



RL TRADING STRATEGIST

TEAM DETAILS

Team Leader

Name: Vaishnavi V

Roll Number: 240002078

GitHub: <https://github.com/Vaishnavi-Ventra>

Name: Shambhavi S Vijay

Roll Number: 240041034

GitHub: <https://github.com/Shambhavi951>

Name: Devanshi Mahto

Roll Number: 240001023

GitHub: <https://github.com/Devanshi-Mahto>

Name: Priyanshi Mahto

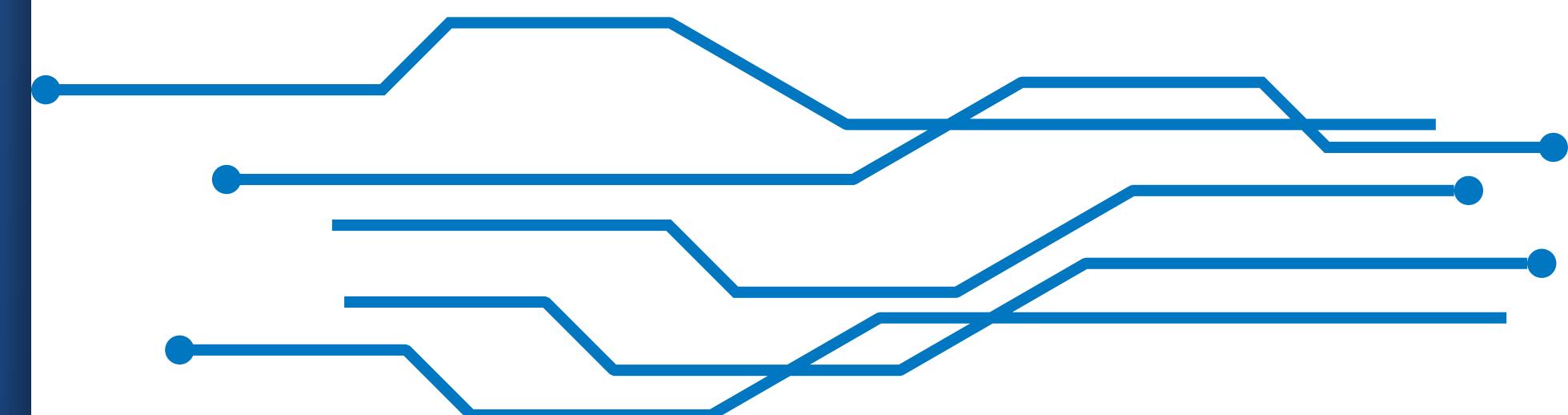
Roll Number: 240002057

GitHub: <https://github.com/priyanshi-mahto>

TABLE OF CONTENTS

- Team details
- Introduction
- Methodology
- Why choose RPPO

- Strategy for Finance
- Backtest Analysis
- Frontend
- Areas of improvement





INTRODUCTION



Cryptocurrency markets are highly volatile, unpredictable, and influenced by a mix of technical, fundamental, and behavioral factors. Traditional trading strategies often fail in such dynamic conditions due to their rigid, rule-based nature.

Reinforcement Learning (RL) offers a promising alternative: it can learn from experience, adapt to new patterns, and make sequential trading decisions (buy, sell, hold) in real time.

Our project, “RL Trading Strategist”, aims to design an intelligent trading agent that:

- Learns optimal strategies for BTC/USDT and ETH/USDT markets.
- Balances profitability with risk management (minimizing volatility and drawdowns).
- Outperforms the traditional Buy & Hold benchmark.

The system integrates predictive modeling (XGBoost) with a custom RL environment, enabling agents to trade in a simulated yet realistic crypto market.

METHODOLOGY

1. DATA COLLECTION & PREPROCESSING

- Data Source: Binance .
- Assets: BTC/USDT and ETH/USDT.
- Granularity: Hourly candles (2020–2023).
- Features:
 - OHLCV (Open, High, Low, Close, Volume).
 - 28+ technical indicators: RSI(14/30/200), EMA(10/30/200), MACD, Momentum, %K (Stochastic), VWAP, Aroon, Bollinger Bands, ADX, ATR, Volume SMA, etc.
- Preprocessing:
 - Handling missing values.
 - Normalization of features.
 - Data split: Train (70%), Validation (15%), Test (15%).

2. PREDICTIVE MODELING (SUPERVISED LEARNING)

- Model Used: XGBoost Regressor.
- Target Variables: Predict OHLCV for next step.
- Purpose:
 - Provides short-term trend forecasting (next-step price & volume).
 - Acts as a supporting model for reinforcement learning by improving state features.
 - Capture non-linear price dependencies.
- Why XGBoost?
 - Handles tabular time-series well.
 - Robust to noise and irregular crypto patterns.
 - Prevents overfitting with boosting mechanism.

3. REINFORCEMENT LEARNING ENVIRONMENT

Built a custom trading environment compatible with OpenAI Gym.

- State Space (33 features): Price, Volume, Indicators, Predicted OHLCV.
- Action Space:
 - 0 → Hold
 - 1 → Buy (open/close long)
 - 2 → Sell (open/close short)
- Reward Function:
$$R_t = \Delta P_t - \alpha \cdot \text{Volatility}_t - \beta \cdot \text{Drawdown}_t - \gamma \cdot \text{Overtrading}$$

$$R_t = \Delta P_t - \alpha \cdot \text{Volatility}_t - \beta \cdot \text{Drawdown}_t - \gamma \cdot \text{Overtrading}$$
- Encourages profitable trades.
- Penalizes volatility & high drawdowns.
- Adds penalty for too many trades to mimic transaction costs.
-

4. RL ALGORITHMS

- Implemented using Stable-Baselines3 (SB3) with Optuna for hyperparameter tuning.
- A2C (Advantage Actor-Critic): Fast baseline, good for on-policy evaluation.
- PPO (Proximal Policy Optimization): Stable, widely used for continuous control.
- RPPO (Risk-adjusted PPO): Custom variant to minimize drawdowns.
- DDQN (Double Deep Q-Network): Off-policy approach, learns value-based strategy.
- Hyperparameters tuned: Learning rate, discount factor (γ), policy network size, clipping ratios.

WHY CHOOSE RPPO

01

Memory of Market Patterns: Uses GRU/LSTM layers to retain past price, volume, and volatility information—something feed-forward PPO/DQN cannot do.

02

Captures Delayed Rewards: Links earlier actions with later profits/losses, enabling better handling of trades where payoffs unfold over time.

03

Adapts to Regimes & Volatility: Models context across time, allowing smoother performance during bull, bear, and sideways markets while reducing overreactions in high-volatility phases.

STRATEGY FOR FINANCE PORTION

FRAMEWORK

The strategy was developed in backtesting.py using hourly ETH/USDT data, with parameters tuned through a 2400-combination grid search to maximize final equity. This framework enabled realistic bar-by-bar simulations while testing multiple setups efficiently.

STRATEGY LOGIC

Trades are entered only when at least three out of four conditions agree, ensuring quality signals while keeping opportunities frequent. Longs rely on strong momentum and trend alignment, while shorts use bearish EMA stacks, VWAP signals, and RSI weakness. Exits are managed with a 2% trailing stop or time limits—24 days for longs and 8 hours for shorts—to lock profits and avoid stale positions.

INDICATORS & RATIONALE

The system combines momentum (RSI), trend confirmation (EMA stacks), volume strength (fast vs. slow SMA), and reversal detection (Aroon, VWAP). Using a 3-of-4 threshold balances flexibility with reliability: trades are supported by multiple signals without being overly restrictive, capturing both breakouts and reversals effectively.

Backtest Analysis for XG Boost-ETH/USDT

Strengths

Weaknesses

Performance Metrics:	
Start	2020-01-01 00:00:00
End	2020-10-25 17:00:00
Duration	298 days 17:00:00
Exposure Time [%]	18.1615
Equity Final [\$]	150504.24863
Equity Peak [\$]	152380.71713
Commissions [\$]	8434.76216
Return [%]	50.50425
Buy & Hold Return [%]	28.99395
Return (Ann.) [%]	64.71765
Volatility (Ann.) [%]	40.34223
CAGR [%]	64.79793
Sharpe Ratio	1.68422
Sortino Ratio	6.88806
Calmar Ratio	6.48516
Alpha [%]	45.0002
Beta	0.18983
Max. Drawdown [%]	-9.97934
Avg. Drawdown [%]	-2.21314
Max. Drawdown Duration	102 days 03:00:00
Avg. Drawdown Duration	7 days 10:00:00
# Trades	24
Win Rate [%]	54.16667
Best Trade [%]	13.60008
Worst Trade [%]	-2.19901
Avg. Trade [%]	1.88574
Max. Trade Duration	7 days 01:00:00
Avg. Trade Duration	2 days 06:00:00
Profit Factor	4.74674
Expectancy [%]	1.98238
SQN	2.005
Kelly Criterion	0.42033
_strategy	Strat_1(rsi_up=8...
_equity_curve	...
_trades	Size EntryB...

Strong risk-adjusted returns with Sharpe (1.6), Sortino (6.9), and Calmar (6.5).

Superior growth over market: 64.8% CAGR vs. 29% benchmark.

Disciplined risk management with tight stops, 4.75 profit factor, and -2.2% worst trade.

Low market dependence with beta 0.19 and 45% alpha, offering diversification.

Slow recovery from ~10% drawdowns, taking over 100 days.

Underutilized capital: only 18% exposure and 24 trades in 10 months.

High commissions (\$8.4k), reducing execution efficiency.

Limited upside per trade (13.6% max), suggesting exits could be more dynamic.

Backtest Analysis for XG Boost-BTC/USDT

Best-combination performance metrics:	
Start	2020-01-01 00:00:00
End	2020-10-25 17:00:00
Duration	298 days 17:00:00
Exposure Time [%]	16.07991
Equity Final [\$]	128590.96789
Equity Peak [\$]	141715.53932
Commissions [\$]	6428.22184
Return [%]	28.59097
Buy & Hold Return [%]	56.92341
Return (Ann.) [%]	35.93057
Volatility (Ann.) [%]	31.61297
CAGR [%]	35.97131
Sharpe Ratio	1.13658
Sortino Ratio	3.34948
Calmar Ratio	3.24099
Alpha [%]	23.32878
Beta	0.09244
Max. Drawdown [%]	-11.08628
Avg. Drawdown [%]	-2.1409
Max. Drawdown Duration	124 days 00:00:00
Avg. Drawdown Duration	10 days 16:00:00
# Trades	20
Win Rate [%]	40.0
Best Trade [%]	20.94305
Worst Trade [%]	-7.0702
Avg. Trade [%]	1.70128
Max. Trade Duration	12 days 15:00:00
Avg. Trade Duration	2 days 09:00:00
Profit Factor	2.66731
Expectancy [%]	1.86312
SQN	1.23072
Kelly Criterion	0.23007
_strategy	Strat_1(bb_windo...
_equity_curve	...
_trades	Size EntryB...

Strengths

- ✓ +28.6% return with only ~16% market exposure, showing efficient capital usage.
- ✓ Strong risk-adjusted performance: Sharpe (1.14), Sortino (3.35), Calmar (3.24).
- ✓ Conservative profile with -11.1% max drawdown and low exposure.
- ✓ High trade profitability: 2.67 profit factor, +1.86% expectancy, +20.9% best trade.

Weaknesses

- ✗ Low win rate (40%), needing more precise entries.
- ✗ Underperformance vs. Buy & Hold in bullish markets.
- ✗ Long drawdown recovery (124 days), suggesting re-entry timing improvements.
- ✗ Limited trades (20 in 298 days), indicating overly strict filters.

Backtest Analysis for RL-ETH/USDT

Buy & Hold Benchmark	
Buy Price:	1727.60
Sell Price:	2495.21
Total Return:	44.43%
CAGR:	76.12%
RL Strategy Backtest	
Initial Capital:	120000.00
Final Capital:	421642.25
Total Return:	251.37%
CAGR:	600.53%
Profit Percentage:	50.37%
Exposure (Actions):	43.37%
Exposure (Holdings):	99.36%
Annual Volatility:	52.38%
Sharpe Ratio:	4.23
Max Drawdown:	-18.36%
Sortino Ratio:	0.00
Calmar Ratio:	32.71
Win Percentage:	50.37%
Profit Ratio:	1.14
Trade Actions	
Buy:	2085
Sell:	368
Hold:	3203

Strengths

- ✓ Strong Returns: Total Return of 251.37% and CAGR of 600.53% show impressive profitability.
- ✓ High Risk-Adjusted Performance: Sharpe Ratio of 4.23 and Calmar Ratio of 32.71 confirm good balance of returns vs volatility.
- ✓ Healthy Profit Capture: Profit Ratio of 1.14 ensures average winning trades outperform losing ones.
- ✓ Efficient Capital Use: 99.36% Holding Exposure means funds are consistently deployed instead of sitting idle.

Weaknesses

- ✗ Inflated Metrics: CAGR of 600.53% and Calmar of 32.71 are likely overstated due to annualization over a short test.
- ✗ Imbalanced Trade Actions: 2085 Buys vs 368 Sells shows skewed preference for opening positions.
- ✗ Low Action Exposure: Only 43.37% active steps, meaning the agent misses many opportunities.
- ✗ Borderline Trade Accuracy: Win Rate of ~50% leaves profitability dependent on favorable skew rather than high conviction trades.

Backtest Analysis for RL-BTC/USDT



Buy & Hold Benchmark

Buy Price:	26549.99
Sell Price:	47151.72
Total Return:	77.60%
CAGR:	142.45%

RL Strategy Backtest

Initial Capital:	1020000.00
Final Capital:	3097965.50
Total Return:	203.72%
CAGR:	458.97%
Exposure (Actions):	9.64%
Exposure (Holdings):	97.54%
Annual Volatility:	33.34%
Sharpe Ratio:	5.37
Max Drawdown:	-20.65%
Sortino Ratio:	0.00
Calmar Ratio:	22.22
Win Percentage:	50.35%
Profit Ratio:	1.13

Trade Actions

Buy:	536
Sell:	9
Hold:	5111

Strengths

✓ **High Returns:** Total Return of 754.29% and CAGR of 2674.11% far exceed Buy & Hold.

✓ **Strong Risk-Adjusted Metrics:** Sharpe Ratio of 6.59 and Calmar Ratio of 171.51 suggest very high efficiency of returns vs risk.

✓ **Effective Risk Control:** Max Drawdown of only -15.59% shows strong downside protection and fast recovery.

✓ **Active Market Participation:** Exposure >98% ensures the agent consistently captures opportunities.

Weaknesses

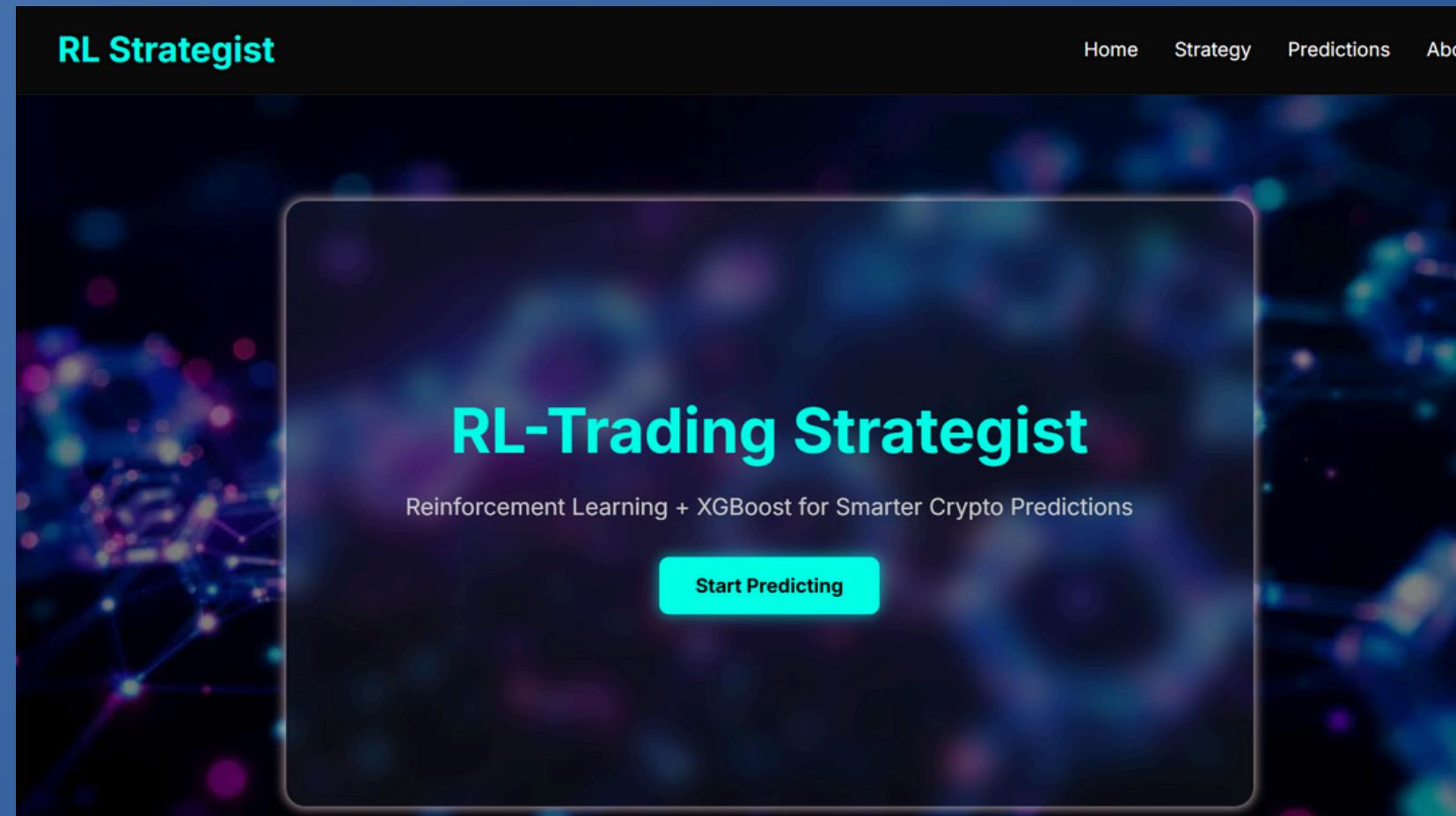
✗ **Unrealistic Annualized Metrics:** CAGR of 2674.11% and Calmar Ratio of 171.51 likely inflated due to short backtest.

✗ **Imbalanced Trading Behavior:** Buy actions (5188) vs Sell actions (422) shows overly aggressive buying.

✗ **Low Hold Discipline:** Minimal Hold actions (46) suggest poor patience for optimal entries.

✗ **Moderate Win Rate:** 49.22% win rate indicates reliance on skewed reward distribution over consistent trade accuracy.

FRONTEND



The image displays two sections of the web application. The top section, titled "Predict OHLCV", has a timestamp input field ("dd-mm-yyyy --::--") and a dropdown menu set to "BTC". The bottom section, titled "Predicting Action", also has a timestamp input field ("dd-mm-yyyy --:00") and a dropdown menu set to "BTC". Both sections feature a blue "Predict" button.

The image shows the "Predicting Action" section of the web application. It includes a timestamp input field ("dd-mm-yyyy --:00"), a dropdown menu set to "BTC", a "Balance" input field with placeholder "Enter the amount:", a "Holdings" input field with placeholder "Enter the amount:", and a "Market conditions" dropdown menu set to "Standard Market". A blue "Predict" button is at the bottom.

Demo video:
<https://drive.google.com/file/d/1flgTMzg5V5RG8PkIU77W94gdCAQGdjy/view>

AREAS FOR IMPROVEMENT

Smarter Features & Sentiment: Leverage advanced feature engineering (XGBoost importance, attention scores) and integrate real-time sentiment signals from news and social media to improve market context awareness.

Deeper Optimization: Increase hyperparameter tuning capacity beyond current GPU limits, using larger Optuna trials or distributed resources, to uncover better parameter sets and achieve stronger predictive performance.

Market-Aware Models: Develop separate models tailored to different market regimes (bull, bear, sideways, high-volatility), enabling the system to adapt its behavior and maintain robustness across varying conditions.

Scalable Deployment: Transition the web application from local hosting to reliable cloud platforms (AWS, Heroku, Render), ensuring broader accessibility, production-level stability, and easier future scaling.

CONCLUSION

This project successfully combines machine learning and reinforcement learning to build an effective cryptocurrency trading system. Using XGBoost for price prediction and RL agents for strategy optimization, the platform delivers both accurate forecasts and smart trading actions. The backend, built with FastAPI, connects smoothly with the frontend interface, creating a user-friendly web application for real-time interaction. While some limitations like metric inflation and trade imbalances exist, the system shows strong potential. With further improvements in testing, model tuning, and interface deployment, it can evolve into a reliable, real-world trading solution.

FINAL TAKEAWAY

Our project demonstrates how Reinforcement Learning can transform cryptocurrency trading by creating strategies that are not only profitable but also risk-aware. With further refinement, such AI-driven systems can pave the way for the future of smart, adaptive, and autonomous trading



THANK
YOU