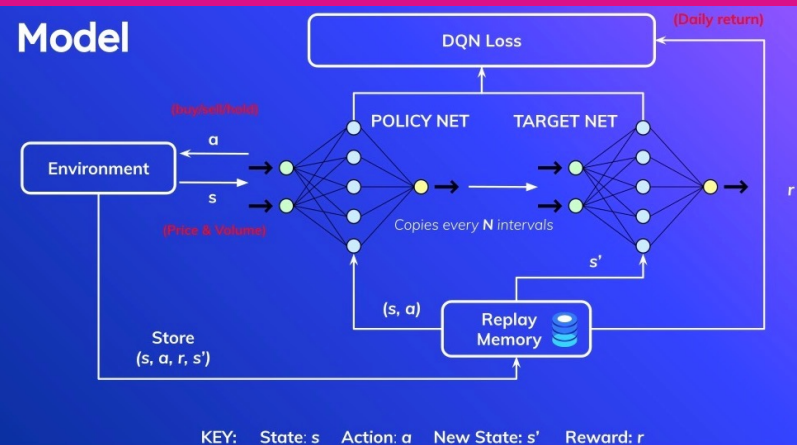
5th March 2022 | University of Derby

RATIONALE:

- *Increase Adoption of DRL in Finance*
- *Trading Dynamics in Stock Market*
- *DRL expert is in high demand*
- *DRL is the hottest topic of in ML applications & ML Research*
- *Most scenarios in Finance lend themselves to MDP, DRL or MARL*

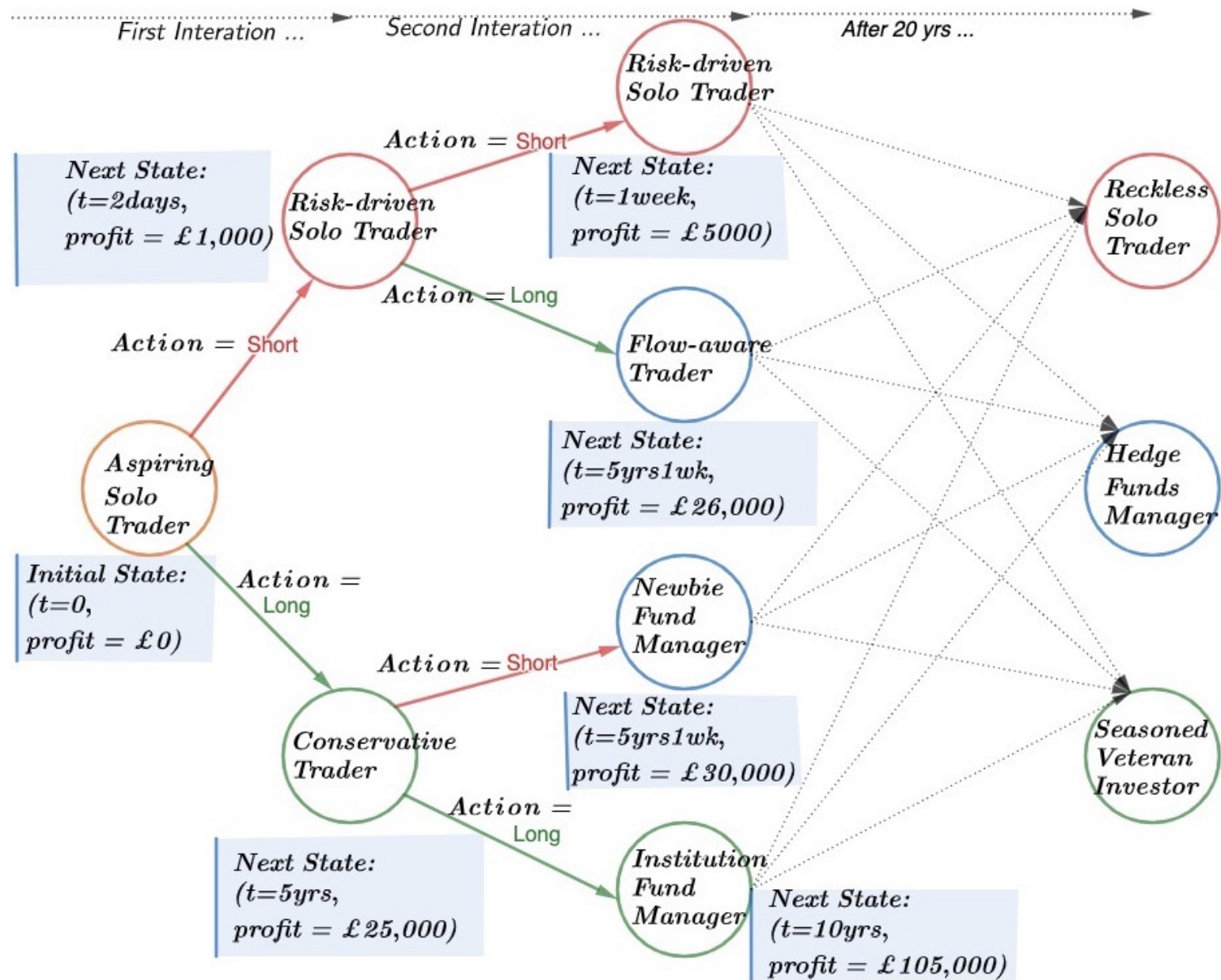


OBJECTIVES



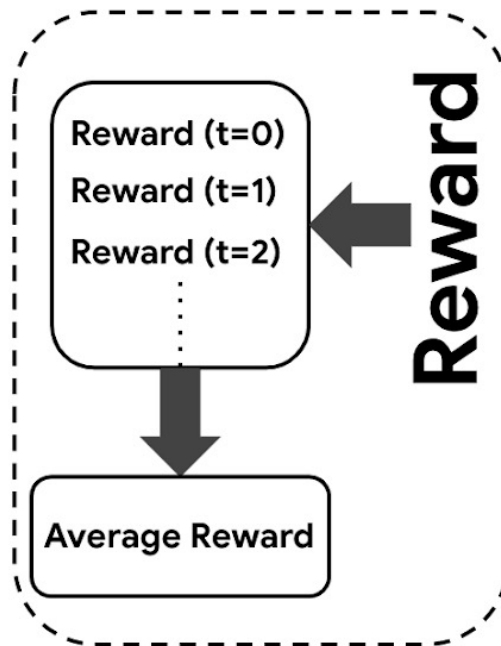
- ❖ Train an agent to trade a particular stock whilst hedging a European call option.
- ❖ Through RL mechanisms, train agents to buy or sell with reasonable accuracy.
- ❖ Harness reinforcement learning with TensorFlow and Keras using Python in order to buy or sell individual stocks using DRL agents.
- ❖ Gain a greater understanding of Actor-Critic methods, Policy Gradient method, Markov Decision process and Dynamic Programming as well as the state-action-reward pattern.
- ❖ Choose and optimize a Q-Network learning parameters and fine-tune its performance.

MDP: Intuitive Understanding

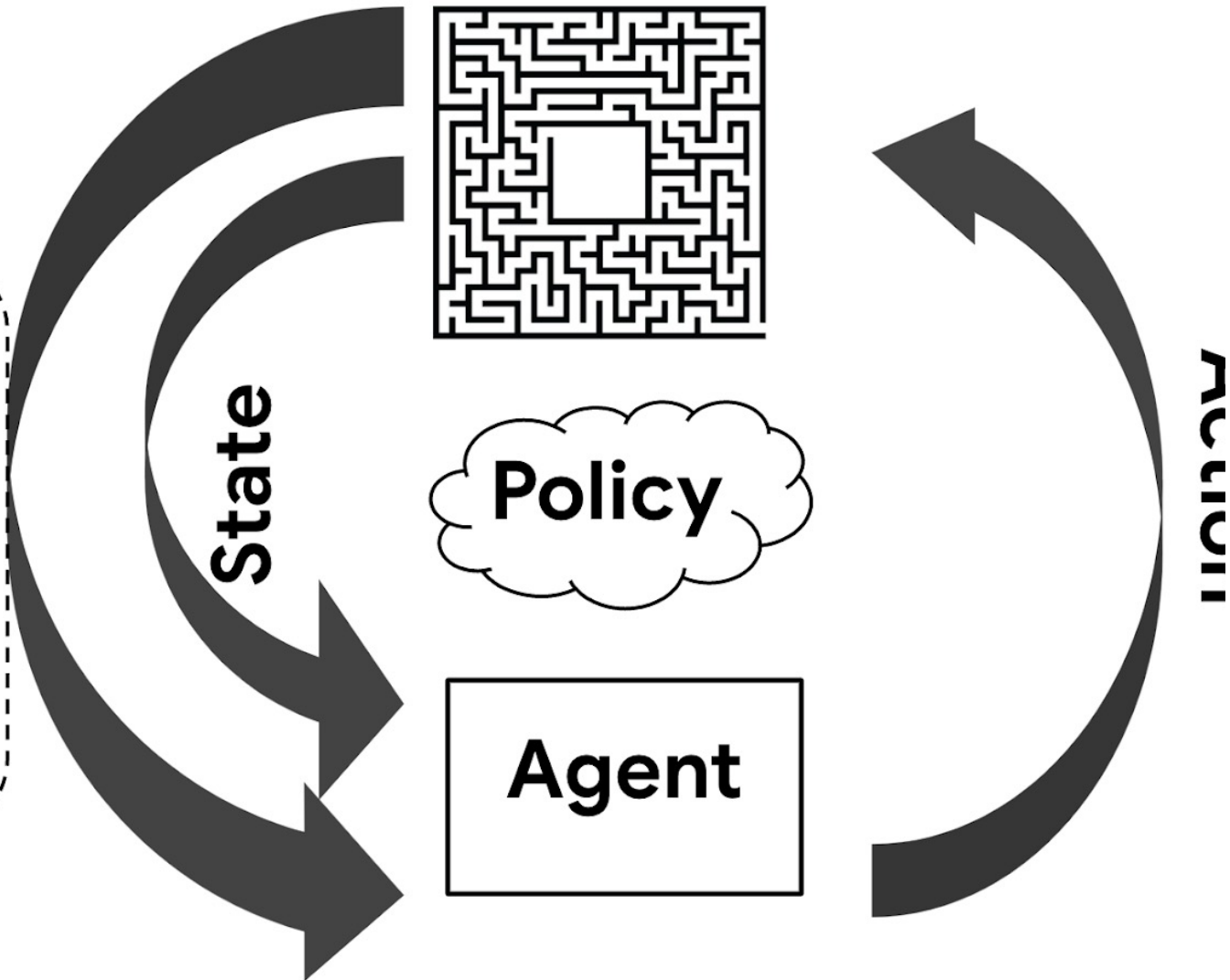


MDP: Intuitive Understanding

Policy Evaluation



Environment



RESEARCH METHODOLOGY

Agility is at the heart of every workflow...

Data Prep & Prelim Feature Engineering & ML models

- ✓ SP500 Stock Data
- ✓ Prelim Analysis
- ✓ Feature Engineering
- ✓ Least Squares
- ✓ Arima & Garch & LSTM
- ✓ Kalman Filters
- ✓ Option Pricing
- ✓ Portfolio Optimisation
- ✓ Heavy Tail & Copulas

Main Focus

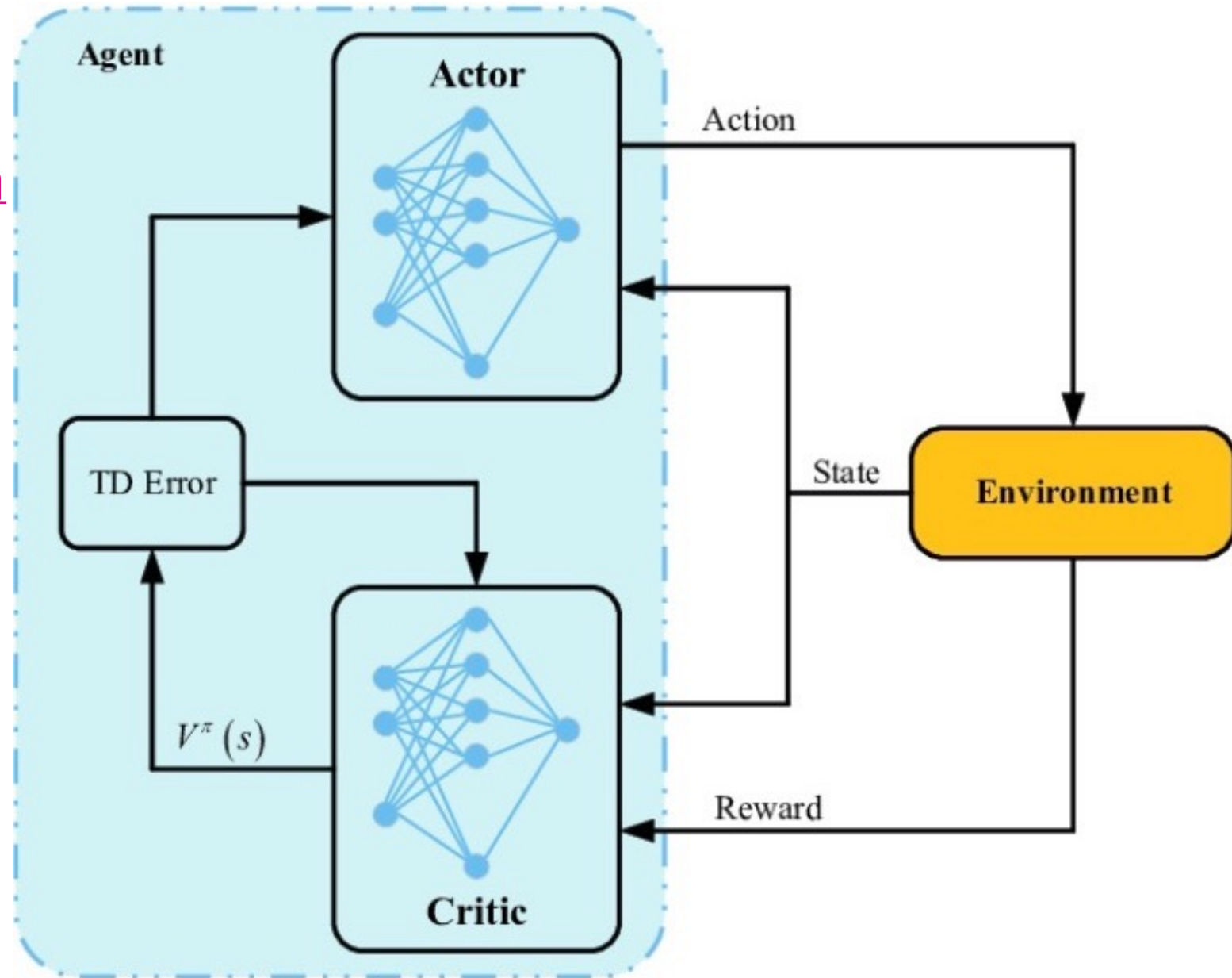
SCENARIOS

- ✓ A2C Algorithm simple
- ✓ A2C Algorithm (custom)
- ✓ DDPG Algorithm (hedging)
- ✓ DQN Algorithm
- ✓ A2C, PPO & DDPG
(multi-Asset Allocation)
- ✓ Multi-Agent
(stock liquidation)

Evaluation & Tuning

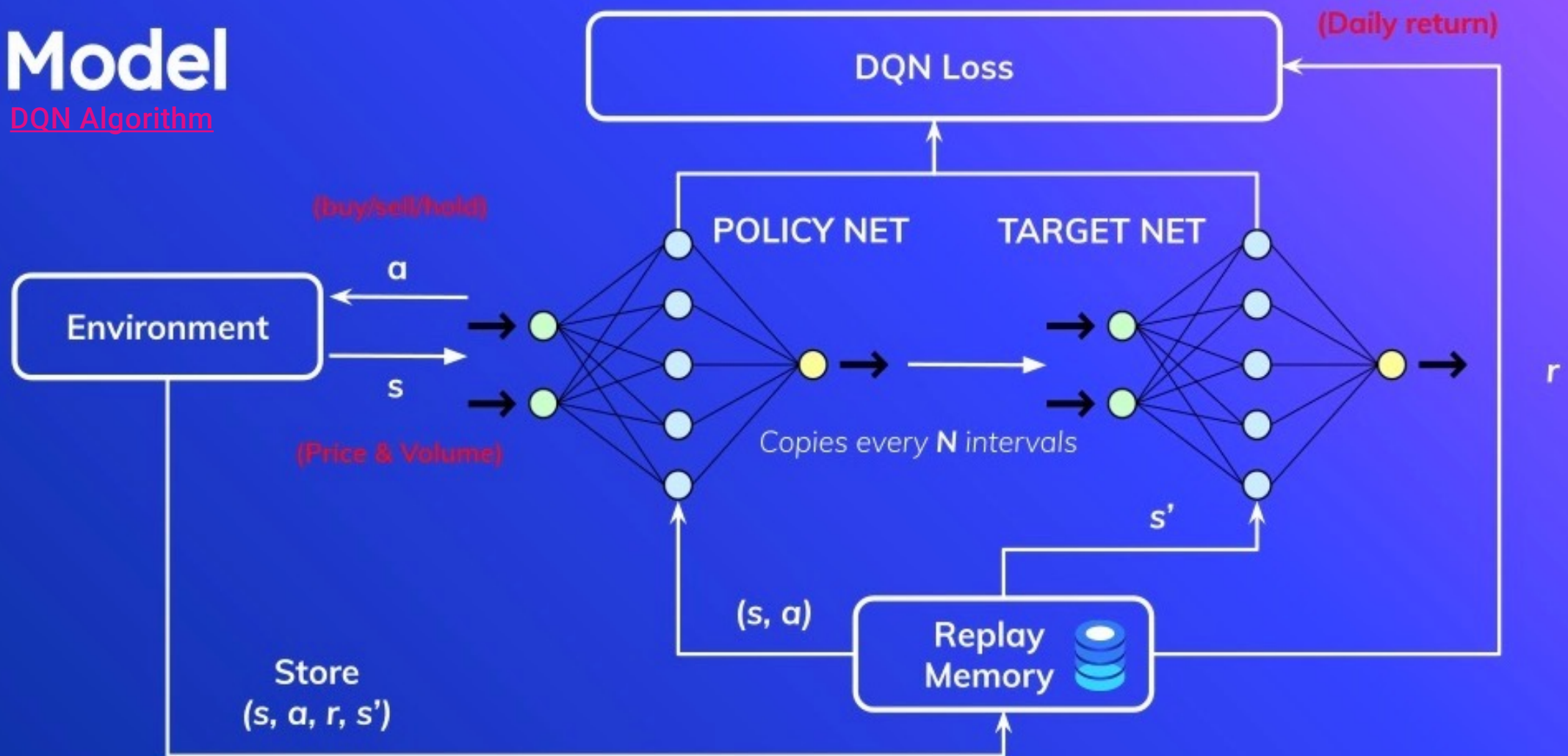
- ✓ Evaluation
- ✓ Tuning
- ✓ Positives
- ✓ Challenges

DRL: Intuitive Understanding Of A2C Algorithm



Model

DQN Algorithm

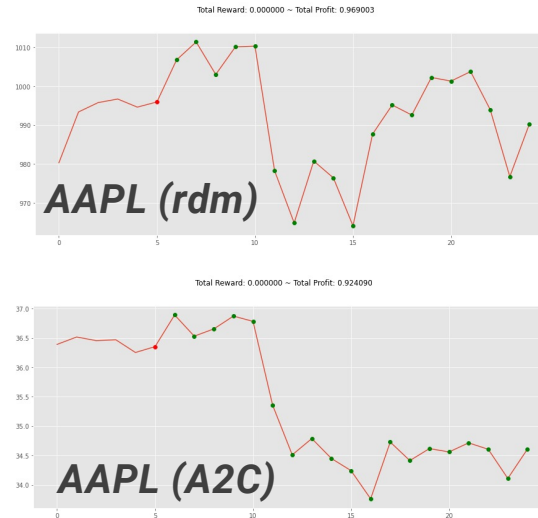


KEY: State: s Action: a New State: s' Reward: r

Scenario 1: A2C Algorithm

- *Ad Hoc results with random agent*
- *Poor performance on evaluation phase*
- *Insufficient features.*
- *Short training time.*

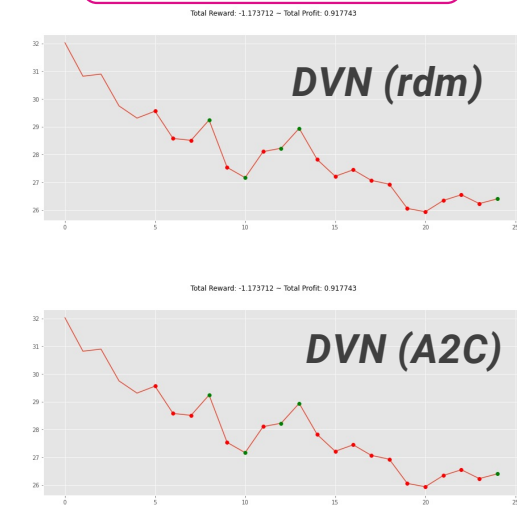
10 Large Cap Companies



Tickers	Var	fps	# Updates	Policy Entropy	Total Profit	Value Loss	Duration (in mins)
SP500	-0.068	362	200000	0.671	0.9518244	22.8	45
AAPL	0.136	409	200000	0.214	0.9240903	0.0267	40
FB	-0.237	446	200000	0.64	0.9826773	0.0242	37
MSFT	-0.018	406	200000	0.69	0.8951777	0.0666	37
GOOGL	0.053	405	200000	0.577	1.0	3.91	41
GOOG	0.00422	403	200000	0.644	0.9134023	4.03	42
AMZN	-0.000466	335	200000	0.652	0.9518244	4.31	49
TSLA	-0.128	431	200000	0.692	1.0	0.018	38
BAC	-61.4	399	200000	0.297	1.0262192	0.0709	41
JPM	-9.06e ⁻⁰⁶	362	200000	0.693	0.123	0.9932655	45

Table 5.6: Top 10 Companies (Experiment: Key Metrics)

Big Winners 2021



Tickers	Var	fps	# Updates	Policy Entropy	Total Profit	Value Loss	Duration (in mins)
DVN	-0.204	391	200000	0.0695	0.9177430	0.187	42
MRO	-5.07	462	200000	0.000232	0.8738668	3.05e ⁻⁰⁷	36
MRNA	0.782	371	200000	0.106	0.9265450	0.0191	44
FTNT	0.00695	366	200000	0.69	0.9585552	0.0639	45
SBNY	-0.772	394	200000	0.000397	0.9630949	4.06e ⁻¹⁰	42
F	-15.7	374	200000	0.541	0.9498389	5.06e ⁻⁰⁵	44
BBWI	-25.9	561	200000	3.74e ⁻⁰⁵	1.0190431	8.6e ⁻⁰⁸	40
NVDA	0.363	431	200000	0.0127	0.8954711	0.0221	38
FANG	-12.9	342	200000	1.48e ⁻⁰⁶	0.9288472	1.78e ⁻⁰⁸	48
NUE	0.447	415	200000	0.00342	0.9826887	0.00113	40

Table 5.7: Big Winners 2021 (Experiment: Key Metrics)

Scenario 2: A2C Algorithm with custom model & Indicators

- *Ad Hoc results with random agent*
- *Good performance on evaluation phase*
- *Perhaps even more features?*
- *Longer training time.*

10 Large Cap Companies



Figure 5.29: Predictions: A2C with custom indicators (AAPL)

Tickers	Var	fps	Total Reward	Policy Entropy	Total Profit	Value Loss	Duration (in mins)
SP500	1	327	0	0.00032	1.141261	0.000499	50
AAPL	1	372	0.929888	0.000776	1.072712	0.000333	44
FB	1	385	7.679962	0.00363	0.795125	0.0046	43
MSFT	1	448	15.722967	0.0179	0.707773	5.9×10^{-6}	37
GOOGL	1	431	181.820007	0.0182	0.863374	0.000708	34
GOOG	1	388	225.887023	0.00238	0.8040905	0.00329	42
AMZN	1	468	254.239	0.00495	0.83714	0.000382	35
TSLA	1	434	3.919991	0.128	0.878455	0.00424	38
BAC	1	417	8.463927	0.00101	1.378979	1.57×10^{-6}	39
JPM	1	423	27.241544	0.693	0.928567	6.03×10^{-8}	38

Table 5.8: Top 10 Companies (Scenario 2: Key Metrics)

Big Winners 2021



Figure 5.28: Predictions: A2C with custom indicators (NUE)

Tickers	Var	fps	Total Reward	Policy Entropy	Total Profit	Value Loss	Duration (in mins)
DVN	1	391	6.740226	0.0648	0.831339	1.11×10^{-6}	34
MRO	1	554	4.583833	0.0944	0.896599	3.18×10^{-5}	30
MRNA	1	368	-0.563000	0.148	0.665498	5.92×10^{-4}	45
FTNT	0.995	505	5.579998	0.115	0.802322	4.34×10^{-2}	33
SBNY	1	390	0.000000	7.36×10^{-5}	1.041657	4.53×10^{-6}	40
F	1	387	-0.8595	0.143	0.588174	1.58×10^{-6}	42
BBWI	1	418	3.060227	0.000234	0.836737	7.09×10^{-10}	39
NVDA	1	314	0.818256	0.00814	0.873397	7.99×10^{-4}	53
FANG	1	436	20.993150	0.137	0.943501	1.01×10^{-5}	38
NUE	1	404	0.000000	0.00592	1.129064	7.28×10^{-6}	41

Table 5.9: Big Winners 2021 (Scenario 2: Key Metrics)

Scenario 3: DDPG Algorithm For Hedging EU option

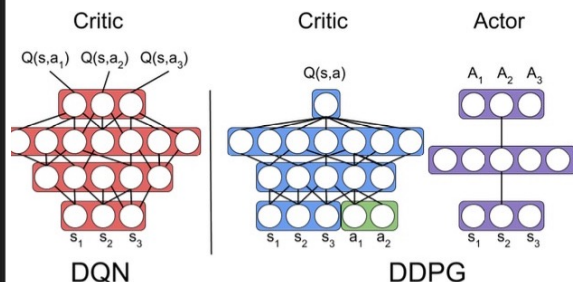
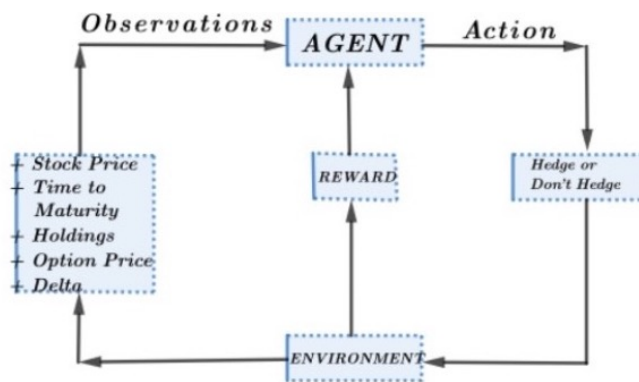


Figure 5.30: DDPG(Deterministic Policy Gradient)

	BSM	RL
Average Hedge Cost (% of Option Price)	91.259	53.425
STD Hedge Cost (% of Option Price)	35.712	61.741

Table 5.10: Comparison: BS vs RL

INTUITIVE UNDERSTANDING



TRAINING & TESTING

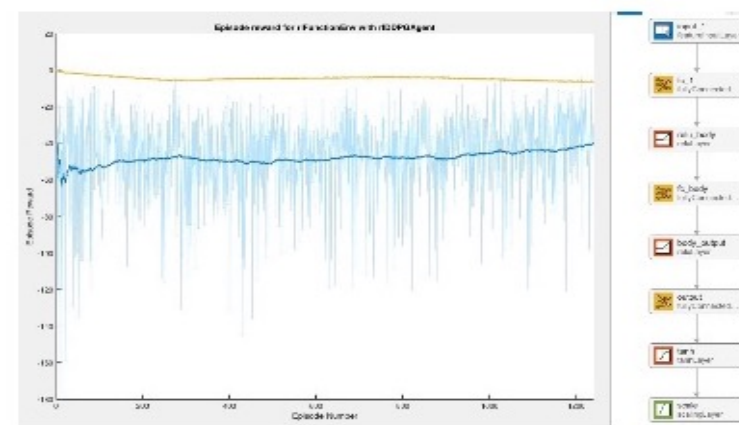


Figure 5.32: Episodic rewards

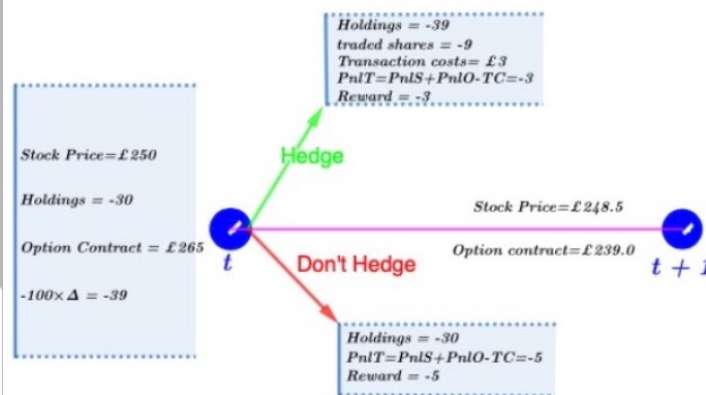


Figure 5.31: Environment - Agent interactions

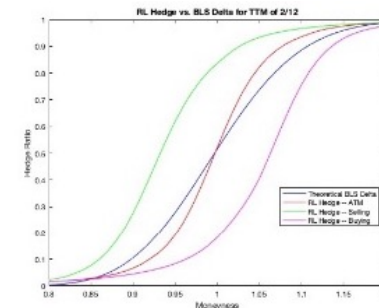
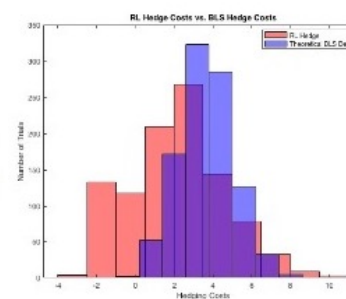


Figure 5.33: Moneyness & Hedging Ratio

Scenario 4: DQN Algorithm for Trading simulations

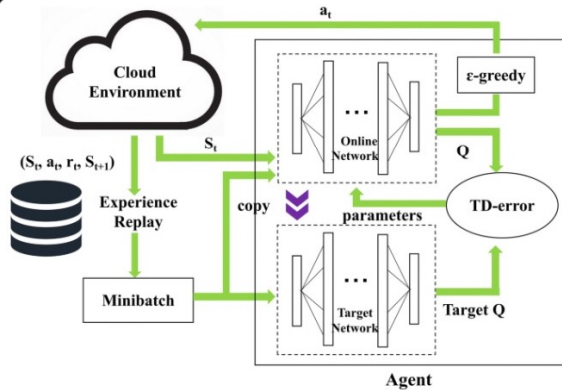


Figure 5.34: Example of DQN Architecture (Peng, Z et al. 2020)

Date	03/01/2012	04/01/2012
# SPY shares	0	784
Cash	\$100000	\$30
NAV	\$100000	\$100116.80
Reward	-	\$146.80

Table 5.11: Checking the environment (one step)

Layer (type)	Output shape	# of params
dense (Dense)	Multiple	14,100
dense 1 (Dense)	Multiple	5050
dense 2 (Dense)	Multiple	102
model: sequential		
Total params: 19,252		
Trainable params: 19,252		
non-trainable params: 0		

Table 5.12: Summary (Q-Network)

DQN Agent with random behaviour

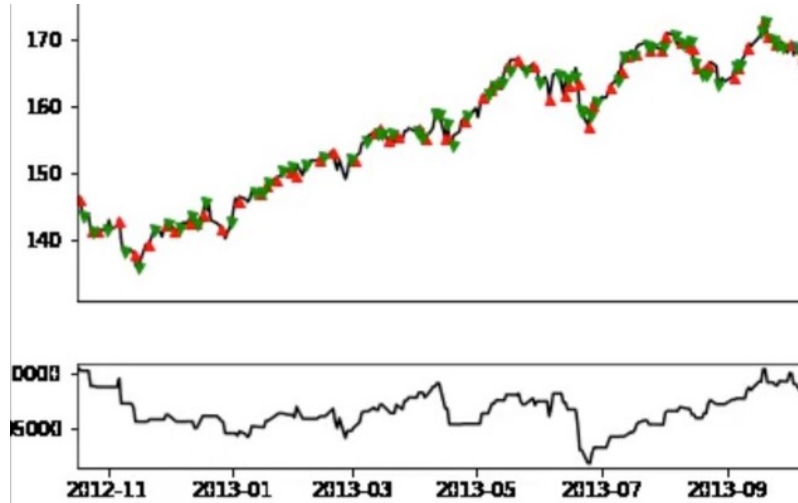


Figure 5.35: Example of random trading simulation

Ticker	Annual Return (in %)			Sharpe Ratio			Tail Ratio		
	Rdm	Agent	BM	Rdm	Agent	BM	Rdm	Agent	BM
SPY	4.12	8.03	12.17	0.491	0.860	0.955	1.169	1.135	0.952
AAPL	3.64	15.66	21.28	0.281	0.839	0.881	1.008	1.065	1.014
MSFT	7.26	11.23	24.60	0.533	1.055	1.079	1.188	12.371	1.092
GOOGL	3.26	19.27	19.27	0.281	0.876	0.876	1.028	0.963	0.963
GOOG	11.50	19.34	19.37	0.707	0.875	0.875	1.138	0.982	0.983
FB	2.04	8.94	22.89	0.201	0.647	0.818	0.935	11.751	0.993
AMZN	9.38	33.56	33.39	0.529	1.122	1.117	1.129	1.058	1.057
TSLA	24.16	31.38	53.17	0.761	0.946	1.063	1.093	1.146	1.067
JPM	4.65	17.82	18.89	0.375	0.931	0.912	1.088	1.080	1.070
BAC	10.18	10.88	25.29	0.585	0.679	0.968	1.145	1.261	1.123

Table 5.14: DQN Agent Performance (Top Cap 2021 Summary)

Note: Rdm → random trader and BM → Benchmark

DQN Agent with custom model & Indicators



Figure 5.36: Example of DQN agent trading (SPY case)

	random	DQN Agent	Benchmark
Annual Return	4.12%	8.02%	12.17%
Cumulative Returns	37.77%	84.47%	148.73%
Annual Volatility	0.091	0.095	0.129
Sharpe Ratio	0.491	0.860	0.955
Calmar Ratio	0.313	0.829	0.603
Stability	0.778	0.840	0.951
Max Drawdown	-0.132	-0.097	-0.202
Omega Ratio	1.134	1.238	1.186
Sortino Ratio	0.684	1.274	1.345
Skew	-0.620	0.121	-0.402
Kurtosis	8.331	8.465	3.279
Tail Ratio	1.169	1.135	0.952
Daily VaR	-0.01	-0.012	-0.016

Table 5.13: Summary (DQN Agent Performance for SPY index)

Ticker	Annual Return (in %)			Sharpe Ratio			Tail Ratio		
	Rdm	Agent	BM	Rdm	Agent	BM	Rdm	Agent	BM
DVN	-7.84	0.0016	-11.99	-0.159	0.156	-0.145	0.907	0.998	0.945
MRO	-2.65	-2.52	-10.34	0.054	0.118	-0.063	0.995	0.957	0.947
MRNA	48.71	-40.90	-11.10	1.024	-1.593	0.258	1.235	0.460	1.030
FTNT	15.79	22.57	22.58	0.720	0.755	0.755	1.070	0.973	0.973
SBNY	1.05	3.58	10.52	0.147	0.269	0.530	0.940	0.997	1.003
F	-2.05	3.72	-2.21	-0.035	0.276	0.0290	0.958	1.015	0.959
BBWI	-0.27	-6.09	-9.03	0.107	-0.035	-0.109	1.071	1.001	0.966
NVDA	14.00	42.91	41.79	0.627	1.141	1.115	1.144	1.198	1.197
FANG	14.00	42.91	41.79	0.627	1.141	1.115	1.144	1.198	1.197
NUE	-2.10	2.68	4.65	-0.028	0.230	0.306	1.020	1.061	1.055

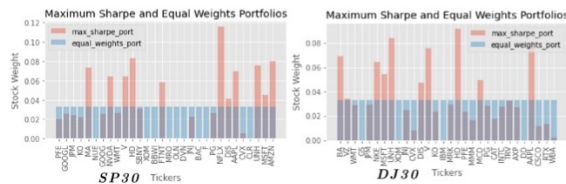
Table 5.15: DQN Agent Performance (Big Winners 2021 Summary)

Scenario 5: A2C ,PPO & DDPG For Multi-Asset Allocation

	Portfolio 1		Portfolio 2 (Max Sharpe Ratio)	
	SP30	DJ30	SP30	DJ30
Annualised return	17.59%	12.97%	24.70%	18.60%
Annualised Variance	2.80%	2.00%	3.20%	2.00%
Annualised Std	16.80%	14.10%	17.90%	14.10%
Sharpe Ratio	0.81	0.64	1.27	1.17

Table 5.16: A Comparison of 2 Portfolios

Note: SP30 (large caps + big winners + extras), DJ30 (30 Dow Jones stocks)



	SP30			DJ30		
	A2C	PPO	DDPG	A2C	PPO	DDPG
Begin total asset	1	1	1	1	1	1
End total asset	4.250	4.264	4.726	3.378	3.034	3.380
Sharpe Ratio	1.205	1.214	1.264	1.119	1.030	1.131
training Duration	2 mins	2 mins	36 mins	2 mins	2 mins	38 mins

Table 5.19: Model Performances on train and test sets

Note: Train (black), Test (red) and total asset (in millions)

	DJ30 Best model (A2C)	SP30 Best model (DDPG)
Annual return	20.482%	30.343%
Cumulative returns	50.292%	78.501%
Annual volatility	24.719%	24.972%
Sharpe ratio	0.88	1.19
Calmar ratio	0.60	0.94
Max drawdown	-34.264%	-32.408%
Omega ratio	1.21	1.28
Stability	0.78	0.90
Sortino ratio	1.24	1.69
Skew	-0.49	-0.26
Kurtosis	14.26	13.93
Tail ratio	0.94	0.90
Daily value at risk	-3.028%	-3.029%
Alpha	0.06	0.15
Beta	0.92	0.91

Table 5.20: Backtest Comparison (Best models: A2C vs DDPG)

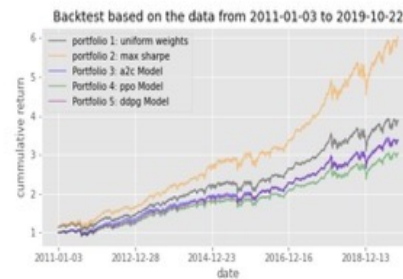


Figure 5.38: Models Performance on train data (DJ30)



Figure 5.39: Models Performance on test data (DJ30)

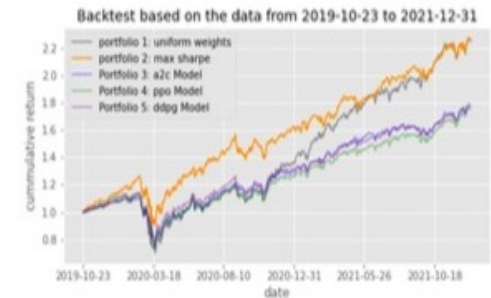


Figure 5.40: Models Performance on train data (SP30)

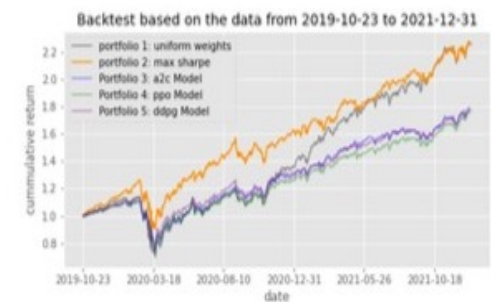
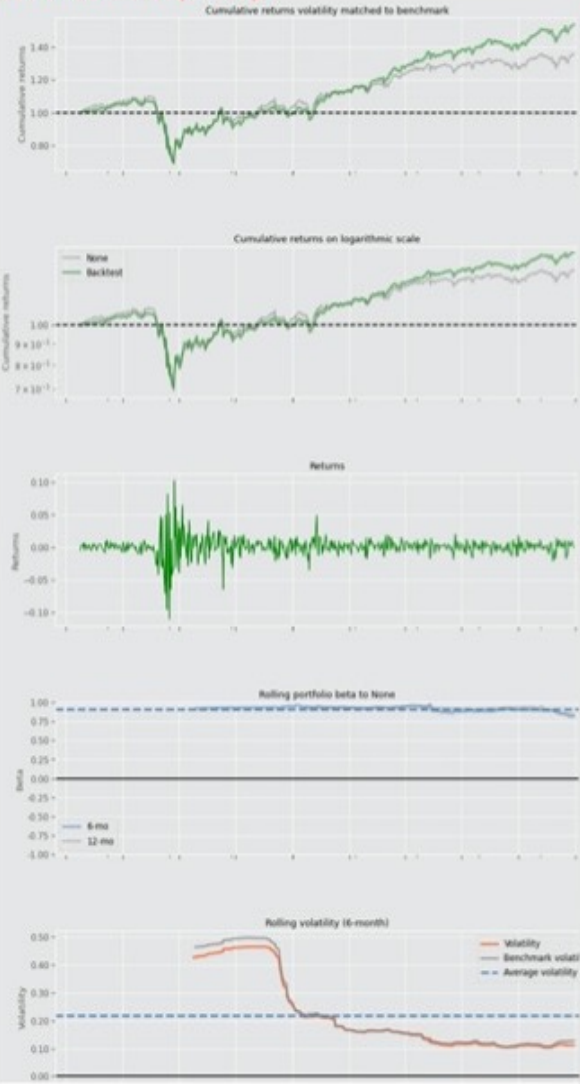
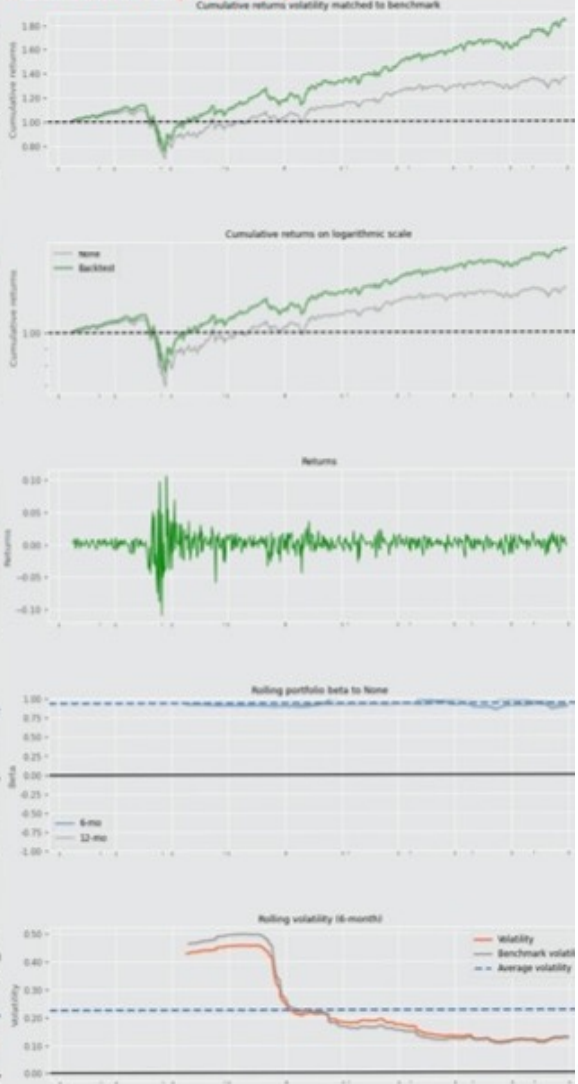


Figure 5.41: Models Performance on test data (SP30)

DJ30:
Best Model (A2C)



SP30:
Best Model (DDPG)



DJ30:
Best Model (A2C)



SP30:
Best Model (DDPG)

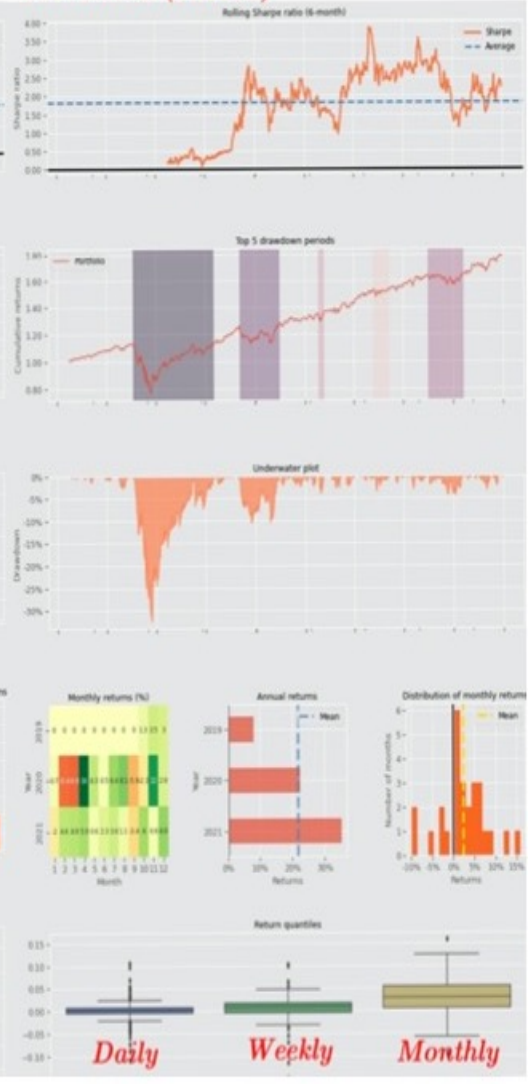


Figure 5.42: Backtest Comparison 2 (Best models: A2C vs DDPG)

Figure 5.43: Backtest Comparison 3 (Best models: A2C vs DDPG)

Scenario 6: Multi-Agent RL for Stock liquidation

DRL environment :
Single Agent

$$R_t = U_t(x_t^*) - U_{t+1}(x_{t+1}^*)$$

MARL environment :
multiple Agents

In a MARL with i agents, the state vector at time t_k is:

$$s_k = [r_{k-D}, \dots, r_{k-1}, r_k, m_k, l_{1,k}, \dots, l_{i,k}]$$

Action a determined by $n_{i,k} = a_{i,k} \times x_{i,k}$

Reward defined as: $R_{i,t} = U_{i,t}(x_{i,t}^*) - U_{i,t+1}(x_{i,t+1}^*)$

Observation defined as: $O_{i,k} = [r_{k-D}, \dots, r_{k-1}, r_k, m_k, l_{i,k}]$

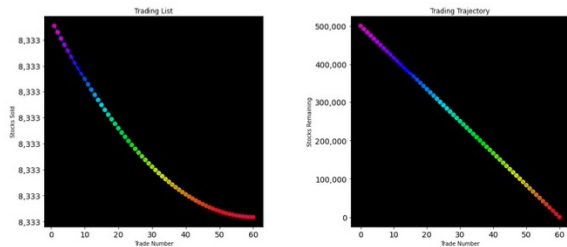


Figure 5.44: Trading List & Trading Trajectory (0.5M;0.5M)

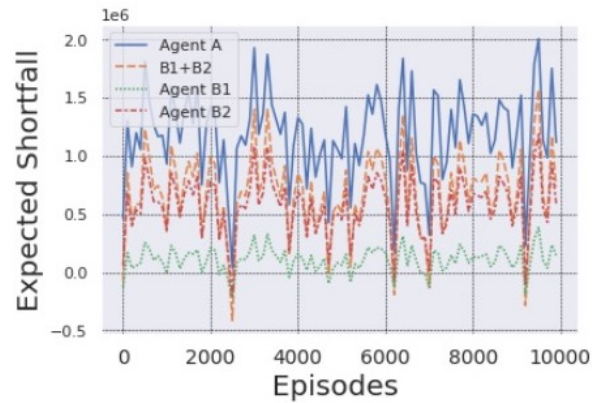


Figure 5.45: Expected Shortfalls (agents different risks (0.5M;0.5M))

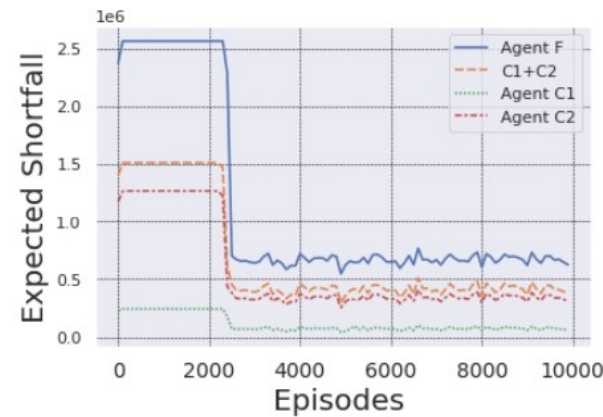


Figure 5.46: Expected Shortfalls (agents same risk level (0.5M;0.5M))

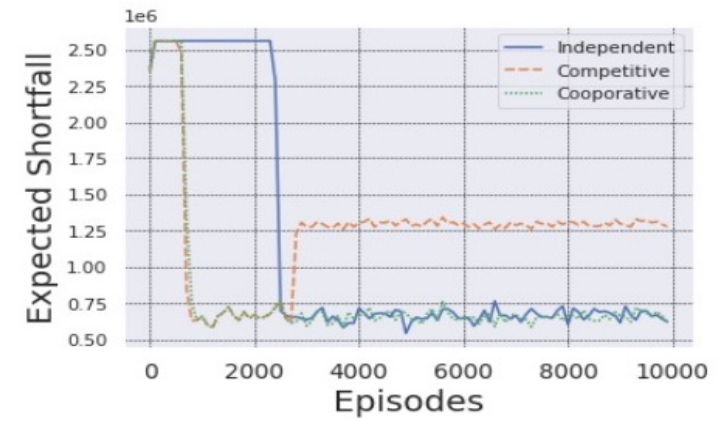


Figure 5.48: Competitive & Cooperative vs Independent

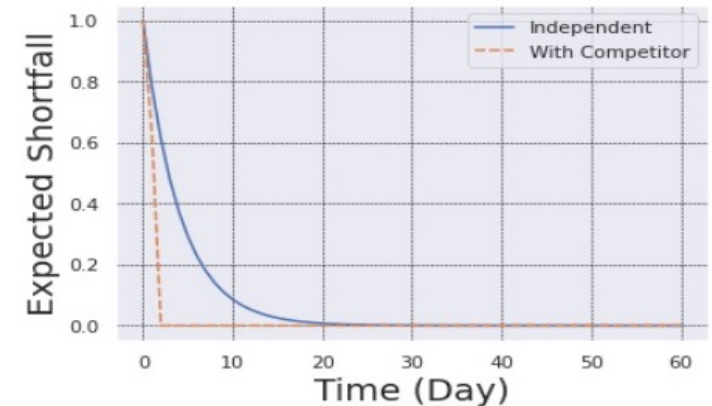


Figure 5.49: Independent Agent vs Competitive Agent

Key Benefits

- DRL is well-suited for trading scenarios and optimisation
- Long-term results is desirable in a financial context
- Better understanding of competition
- Determine optimal hedging strategies
- Possible emulation of human performance & surpassing
- Drug Discovery, Self-Driving cars, chemical reaction

KALMAN FILTERS

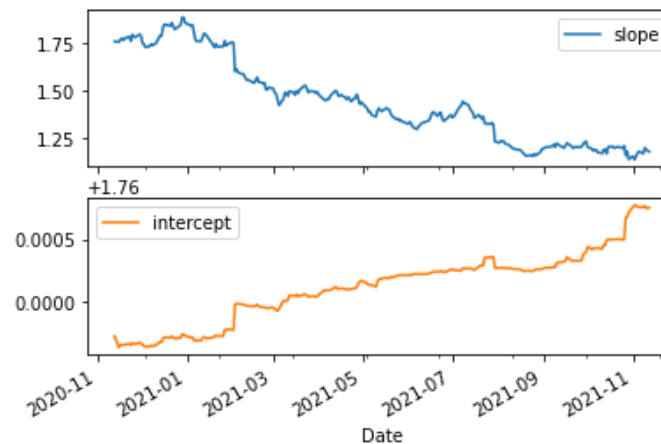
1. Continuous Vars (HMM)
2. Guassian Assumptions
3. Matrix Computations

Future Work

Question: Could we combine HMM, Kalman Filters & DRL to form a new system called HRDKF for the purpose of emulating a Multi-Agent trading environment with parameter sharing?

HMM example

1. Useful & Powerful
2. Ease of parameter sharing
3. Inference reduction ($O(n^2)$ to $O(n)$)

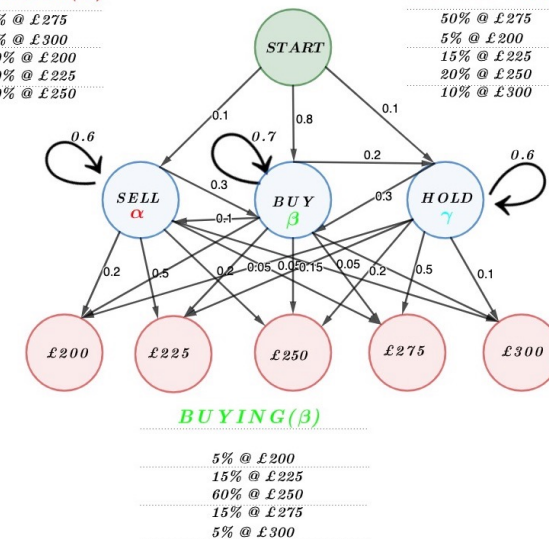


SELLING (α)

5% @ £275
5% @ £300
20% @ £200
50% @ £225
20% @ £250

HOLDING (γ)

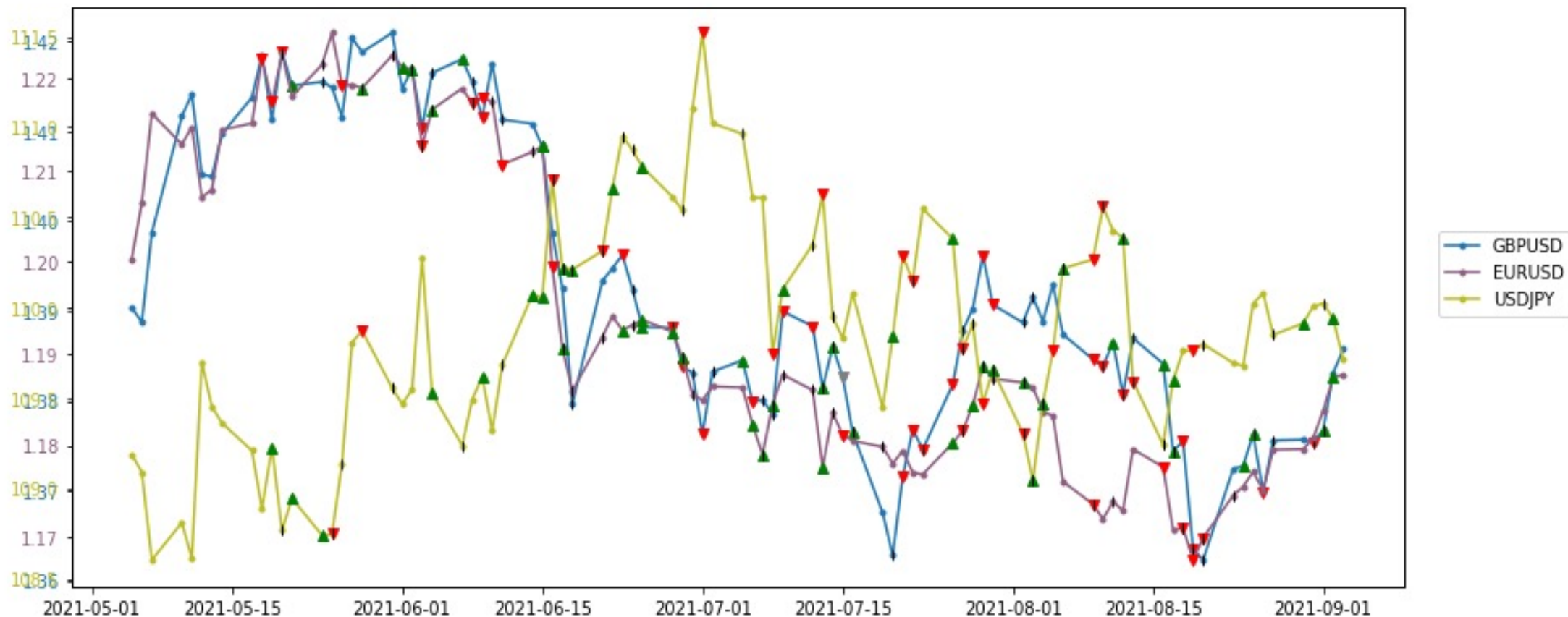
50% @ £275
5% @ £200
15% @ £225
20% @ £250
10% @ £300



FINAL NOTE

Indeed, DRL are being utilized with great Success in Finance, leveraging today's computer capabilities and DRL innovative algorithms.

Balance: 991908.445485 GBP ~ Equity: 991617.873220 ~ Margin: 300.531556 ~ Free Margin: 991317.341663 ~ Margin Level: 3299.546598



THANKS

YOU MIGHT ALSO ENJOY


- Library Authors & Maintainers (Tensorflow, Keras, Rllib, yfinance, FinRL...)
 - <https://github.com/AI4Finance-Foundation/FinRL>
 - Black, F., Litterman, R.: Global portfolio optimization. Financial Analysts (1992)
 - Mnih, V. (2016). Asynchronous Methods for Deep Reinforcement Learning.
<https://arxiv.org/pdf/1602.01783.pdf> [Accessed on 27/06/2021]
 - Schulman, J. et al. (2017) Proximal policy optimization algorithms.
[Online] Available at: <https://arxiv.org/pdf/1707.06347.pdf>
 - <https://github.com/WenhangBao/Multi-Agent-RL-for-Liquidation>
 - https://github.com/matlab-deep-learning/reinforcement_learning_financial_trading
 - <https://uk.mathworks.com/help/finance/hedging-option-using-reinforcement-learning.html>
 - Cao et al (2021). Deep Hedging of Derivatives Using Reinforcement Learning
<https://arxiv.org/pdf/2103.16409.pdf>
 - **Castle Labs**: Computational STochastic optimization and Learning
<https://castlelab.princeton.edu>
 - WhiRL is a machine learning research group in the Department of Computer Science at the University of Oxford
<http://whirl.cs.ox.ac.uk/index.html>
- PyMarl : Python Multi-Agent Reinforcement Learning framework
- <https://github.com/AminHP/gym-anytrading>
 - <https://github.com/AminHP/gym-mtsim>
- 
- Image from Castle Labs*



Image from Castle Labs

THANKS

ALL RESOURCES FROM THIS PROJECT
WILL IMMEDIATELY AVAILABLE TO THE GENERAL PUBLIC:

- GITHUB REPO:

<https://github.com/youms56/RLProjectThesisFinalProject.git>

- VIDEO ON YOUTUBE:

https://www.youtube.com/channel/ucb_4yl3smasdjquqlhzhv7q

- ONEDRIVE ACCOUNT:

my.sharepoint.com/:f:/g/personal/100520418_unimail_derby_ac_uk/EoWUmbTM9jFNmiaqF6rs2okB64avqeLcoFk4Zn8FnAlX4Q?e=vNRrCU

- GOOGLE DRIVE ACCOUNT:

<https://drive.google.com/drive/folders/12rOEjA9GM20giTejxyQET42pPZnie0gc?usp=sharing>

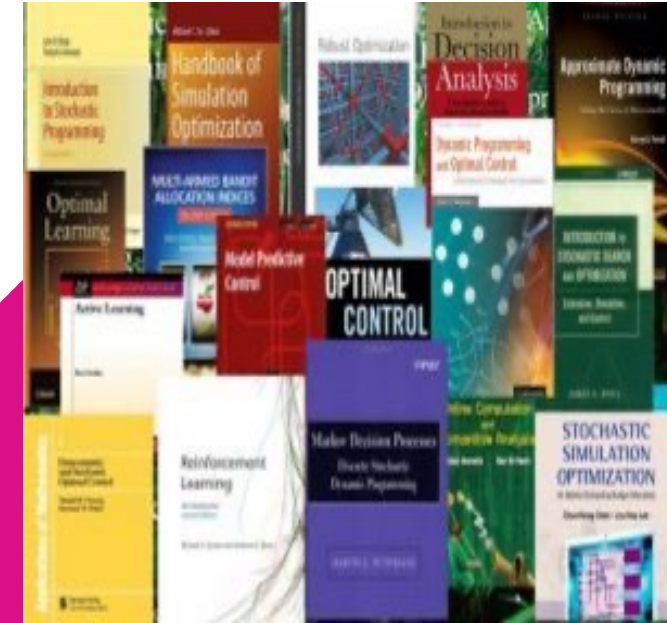


Image from Castle Labs