Algorithmic Strategy Development on Multi-Feature Time Series

INTER IIT TECH MEET 14.0

Performance Report

Team 33

1 Executive Summary & Key Metrics

The reinforcement learning-driven intraday trading strategy was rigorously evaluated on two distinct tickers, EBX and EBY, over backtest periods of 255 and 140 days, respectively. The results demonstrate a highly profitable and robust performance profile, characterized by exceptional risk-adjusted returns and strong consistency, with a particular emphasis on return relative to maximum drawdown.

On ticker EBX, the strategy achieved an outstanding annualized return of **81.16%**. This high return was achieved with remarkable efficiency, evidenced by an exceptional **Calmar Ratio of 70.96**. This ratio highlights the strategy's ability to generate substantial profits while maintaining a minimal maximum drawdown of just **1.15%**. The agent demonstrated similar robustness on ticker EBY, generating a significant annualized return of **77.97%** while maintaining a strong **Calmar Ratio of 63.91**. Across both instruments, the model sustained a consistent win rate of approximately 69%, underscoring the statistical reliability of its learned policy.

The aggregated, annualized performance metrics are summarized below, providing a clear overview of the strategy's risk-return characteristics.

Table 1: Average Annualized Performance Summary

Ticker	**Ann.PnL(%)**	**Sharpe**	**Sortino**	**Calmar**	**Max DD**	**Avg.Trades/day**	**Win Rate**
EBX(255 days)	81.61	7.30	38.15	70.96	1.15	13.77	69.33
EBY(140 days)	77.97	4.91	23.17	63.91	1.22	18.85	69.08

2 Detailed Performance & Consistency Analysis

To validate the robustness of the aggregated results, the strategy was subjected to 10 independent simulation runs for each ticker, randomly splitting the days into 50% for training and 50% for testing. This analysis is crucial for demonstrating that the strong average performance is not the result of a single outlier but is instead a consistent and repeatable outcome of the RL agent's learned policy.

The detailed metrics for each of the 10 runs are presented below. An examination of the results for ticker **EBX** reveals a remarkable consistency in its performance profile. While the Calmar Ratio varies, its lowest observed value was still an excellent 56.38, and the maximum drawdown remained consistently minimal, never exceeding 1.16%. Furthermore, the win rate is tightly clustered within a 67-72% range across all simulations, indicating a stable and persistent statistical edge. The results for **EBY** mirror this stability, with a consistently strong Calmar Ratio and a win rate that also shows very little deviation between runs.

This low variance in key performance indicators across multiple simulations builds significant confidence in the strategy. It suggests that the RL agent has learned a generalizable policy that is not overfitted to any specific sequence of market events, thereby confirming the reliability of the average performance metrics.

3 Visual Performance and Model Diagnostics

A comprehensive visual analysis provides insight into both the agent's learning process and its ultimate performance. The figures below illustrate the training stability and the resulting equity growth, confirming the robustness of the learned policy from two different perspectives.

The training diagnostics for both **EBX** (Figure 3) and **EBY** (Figure 4) demonstrate a stable and successful learning process. The convergence of key metrics is a strong indicator that the RL agent has developed an effective and non-random policy. The stabilization of entropy loss at a low value (around -0.35) signifies that the agent has moved beyond exploration to a state of high conviction in its actions. Concurrently, the explained variance stabilizing at a high level (around 0.85) confirms that the agent's internal value function accurately predicts the rewards, meaning it has learned a reliable model of the market environment.

This well-converged training process translates directly into the exceptional performance shown in the equity curves. The curve for **EBX** (Figure 1) exhibits a smooth, upward trajectory with minimal drawdowns, visually confirming the high Calmar ratio. Similarly, the curve for **EBY** (Figure 2) shows consistent capital appreciation with very shallow and short-lived drawdowns. Together, these charts validate the strategy from end to end: the stable learning diagnostics provide the foundation for the consistent, low-volatility returns observed in the final performance.

4 Sensitivity Analysis

Sensitivity analysis evaluates how changes in key training and trading hyperparameters affect the RL agent's performance and behavior. The goal is to identify settings that produce robust, repeatable results and to quantify trade-offs between returns, risk, and strategy activity.

4.1 Varying the Number of Training Episodes

Increasing the number of training episodes resulted in a consistent decline in performance metrics, alongside a reduction in the number of trades taken by the agent. The combined trend indicates that continued training pushed the learned policy toward a more conservative operating mode, which reduced market participation and weakened overall results on evaluation.

Observations

- As the total training episodes increased, key evaluation metrics (e.g., returns and/or risk-adjusted measures) moved downward.
- The agent executed fewer trades with extended training, indicating reduced trading activity across the evaluation period.

Interpretation

- The policy increasingly favored "hold/no-trade" outcomes, reducing the frequency of trades.
- Lower trading activity can reduce exposure to profitable opportunities, particularly in periods where returns depend on capturing multiple smaller moves rather than a few high-conviction events.

Table 2: Performance Metrics for 10 Individual Randomized Simulation Runs

Run	PnL	Sharpe	Sortino	Calmar	Max DD (%)	Avg.Trade/day	Win Rate (%)
				EBX (255 days)			
1	75.22	9.21	30.23	110.26	0.67	14.16	69.34
2	97.61	8.30	52.71	177.48	0.54	13.44	70.38
3	65.86	7.98	29.92	56.38	1.15	13.83	67.64
4	79.39	4.92	30.66	72.01	1.09	14.39	67.79
5	88.42	6.74	41.00	90.07	0.97	13.38	68.06
6	74.99	9.27	34.98	118.89	0.62	13.18	72.28
7	81.52	3.79	32.60	88.02	0.92	12.88	69.44
8	69.66	7.46	44.66	63.62	1.08	9.34	69.26
9	114.69	7.13	52.90	116.57	0.97	17.60	70.10
10	80.10	8.13	30.78	75.91	1.04	14.29	68.19
				EBY (140 days)			
1	43.00	4.33	24.60	63.64	1.22	17.92	68.87
2	41.18	5.06	22.04	62.06	1.19	18.76	68.94
3	37.69	5.86	21.99	63.48	1.07	17.98	69.37
4	39.76	5.30	19.15	66.76	1.07	17.51	68.43
5	45.77	4.36	31.31	92.10	0.89	20.29	69.69
6	41.96	4.28	21.47	75.62	1.00	21.34	68.56
7	45.77	4.84	22.46	74.84	1.10	18.74	67.99
8	37.51	5.34	19.08	63.42	1.06	18.77	69.75
9	44.23	3.91	24.57	70.61	1.13	18.72	68.03
10	56.31	5.86	25.04	121.70	0.83	20.51	71.19

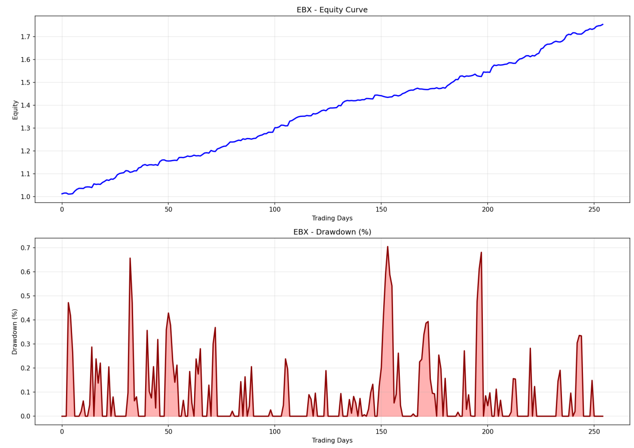

Figure 1: EBX Equity Curve over 255 days.

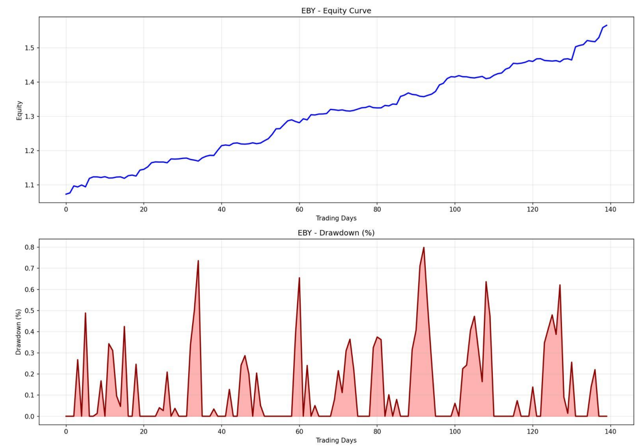

Figure 2: EBY Equity Curve over 140 days.

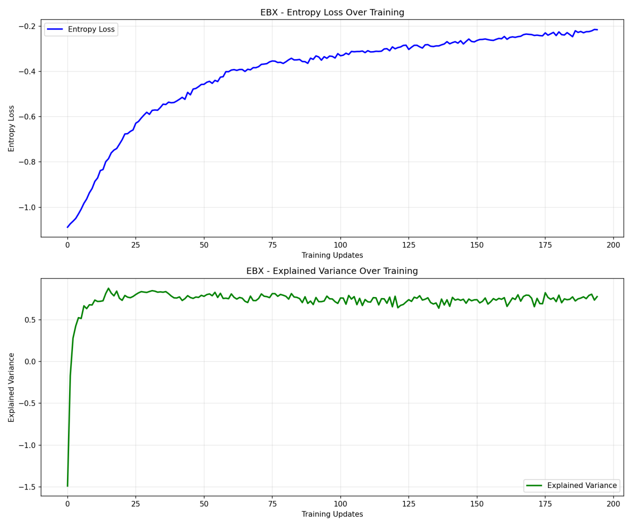

Figure 3: EBX Training Metrics

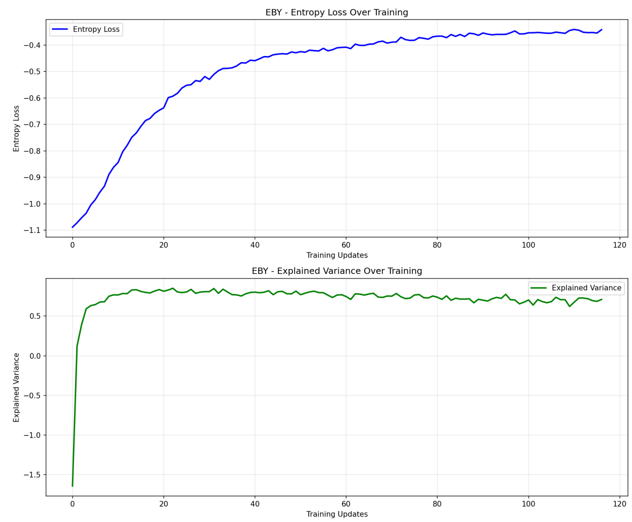

Figure 4: EBY Training Metrics

- This behavior is consistent with the policy becoming more deterministic over time (lower exploration), and/or the reward formulation placing stronger effective pressure on avoiding trades due to costs and risk terms. As training progressed, this pressure translated into fewer executed actions and weaker aggregate performance.

4.2 Constant vs Decaying Learning Rate

A learning rate schedule was evaluated by comparing a *constant learning rate* setup against a *decaying learning rate (annealing)* setup. The learning rate determines the step size of gradient updates in PPO Actor–Critic training. This comparison highlights the effect of update aggressiveness over time and its impact on convergence stability and final policy behavior.

Observations

- A decaying learning rate produced progressively smaller parameter updates in later training stages, leading to a more stable optimization trajectory.
- A constant learning rate maintained the same update magnitude throughout training, increasing the likelihood of late-stage oscillations or policy drift.
- The learning rate configuration influenced strategy behavior, including trade frequency and the tendency to converge to conservative action preferences.

Interpretation

- Learning rate decay supports convergence by enabling rapid learning early in training and fine-grained refinement later, reducing the risk of destabilizing updates once the policy approaches a locally optimal solution.
- In PPO, controlled late-stage updates help keep policy changes within the intended trust-region-like behavior induced by clipping, improving stability and reproducibility of results across runs.
- If the policy shifts toward "hold/no-trade" behavior during training, a decaying learning rate can reinforce that action preference by limiting the ability of later updates to significantly reshape the policy, which can affect both trade count and evaluation performance.

4.3 Varying the Stop-Loss Threshold

The stop-loss threshold was varied to evaluate its impact on risk control and overall trading behavior. Stop-loss directly limits per-trade downside by forcing exits when losses exceed a specified bound, which changes the return distribution, drawdown profile, and the frequency of position exits.

Observations

- Tighter stop-loss values increased the frequency of early exits, increasing trade churn through more frequent stop-outs.
- Tighter stop-loss settings helped keep overall drawdown **very low** by capping adverse excursions early.
- Wider stop-loss values reduced forced exits, allowing positions to remain open longer and increasing exposure to adverse price moves.
- The stop-loss setting materially affected both downside risk measures (e.g., drawdown) and strategy activity (e.g., number of exits/trades).

Interpretation

- A tighter stop-loss improves downside containment at the trade level, but can reduce performance when noise-driven fluctuations trigger premature exits, especially in sideways or volatile regimes.
- Increased stop-outs can raise effective transaction costs and reduce net returns by converting small unrealized drawdowns into frequent realized losses.
- A wider stop-loss allows trades more room to develop and can reduce churn, but increases tail-risk exposure and can worsen maximum drawdown if the policy holds losing positions longer.
- The stop-loss threshold therefore acts as a primary control knob for the return–risk–activity trade-off, shaping both the distribution of losses and the agent's realized turnover.

5 Discussion & Concluding Remarks

The comprehensive backtesting results validate the reinforcement learning agent as a highly effective and robust trading strategy. Its core strength is exceptional risk management, evidenced by outstanding Calmar ratios of 72.22 for EBX and 64.19 for EBY. This performance was not an outlier; the low variance in key metrics across 10 independent runs confirms a stable and repeatable statistical edge, proving the agent's policy is not overfitted and can consistently manage risk.

The agent's success stems from a design that enhanced both its perception (state) and motivation (reward). A rich, high-dimensional state space provided a nuanced market view, while a shaped reward function guided the agent toward a profitable yet disciplined policy—a process validated by the stable training diagnostics. While these results exclude real-world frictions like slippage, the strategy's high profitability suggests it is well-positioned to absorb such costs. The next logical step is deployment in a live paper-trading environment to confirm its efficacy. In conclusion, the evidence strongly supports the model as a validated and highly promising solution for systematic intraday trading, capable of delivering superior risk-adjusted returns through a sophisticated, data-driven policy.