

# Stock Trading Automation with Deep Reinforcement Learning

Joey Yip  
CSCE 790



# Problem Statement

- Learning to predict market behavior based on historical data remains an *open challenge*
- Partially observable factors
  - Inherent complexity
  - Adaptive
  - Temporal nature
- Hard to model the *true state* as well as the *transition dynamics*
- Iterative process of stock trading and learning to improve and optimize trading strategies (policies) over time make this a well-suited problem for Deep Reinforcement Learning (DRL)
- **Objective:** *Maximize* the cumulative return for a portfolio

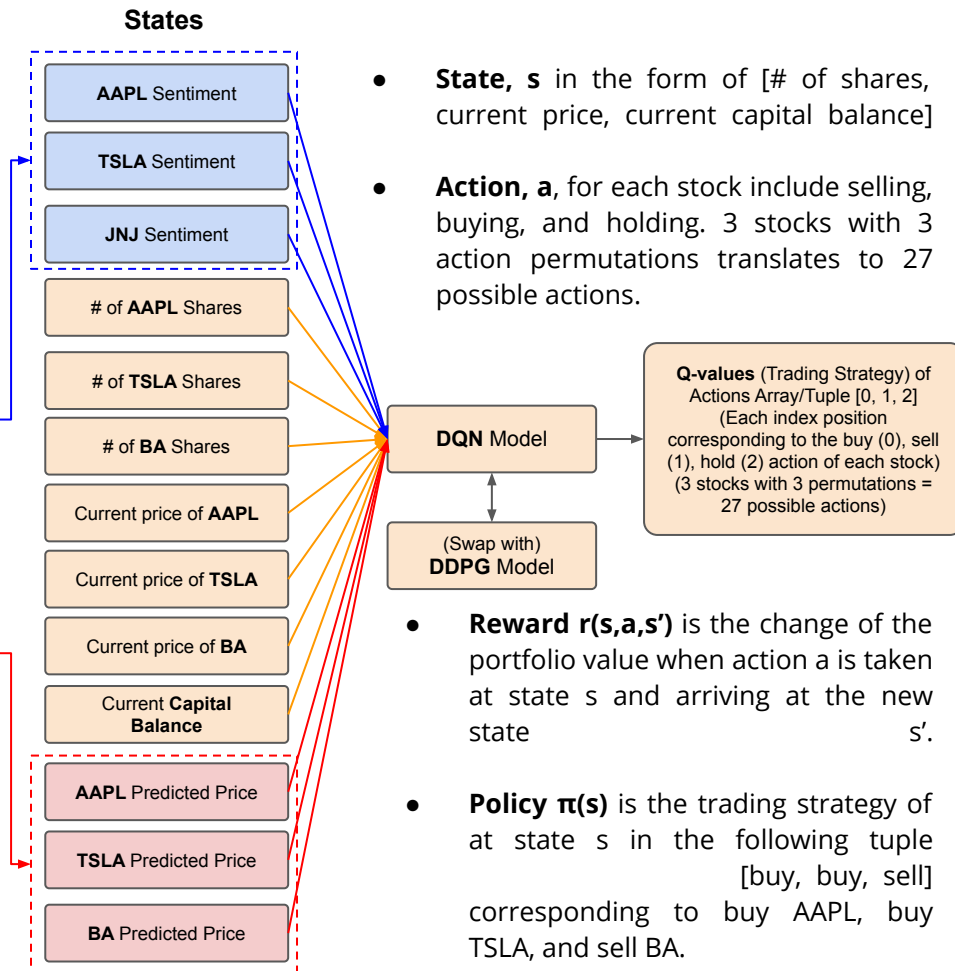
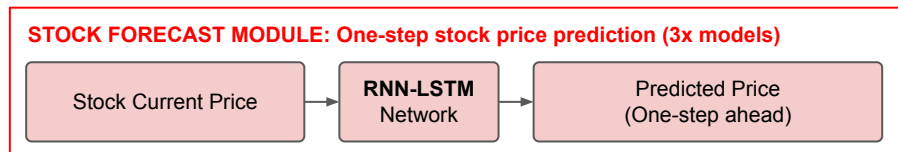
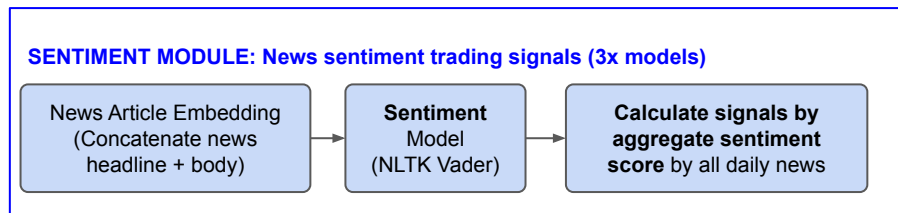
# Related Work & Build Up

- Recent research on **Deep Deterministic Policy Gradient (DDPG)** and **Ensemble Strategy** shown promising results on outperforming the Dow Jones Industrial Average and in **accumulated portfolio return**.

However, these approaches only consider the **historical daily prices** as the states.  
Stock prices often driven by public sentiment towards news events.

- Therefore, this project proposes the additional trading indicators
  1. news sentiment metadata, and
  2. one-step stock prediction into the existing states (to capture stock trend movements)
- **Rationale:** Train an agent that will trade not only by historical prices, but also **take in account additional trading signals** when deciding to buy, sell or hold a particular stock.

# Proposed Approach



# Experiments

## **Deep Q Network (DQN) with Experience Replay** - Discrete action values [0, 1, 2]

- DQN with historical stock prices (Baseline)
- DQN with historical stock prices + **Onestep Forecast**
- DQN with historical stock prices + **News Sentiment**
- DQN with historical stock prices + **Onestep Forecast** + **News Sentiment**

## **Deep-deterministic Policy Gradient (DDPG)** - Continuous action values [0.5, 1.3, 1.9]

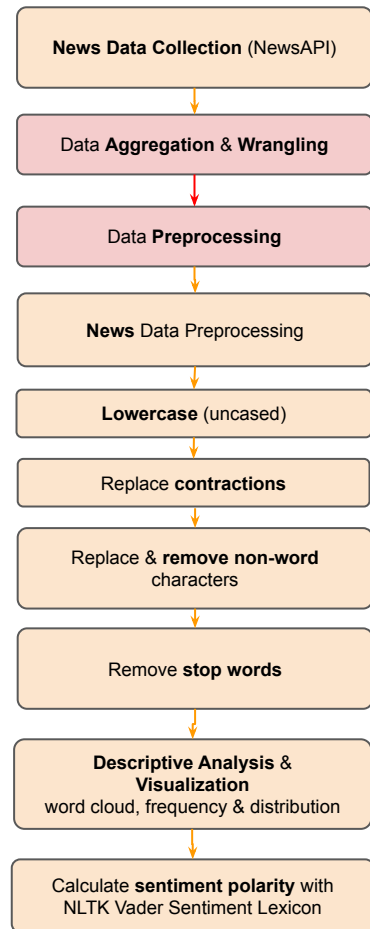
- DDPG with historical stock prices (Baseline)
- DDPG with historical stock prices + **Onestep Forecast**
- DDPG with historical stock prices + **News Sentiment**
- DDPG with historical stock prices + **Onestep Forecast** + **News Sentiment**

# Datasets

- **News** data (headlines and descriptions) from NewsAPI (news)
- **Financial** data from AlphaVantage (daily prices)
- 10/1/20 - 11/29/20
- AAPL (Apple), TSLA (Tesla), BA (Boeing)

Experienced volatility in the past month due to major news events such as **Apple released iPhone 12, Presidential elections**, and **COVID-19** which makes an interesting task for a DRL agent to optimize the “best” trading strategy.

# Sentiment Module



## Apple

Analyzing AAPL Stock Sentiment				
	title	description	description_sentiment	AAPL_Sentiment
0	apple iphone 12 event preview 4 new iphones	apple new iphones come next week four new flag...	0.0000	0.0000
1	judge let apple block unreal engine make unblo...	trial epic games legal battle apple app store ...	-0.2732	-0.7136
2	united states judge makes mixed ruling epic in...	federal judge california friday ruled injuncti...	-0.4404	-0.0189
3	apple extend free subscription video	apple inc aapl said thursday would extend free...	0.6124	1.2248
4	microsoft bringing xcloud ios via web	microsoft working browser based xcloud solutio...	0.3182	0.3182
...	...	...	...	...
3088	cardinals vs patriots live stream watch nfl we...	get cardinals vs patriots live stream watch nf...	0.8481	0.8481
3089	chiefs vs buccaneers live stream watch nfl wee...	game makings classic get kansas city chiefs vs...	0.0000	0.0000
3090	ireland vs georgia live stream watch nations c...	get ireland vs georgia live stream watch autom...	0.7964	1.3070
3091	f1 bahrain grand prix live stream watch 2020 b...	watch f1 bahrain grand prix live stream catch ...	0.4588	0.9176
3092	mike tyson vs roy jones jr live stream watch f...	get tyson vs jones live stream ad watch fight ...	-0.0258	-0.4076

## Boeing

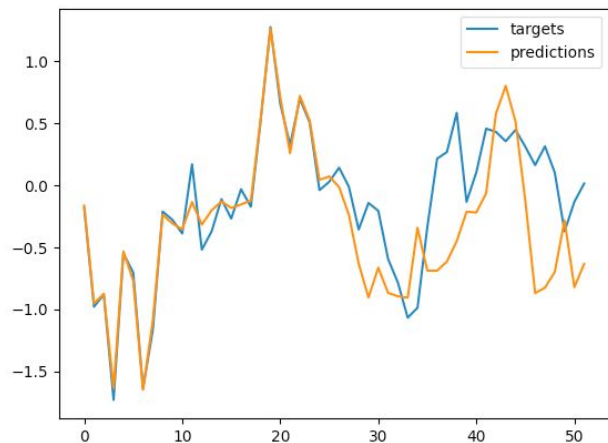
Analyzing BA Stock Sentiment				
	title	description	description_sentiment	BA_Sentiment
0	exclusive boeing talks alaska airlines potenti...	boeing co ba n discussions sell 737 max jets a...	0.0258	0.1538
1	airbus ramps deliveries september seen targeti...	airbus air pa reached highest number monthly d...	-0.5267	-0.5267
2	norwegian air considers shutting gatwick hub l...	norwegian air considering shutting base gatwic...	-0.9100	-0.9100
3	norwegian air considers gatwick lay	norwegian air considering laying staff base ga...	-0.9100	-0.9100
4	tiny sun country airlines genius decision fly ...	summary list placement last december month cov...	-0.4215	-0.0196
...	...	...	...	...
1632	trump claims came vaccines biden injured playi...	trump continued push unverified claims stolen ...	-0.4939	-0.7202
1633	covid vaccines airlines face key mission bloom...	covid vaccines airlines face key missionbloomber...	0.0000	0.0000
1634	amazon expanding logistics empire like never p...	amazon hired 420000 employees brace surge onli...	0.5267	0.7630
1635	covid 19 vaccine success cause optimism rollou...	every american could access covid 19 vaccine e...	-0.2732	0.4452
1636	meet 50 ceos investors hackers helped lead cyb...	summary list placementin february san francisc...	0.0000	0.0000

## Tesla

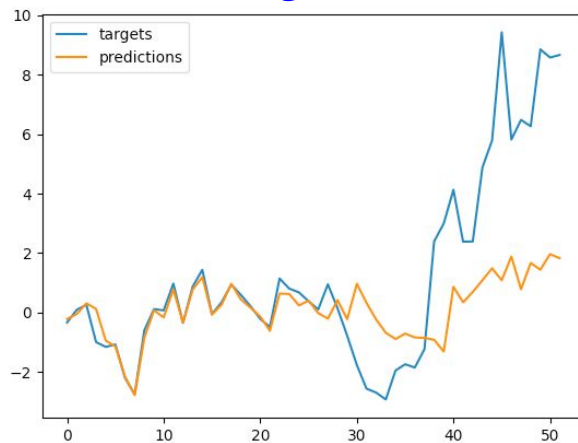
Analyzing TSLA Stock Sentiment				
	title	description	description_sentiment	TSLA_Sentiment
0	tesla finally launches two factor authentication	long string delays tesla finally implemented c...	0.3400	0.3400
1	tesla shot producing half million cars 2020 memo	tesla inc tsla chance producing 500000 cars ye...	0.2500	0.2500
2	tesla co founder j b straubel wants build star...	tesla co founder j b straubel wants build star...	0.2023	0.2023
3	spacex launched enough satellites starlink's u...	timeline uncertain newly launched satellites m...	-0.1531	-0.0251
4	spacex doubled clearing village building rocke...	summary list placementspacex made calls final ...	0.5574	0.4294
...	...	...	...	...
909	supreme court hear first big cfaa case	supreme court hear arguments monday case could...	-0.8360	-0.2786
910	inside story henrik fisker pulled \$ 3 billion ...	summary list placementhenrik fisker study resi...	0.0258	0.0258
911	equity markets devolved casinos former wall st...	summary list placementthe belief housing price...	0.2732	0.0709
912	volkswagen may release small ev little \$ 24 000	future volkswagen may pretty affordable electr...	0.4939	0.4939
913	here's us economy could transform biden appoin...	former federal reserve chair set first person ...	0.2023	0.4046

# LSTM Onestep Forecast Module with Sliding Window of 10 (prices of the past 10 days)

## Apple Forecast



## Boeing Forecast



## Tesla Forecast





# States Representation (before normalization)

	timestamp	AAPL	BA	TSLA	AAPL_Forecast	BA_Forecast	TSLA_Forecast	AAPL_Sentiment	BA_Sentiment	TSLA_Sentiment
0	2020-10-09	115.28	169.9300	430.1300	115.626106	170.083801	429.269348	16.8374	22.9691	6.6002
1	2020-10-12	120.06	167.7500	442.0000	120.161232	166.785553	441.086578	41.4626	-0.8622	-0.4947
2	2020-10-13	125.27	165.3600	443.3500	125.193718	165.056900	445.059906	29.0896	0.3892	1.8425
3	2020-10-14	121.00	162.7000	449.7800	121.372795	163.192902	449.106537	27.1652	11.2219	6.3633
4	2020-10-15	118.72	161.0100	450.3100	118.252190	160.192429	452.636047	26.4232	0.5879	0.3658
5	2020-10-16	121.28	171.2200	454.4400	121.442116	168.023529	449.427612	23.1806	9.8061	3.7285
6	2020-10-19	119.96	169.0900	446.2400	120.080605	167.964294	446.013367	37.3280	15.2290	10.5382
7	2020-10-20	116.20	168.2900	431.7500	116.770760	165.543274	435.844879	27.9774	0.1049	10.9341
8	2020-10-21	116.67	166.4400	422.7000	116.963181	166.546005	421.197632	34.7908	7.1339	18.4842
9	2020-10-22	117.45	164.7000	441.9200	116.367424	163.934662	408.478210	34.1368	-0.0528	8.1257
10	2020-10-23	116.39	169.9900	421.8400	114.836533	162.752960	424.281433	35.7615	6.8626	3.8984
11	2020-10-27	115.49	159.1600	423.7600	110.229721	162.650406	406.162750	32.9238	6.5642	10.6035
12	2020-10-28	115.05	153.0000	416.4800	111.891815	170.087463	432.457062	24.4499	-5.9386	9.2907
13	2020-10-29	112.37	148.0300	409.9600	110.486961	166.081665	435.012756	31.9855	13.3517	8.5056
14	2020-10-30	111.06	147.2000	406.8954	110.302361	162.581406	428.590118	13.9098	0.9882	6.4001
15	2020-11-02	109.11	145.7500	394.0000	110.218719	159.790070	435.555481	27.9320	-0.7152	-0.6954
16	2020-11-03	109.66	151.8300	409.7300	114.109825	158.467392	441.628174	22.7629	1.8586	2.6818
17	2020-11-04	114.14	153.1700	430.6200	111.724678	159.610474	459.091827	16.3490	-0.0912	1.2712
18	2020-11-05	117.95	152.4800	428.3000	111.724220	158.814545	462.927277	27.5429	3.1733	8.9129
19	2020-11-06	118.32	156.3300	436.1000	112.229881	158.701111	463.035004	25.8112	0.9215	6.5744
20	2020-11-10	115.55	182.8000	420.0900	114.995636	155.834671	418.607208	26.0223	23.9995	2.2220
21	2020-11-11	117.19	189.8900	416.4500	114.958824	169.473160	403.811859	29.5164	14.5148	2.0514
22	2020-11-12	119.62	178.9500	415.0500	116.023232	166.197113	415.180389	34.5058	7.3910	5.0093
23	2020-11-13	119.44	178.9997	410.8500	120.454453	168.383270	452.657074	34.6146	13.6549	14.6865
24	2020-11-16	118.92	194.5600	408.9300	121.996758	170.832382	464.674011	23.1066	16.9128	13.4888

# RL Models

---

## Algorithm 1 Deep Q-learning with Experience Replay

---

```

Initialize replay memory  $\mathcal{D}$  to capacity  $N$ 
Initialize action-value function  $Q$  with random weights
for episode = 1,  $M$  do
    Initialise sequence  $s_1 = \{x_1\}$  and preprocessed sequenced  $\phi_1 = \phi(s_1)$ 
    for  $t = 1, T$  do
        With probability  $\epsilon$  select a random action  $a_t$ 
        otherwise select  $a_t = \max_a Q^*(\phi(s_t), a; \theta)$ 
        Execute action  $a_t$  in emulator and observe reward  $r_t$  and image  $x_{t+1}$ 
        Set  $s_{t+1} = s_t, a_t, x_{t+1}$  and preprocess  $\phi_{t+1} = \phi(s_{t+1})$ 
        Store transition  $(\phi_t, a_t, r_t, \phi_{t+1})$  in  $\mathcal{D}$ 
        Sample random minibatch of transitions  $(\phi_j, a_j, r_j, \phi_{j+1})$  from  $\mathcal{D}$ 
        Set  $y_j = \begin{cases} r_j & \text{for terminal } \phi_{j+1} \\ r_j + \gamma \max_{a'} Q(\phi_{j+1}, a'; \theta) & \text{for non-terminal } \phi_{j+1} \end{cases}$ 
        Perform a gradient descent step on  $(y_j - Q(\phi_j, a_j; \theta))^2$  according to equation 3
    end for
end for

```

---

## Deep Q Network (DQN)

## Deep-deterministic Policy Gradient (DDPG)

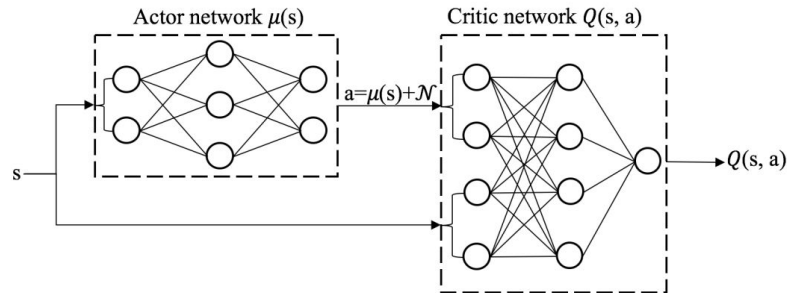


Figure 2: Learning network architecture.

---

## Algorithm 1 DDPG algorithm

---

- 1: Randomly initialize critic network  $Q(s, a | \theta^Q)$  and actor  $\mu(s | \theta^\mu)$  with random weight  $\theta^Q$  and  $\theta^\mu$ ;
- 2: Initialize target network  $Q'$  and  $\mu'$  with weights  $\theta^{Q'} \leftarrow \theta^Q, \theta^{\mu'} \leftarrow \theta^\mu$ ;
- 3: Initialize replay buffer  $R$ ;
- 4: **for** episode = 1,  $M$  **do**
- 5:   Initialize a random process  $\mathcal{N}$  for action exploration;
- 6:   Receive initial observation state  $s_1$ ;
- 7:   **for**  $t = 1, T$  **do**
- 8:     Select action  $a_t = \mu(s_t | \theta^\mu) + \mathcal{N}_t$  according to the current policy and exploration noise;
- 9:     Execute action  $a_t$  and observe reward  $r_t$  and state  $s_{t+1}$ ;
- 10:    Store transition  $(s_t, a_t, r_t, s_{t+1})$  in  $R$ ;
- 11:    Sample a random minibatch of  $N$  transitions  $(s_i, a_i, r_i, s_{i+1})$  from  $R$ ;
- 12:    Set  $y_i = r_i + \gamma Q'(s_{t+1}, \mu'(s_{i+1} | \theta^{\mu'}) | \theta^{Q'})$ ;
- 13:    Update critic by minimizing the loss:  $L = \frac{1}{N} \sum_i (y_i - Q(s_i, a_i | \theta^Q))^2$ ;
- 14:    Update the actor policy by using the sampled policy gradient:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_i \nabla_a Q(s, a | \theta^Q) |_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s | \theta^\mu) |_{s_i};$$

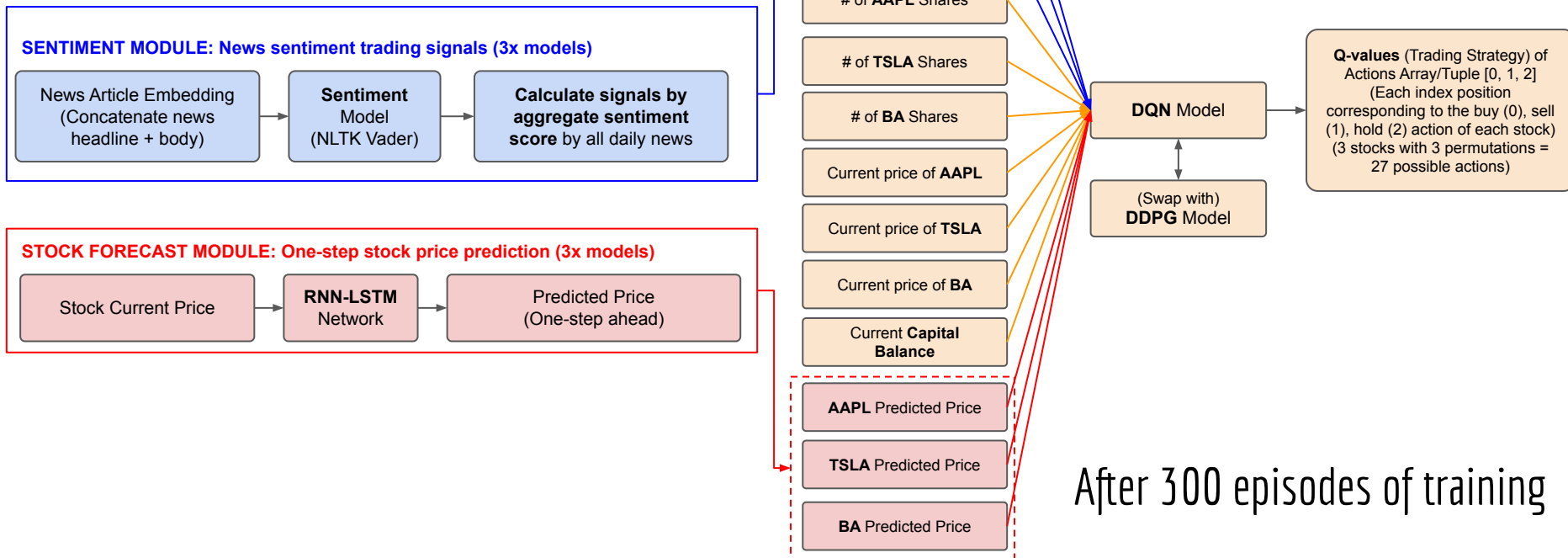
- 15:   Update the target networks:

$$\theta^{Q'} \leftarrow \tau \theta^Q + (1 - \tau) \theta^{Q'},$$

$$\theta^{\mu'} \leftarrow \tau \theta^\mu + (1 - \tau) \theta^{\mu'}.$$

- 16:   **end for**
  - 17: **end for**
-

# Putting it all together



# Results

<b>Models</b> (Epsilon-greedy policy of 0.6)	<b>Initial Capital</b>	<b>Cumulative Portfolio Return</b>	<b>Return of Investment</b>
<b>DQN</b>	10,000	14,846.29	~1.5x
<b>DQN + Onestep Forecast</b>	10,000	<b>19,500.61</b>	<b>~1.9x</b>
<b>DQN + Sentiment</b>	10,000	17,710.09	~1.7x
<b>DQN + Onestep Forecast + Sentiment</b>	10,000	18,112.42	~1.8x
<b>DDPG</b>	10,000	<b>14,185.75</b>	<b>~1.4x</b>
<b>DDPG + Onestep Forecast</b>	10,000	13,185.75	~1.3x
<b>DDPG + Sentiment</b>	10,000	12,560.93	~1.25x
<b>DDPG + Onestep Forecast + Sentiment</b>	10,000	13,061.52	~1.3x

# Discussion

- **DQN + Onestep Forecast** performed the best with an average of ~1.9x return!
  - LSTM models were able to capture the stock prices & trends
  - Note: Multistep forecast resulted in flat-line trend (recurrent nature)
- **News sentiment signals** do not seem to perform well
  - Mix signals aggregated from the daily news
  - Search keywords affect the quality of datasets being collected
    - Keywords engineering
  - Data scrutinization on irrelevant and fake news

Analyzing AAPL Stock Sentiment					
	title	description	... description_sentiment	AAPL_Sentiment	
0	apple iphone 12 event preview 4 new iphones	apple new iphones come next week four new flag...	...	0.0000	0.0000
1	judge let apple block unreal engine make unblo...	trial epic games legal battle apple app store ...	...	-0.2732	-0.7136
2	united states judge makes mixed ruling epic in...	federal judge california friday ruled injuncti...	...	-0.4404	-0.0189
3	apple extend free subscription video	apple inc aapl said thursday would extend free...	...	0.6124	1.2248
4	microsoft bringing xcloud ios via web	microsoft working browser based xcloud solutio...	...	0.3182	0.3182
...	...	...	...	...	...
3088	cardinals vs patriots live stream watch nfl we...	get cardinals vs patriots live stream watch nf...	...	0.8481	0.8481
3089	chiefs vs buccaneers live stream watch nfl wee...	game makings classic get kansas city chiefs vs...	...	0.0000	0.0000
3090	ireland vs georgia live stream watch nations c...	get ireland vs georgia live stream watch autum...	...	0.7964	1.3070
3091	f1 bahrain grand prix live stream watch 2020 b...	watch f1 bahrain grand prix live stream catch ...	...	0.4588	0.9176
3092	mike tyson vs roy jones jr live stream watch f...	get tyson vs jones live stream ad watch fight ...	...	-0.0258	-0.4076

- **DDPG** did not perform well in this datasets compared to DQN
  - Small available datasets (Newsapi allows scraping up to 1 month old news)
  - Data hungry

# Conclusion & Future Work

- This project built-off existing DDPG work
  - Preliminary approach to augment state-of-the-art DRL for stock trading with news sentiment signals and onestep forecast
- Expand on bigger datasets
- Good quality datasets definitely affect the model outputs!

## References

- [1] Xiong, Z., Liu, X. Y., Zhong, S., Yang, H., & Walid, A. (2018). Practical deep reinforcement learning approach for stock trading. arXiv preprint arXiv:1811.07522.
- [2] Wu, X., Chen, H., Wang, J., Troiano, L., Loia, V., & Fujita, H. (2020). Adaptive Stock Trading Strategies with Deep Reinforcement Learning Methods. Information Sciences.
- [3] Yang, H., Liu, X. Y., Zhong, S., & Walid, A. (2020). Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy. Available at SSRN.
- [4] Rehman, A. U., Malik, A. K., Raza, B., & Ali, W. (2019). A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. Multimedia Tools and Applications, 78(18), 26597-26613.
- [5] Moghar, A., & Hamiche, M. (2020). Stock Market Prediction Using LSTM Recurrent Neural Network. Procedia Computer Science, 170, 1168-1173.
- [6] Bharathi, S., & Geetha, A. (2017). Sentiment analysis for effective stock market prediction. International Journal of Intelligent Engineering and Systems, 10(3), 146-154.
- [7] Carta, S., Corrigan, A., Ferreira, A., Podda, A. S., & Recupero, D. R. (2020). A multi-layer and multi-ensemble stock trader using deep learning and deep reinforcement learning. Applied Intelligence, 1-17.