Profit Is in the Past: A Sliding-Window, Sentiment-Enhanced Predictive-ML Reinforcement-Learning Trader — Case Study on Technology Stocks

Mustafa Ercengiz and Hüseyin Sezen

Istanbul Technical University, Maslak, 34467 Sarıyer/Istanbul, Turkey

Abstract. This study introduces a novel hybrid reinforcement learning (RL) model to automate stock trading by combining predictive machine learning (ML) models with sentiment scores. The proposed model employs a dynamic sliding-window approach, where the stock regressor is retrained with the last 30 days of stock prices to predict the next week, and the RL agent is updated on a weekly basis with regressor and instant news sentiment scores in order to make trading decisions on minute-by-minute data. The efficiency of the system was tested using multiple large-cap stocks, namely NVDA, TSLA, MSFT, GOOGL, AAPL, and META.The experimental setup utilized a vast set of news headlines together with minute-by-minute closing prices from December 19, 2024, to March 19, 2025. A Monte Carlo simulation with 30 randomized runs was implemented on NVIDIA (NVDA) to test the robustness and reliability of the developed methodology. The outcome, with a mean return of +14.61%, was significantly better compared to the traditional buy-and-hold strategy, which had a return of 5.9%. Initial experiments performed without the sliding window mechanism resulted in a loss of 25%, proving the crucial requirement for regular model adaptation in dynamic market scenarios.scenarios.

Keywords: Reinforcement Learning, Algorithmic Trading, Stock Market Prediction, Sentiment Analysis, Financial Time Series, Sliding Window, Machine Learning, LSTM, Monte Carlo Simulation, NVIDIA, TSLA, MSFT, META, GOOG, Computational Finance

1 Introduction

In modern stock exchange, profitable trading increasingly depends on the ability to fuse heterogeneous information,price micro-structure, news flow, and shifting regimes—into a single adaptive policy. Reinforcement learning (RL) is a natural fit for this sequential decision problem, yet agents often disrupt when market conditions change or when news sentiment reverses direction. To address this, we recently proposed a *hybrid* framework that combines a sliding-window machine learning forecaster for one-minute returns, and an RL policy whose state vector augments price features with sentiment scores, both retrained on a rolling schedule. While an intensive Monte-Carlo campaign on NVIDIA (NVDA) confirms the method's statistical robustness, a broader question remains:

Can a single parameterisation of this hybrid agent generalise across multiple large-cap names that differ in liquidity, news cadence and volatility?

To establish a practical baseline, the model without re-tuning hyper-parameters applied the *same* six additional high-capitalisation tickers: META, INTC, MSFT, GOOGL, AAPL, and TSLA. For each stock, the agent ingested minute by minute prices and aligned news headlines between 12 December 2024 and 19 March 2025, yielding a uniform test set of 77,082 observations enriched with sentiment scores. Table 1 summarises these single-run results.

Table 1. Single-run performance of the proposed strategy versus a buy-and-hold benchmark (initial $10 000 capital).

Ticker	Strategy Return	Buy&Hold Return	NLP Rows	Final Value ($)	Trades
META	92.40%	−3.88%	77 082	19 239.79	542
INTC†	15.73%	30.96%	77 082	11 573.01	274
MSFT	12.66%	0.72%	77 082	11 266.29	216
GOOGL	10.99%	−1.81%	77 082	11 099.47	490
AAPL	9.17%	−3.43%	77 082	10 917.47	432
TSLA	6.18%	−1.85%	77 082	10 617.76	281

† INTC used dummy news data because the original news file was not sufficient compared to other stocks, likely affecting its relative performance.

The hybrid agent outperformed the passive benchmark on five out of seven entities.A particularly strong example of META, which came close to doubling its initial investment, is INTC the single negative outlier; its carefully crafted headlines deprive the model of genuine sentiment indicators highlighting the importance of news sentiment scores. In short, these early cross-asset findings support the functionality of the framework and motivate a deeper Monte Carlo evaluation of the Nvdia stock presented later in the article.

2 Literature Review

Sentiment analysis, reinforcement learning, and machine learning algorithms have become increasingly dominant in the prediction of the direction of the stock price. Zeng and Jiang [1] demonstrated that the combination of a FinBERT pre-trained language model and an LSTM-based deep network in the extraction of sentiment from financial news had higher prediction accuracy than ARIMA, single LSTM, and BERT models. Similarly, Siek and Chandra [2] applied VADER sentiment analysis based on lexicons to financial news and job postings, mapping them to the direction of the stock price and confirming the predictive value of sentiment signals. Nandkumar and Shaik [3] also proposed a spatial federated learning framework that stores news on the blockchain, enhancing the privacy and trustworthiness of sentiment-based stock predictions through decentralized and verifiable processing.

Deep reinforcement learning methods are also surpassing traditional methods in finance. Zou et al. [4] suggested the CLSTM-PPO model that obtained impressive profits in cumulative returns and Sharpe ratios in five major stock markets. Zhang et al. [5] suggested a double deep self-rewarding Q network (SRDDQN) that achieved extremely high cumulative returns - up to 1124% on the IXIC index - incorporating expert reward signals during training, with improved learning stability and profitability. Reinforcement learning systems with sliding window methods also enable adaptively training on recent data segments, with greater responsiveness to market conditions. In Huynh's [6] experiments, a DQN agent using this approach performed better than conventional buy-and-hold and moving-average methods on return, Sharpe ratio, and maximum drawdown.

Machine learning approaches remain central to financial time series forecasting. Our work is closest to Tran et al. [7], who achieved up to 93% prediction accuracy on Vietnamese stock data using LSTM networks and technical indicators SMA, MACD, and RSI, showing LSTM's strength in capturing complex temporal patterns. In another study, Wang [8] explicitly compared LSTM and Random Forest models on Tesla stock data, citing the superiority of LSTM in modeling long-term dependencies with Random Forest offering an advantage in training efficiency. Hybrid models are gaining traction as well—Li et al. [9] combined genetic programming and LSTM for Chinese stock trend prediction, with their SGP-LSTM model generating a 31% annual excess return over the CSI 300 index.

Collectively, they underscore the potential of sentiment analysis, deep reinforcement learning, and advanced machine learning models in significantly improving the accuracy and profitability of stock market prediction.

Table 2. Literature Review

Study	Key Contribution
Ardiansah et al. [10]	News sentiment enables real-time stock trend inference.
Li et al. [11]	Secure and private sentiment analysis system for financial markets.
Yang et al. [12]	RL significantly improves trading decisions vs traditional methods.
Deng et al. [13]	Deep RL enhances stock trading performance.
Zhou et al. [14]	Sliding window improves RL model adaptability.
Zhang et al. [15]	Sliding window boosts DRL model stability.
Patel et al. [16]	Machine learning outperforms classical models in prediction.
Fischer & Krauss [17]	LSTM achieves high prediction accuracy in stock forecasting.

3 Methodology

A supervised one-step-ahead return forecaster with a tabular Q-learning agent that decides whether to *buy*, *sell*, or *hold* a single share in a minute-bar environment.

3.1 Data Alignment

For each stock two synchronous streams were merged: (i) minute mid-prices P_t ($t = 1,\dots,T$) and (ii) news headlines time-stamped in UTC. The latest headline *strictly prior* to the bar close ($t - 15\,\text{min} < \tau < t$) is converted into a sentiment score $S_t \in [0,1]$ (Section 3.2). Missing headlines receive the neutral prior $S_t = 0.5$.

3.2 Sentiment Classification

TF–IDF logistic regressor used on a hand-labelled corpus of $\approx 10^4$ headlines.[1] The predicted posterior $\hat{S}_t = \Pr(\text{positive} \mid \text{headline})$ is clipped to $[0.01, 0.99]$ to prevent extreme rewards.

3.3 Feature Vector

Define the one-minute percentage return

$$r_{t+1} = 100\,\frac{P_{t+1} - P_t}{P_t}. \tag{1}$$

The feature vector $\mathbf{x}_t \in \mathbb{R}^{23}$ stacks *price shocks*, *volatility*, *momentum*, and *sentiment lags*:

$$\mathbf{x}_t = \big[\Delta P_{t:t-5},\, \sigma_t^{20},\, M_t^5,\, S_{t:t-5},\, \bar{S}_t^{10},\, \sigma_{S,t}^{10}\big]^\top. \tag{2}$$

3.4 Return Forecaster

Five different models—OLS, RF, XGBoost, MLP, and LSTM—inside a rolling window of N_{ML} days were trained. Hyper-parameters are tuned on a *pre-tail validation* slice to avoid RL leakage:

$$\boldsymbol{\theta}^\star = \arg\min_{\boldsymbol{\theta}} \text{RMSE}_{\text{val}}\big(f_{\boldsymbol{\theta}}(\mathbf{x}), r\big). \tag{3}$$

The best model outputs $\hat{r}_t = f_{\boldsymbol{\theta}^\star}(\mathbf{x}_t)$.

3.5 State Space

Each minute the environment emits

$$s_t = \big(b_p(\Delta P_t),\, b_s(S_t),\, b_m(\hat{r}_t),\, p_t\big), \tag{4}$$

where b_* maps its argument into $K_* = 5$ equiprobable bins and $p_t \in \{0,1\}$ flags the current long position.

[1] Training details in Appendix A.

3.6 Reward Function

Let P_{entry} be the last fill price and c_a the action-specific transaction cost. For $a_t \in \{\text{Buy}, \text{Sell}, \text{Hold}\}$,

$$r_t^{\text{RL}} = \begin{cases} 100\,\dfrac{P_{t+1}-P_t}{P_t} - c_{\text{buy}} + \beta_m \hat{r}_t + \beta_s S_t, & \text{Buy}, \\ 100\,\dfrac{P_t - P_{\text{entry}}}{P_{\text{entry}}} - c_{\text{sell}} - \beta_m \hat{r}_t + \beta_s (1-S_t), & \text{Sell}, \\ 100\,\dfrac{P_{t+1}-P_t}{P_{\text{entry}}} - c_{\text{hold}} - \lambda h_t, & \text{Hold}. \end{cases} \tag{5}$$

3.7 Q-Learning

The Q-table updates as

$$Q_{t+1}(s_t,a_t) = Q_t(s_t,a_t) + \alpha_t \left[r_t^{\text{RL}} + \gamma \max_{a'} Q_t(s_{t+1},a') - Q_t(s_t,a_t) \right], \tag{6}$$

with $\alpha_t = 0.1$ and $\gamma = 0.95$. Exploration follows an ε-greedy schedule $\varepsilon_t = \max\left(0.01, 0.9 \times 0.995^t\right)$.

3.8 Rolling Training Protocol

For each window

1. Fit $f_{\theta}^{(w)}$ on the past N_{ML} days.
2. Initialise $Q^{(w)} \leftarrow Q^{(w-1)}$ and train on the next N_{RL} days.
3. Freeze both models and evaluate on a *forward* test slice, logging terminal wealth $V_T = \text{cash}_T + n_T P_T$.

Reliability is assessed via 30 Monte-Carlo runs with bootstrap resampling.

4 Data

The response variable is the one-step-ahead percentage change in the closing price, `target_pct_chg`, defined as:

$$\texttt{target\_chg}_t = 100\,\frac{Ct+1-C_t}{C_t}, \tag{7}$$

where C_t is the closing price at time t.

In addition to the target, features from price data and sentiment scores were extracted.

Sentiment Score

`nlp_sentiment`: continuously valued sentiment score (0–1 range), retrieved via Polygon.io's Ticker News API. The underlying methodology employs an LLM-powered pipeline to extract company-level sentiment from news headlines in real time [**Dolphin2024**]. This approach enables richer per-ticker sentiment representation compared to simple thresholded or binary models.

Price-based Features

$$\texttt{current\_pct\_chg}_t = 100\,\frac{C_t - C_{t-1}}{C_{t-1}}, \tag{8}$$

$$\texttt{volatility}_t = \mathrm{std}\big(\texttt{current\_chg}t - 19 : t\big), \tag{9}$$

$$\texttt{momentum}_t = 100\,\frac{C_t - C_{t-5}}{C_{t-5}}. \tag{10}$$

Relative Strength Index (RSI)

$$\Delta_t = C_t - C_{t-1}, \tag{11}$$

$$\texttt{gain}_t = \max(\Delta_t, 0), \qquad \texttt{loss}_t = \max(-\Delta_t, 0), \tag{12}$$

$$\texttt{avg\_gain}_t = \mathrm{mean}\big(\texttt{gain}_{t-13:t}\big), \tag{13}$$

$$\texttt{avg\_loss}_t = \mathrm{mean}\big(\texttt{loss}_{t-13:t}\big), \tag{14}$$

$$\mathrm{RS}_t = \frac{\texttt{avg\_gain}_t}{\texttt{avg\_loss}_t}, \tag{15}$$

$$\mathrm{RSI}_{14,t} = 100 - \frac{100}{1 + \mathrm{RS}_t}. \tag{16}$$

MACD and Signal Line

$$\texttt{EMA\{12,t\}} = \mathrm{EMA}_{12}(C_t), \qquad \texttt{EMA\{26,t\}} = \mathrm{EMA}_{26}(C_t), \tag{17}$$

$$\texttt{MACD\_t} = \texttt{EMA\{12,t\}} - \texttt{EMA\{26,t\}}, \tag{18}$$

$$\texttt{MACD\_t} = \mathrm{EMA}\,9\big(\texttt{MACD\_t}\big), \tag{19}$$

$$\texttt{MACD\_hist\_t} = \texttt{MACD\_t} - \texttt{MACD\_signal\_t}. \tag{20}$$

Bollinger Bands

$$\texttt{MA\{5,t\}} = \mathrm{mean}\big(C_{t-4:t}\big), \qquad \texttt{MA\{15,t\}} = \mathrm{mean}\big(C_{t-14:t}\big), \tag{21}$$

$$\texttt{ma\_ratio\_t} = \frac{\texttt{MA\{5,t\}}}{\texttt{MA\{15,t\}}}, \tag{22}$$

$$\texttt{BB\_upper,t} = \texttt{MA\{15,t\}} + 2\,\mathrm{std}\big(C_{t-14:t}\big), \tag{23}$$

$$\texttt{BB\_lower,t} = \texttt{MA\{15,t\}} - 2\,\mathrm{std}\big(C_{t-14:t}\big). \tag{24}$$

Lag and Return Features

,

cum_return{15,t} = $\left(\prod_{i=t-14}^{t}(1+r_i)\right)-1$,

macd_hist_slope_t $=$ MACD_t $-$ MACD_hist{t-1,

log_return_t $= \ln\left(\frac{C_t}{C_{t-1}}\right)$.

Signal Feature

– hold_signal: true when (MACD_hist_t $> 0) \wedge ($macd_hist_slope_t $> 0)$.

Signal Feature

– hold_signal: true when $(\texttt{MACD\_hist}_t > 0) \wedge (\texttt{macd\_hist\_slope}_t > 0)$.

These engineered variables serve as inputs to the predictive modelling framework.

5 Findings

Using minute-bar price data in combination with news sentiment scores between December 19, 2024, and March 19, 2025, a sentiment-augmented reinforcement learning (RL) agent based on sliding-window adaptation proved highly effective for high-frequency stock trading. The approach was tested on six technology stocks and managed to outperform the buy-and-hold benchmark in five instances (namely, MSFT, GOOGL, AAPL, TSLA, and META), with META recording the highest return of +92.40% against a benchmark loss of 3.88%. The only stock that faltered, INTC, was based on surrogate news data, thus highlighting the absolute importance of genuine sentiment signals. The requirement of the sliding-window component was also validated by ablation testing on NVIDIA stock, which saw an average loss of around 25% when this was removed. Also, a Monte Carlo simulation over 30 iterations on NVDA resulted in a high mean weekly return of +14.61% further attesting to the generalizability of the proposed approach. The number of trades placed varied from 216 (for MSFT) to 542 (for META), indicating the ability of the model to adjust to asset-specific volatility as well as its proficiency for capturing short-term trading opportunities. Overall, the results support the supposition that reinforcement learning, augmented by sliding-window adaptation and real-time sentiment analysis, forms a powerful and generalizable framework for adaptive algorithmic trading.

6 Conclusion and Future Directions

Sentiment-aware reinforcement learning approach for high-frequency trading with a sliding-window model, adaptive Q-learning, and real-time news sentiment. The experiments proved, it performs significantly better compared to baseline approaches on big-cap equities, with a +92.4% return on META. Monte Carlo simulations and ablation studies highlighted the importance of model updates and changes in sentiment. The hybrid model has potential in unstable markets. Additional studies specified under live market conditions can advance the research by extending evaluation of the model architecture.

References

[1] Qingyun Zeng and Tingsong Jiang. "Financial sentiment analysis using FinBERT with application in predicting stock movement". In: *arXiv preprint arXiv:2306.02136* (2023).

[2] Michael Siek and Vincentius David Chandra. "Analysis of News Sentiment for Stock Price Prediction Using VADER Sentiment". In: *2024 6th International Conference on Cybernetics and Intelligent System (ICORIS)*. 2024, pp. 1–6. DOI: 10.1109/ICORIS60231.2024.10549729.

[3] Nitin Nandkumar Sakhare and Imambi S. Shaik. "Spatial federated learning approach for the sentiment analysis of stock news stored on blockchain". In: *Spatial Information Research* 32.1 (2024), pp. 13–27. DOI: 10.1007/s41324-023-00529-x.

[4] Yubo Zou et al. "A Novel Deep Reinforcement Learning Based Automated Stock Trading System Using Cascaded LSTM Networks". In: *arXiv preprint arXiv:2212.02721* (2022).

[5] Yuling Zhang et al. "A Self-Rewarding Mechanism in Deep Reinforcement Learning for Trading Strategy Optimization". In: *Mathematics* 12.8 (2024), p. 1124. DOI: 10.3390/math12081124.

[6] Nhat-Thien Huynh et al. "A Sliding Window-based Approach using Deep Q-Network for Efficient Stock Trading". In: *Proceedings of the 2023 5th International Conference on Machine Learning and Natural Language Processing*. 2023, pp. 1–7.

[7] Phuoc Tran et al. "Applying machine learning algorithms to predict the stock price trend in the stock market – The case of Vietnam". In: *Scientific Reports* 14.1 (2024), p. 4633. DOI: 10.1038/s41598-024-55245-z.

[8] Yuhan Wang. "A Comparative Study of LSTM and Random Forest in Stock Price Prediction". In: *Proceedings of the 2024 International Conference on Advanced Algorithms and Signal Image Processing (AASIP 2024)*. Vol. 254. Advances in Engineering Research. 2024, pp. 167–172. DOI: 10.2991/978-94-6463-398-6_26.

[9] Jiaqi Li et al. "A Hybrid SGP-LSTM Model for Stock Trend Prediction". In: *International Journal of Financial Studies* 12.1 (2024), p. 25. DOI: 10.3390/ijfs12010025.

[10] Muhammad Refa Ardiansah, Salman Al Ayub, and Anisa Aditsania. "News Sentiment Analysis for Stock Trend Prediction in Real-time". In: *2021 International Conference on ICT for Smart Society (ICISS)*. 2021, pp. 1–6. DOI: 10.1109/ICISS53185.2021.9533221.

[11] Xiang Li et al. "A Secure and Privacy-Preserving Sentiment Analysis Framework for Financial Markets". In: *IEEE Transactions on Knowledge and Data Engineering* 35.7 (2023), pp. 7164–7177. DOI: 10.1109/TKDE.2022.3168855.

[12] Haoran Yang, I-Cheng Liu, and Wei-Yang Lin. "A Deep Reinforcement Learning Approach for Algorithmic Trading". In: *IEEE Access* 8 (2020), pp. 177435–177447. DOI: 10.1109/ACCESS.2020.3026844.

[13] Yifei Deng et al. "Deep direct reinforcement learning for financial signal representation and trading". In: *Proceedings of the 22nd international conference on Multimedia modeling*. 2016, pp. 237–248.

[14] Zinuo Zhou et al. "Sliding-Window based Reinforcement Learning for Stock Trading". In: *arXiv preprint arXiv:2205.15858* (2022).

[15] Zhipeng Zhang et al. "Enhancing Deep Reinforcement Learning for Stock Trading with Sliding Window". In: *IEEE Transactions on Neural Networks and Learning Systems* 32.8 (2021), pp. 3507–3519. DOI: 10.1109/TNNLS.2020.3026369.

[16] Jigar Patel and Sah Shah. "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques". In: *Expert Systems with Applications* 42.1 (2015), pp. 259–268. DOI: 10.1016/j.eswa.2014.07.040.

[17] Thomas Fischer and Christopher Krauss. "Deep learning with long short-term memory for financial market predictions". In: *European Journal of Operational Research* 270.2 (2018), pp. 654–669. DOI: 10.1016/j.ejor.2017.11.054.