KHARAGPUR DATA SCIENCE HACKATHON

Zelta Labs

ZELTA TECH'S PROBLEM STATEMENT

Problem Statement: Algorithmic Trading Model Development for BTC/USDT Crypto Market

DOMAIN DESCRIPTION

Algorithmic trading in the cryptocurrency market involves deploying machine learning (ML) models and computer programs, integrating financial market expertise, data analysis, statistical modeling, and programming skills. *The objective is to unlock the potential of ML-based algorithmic trading, aiming to consistently generate robust returns, outperform benchmarks, and safeguard capital in the dynamic cryptocurrency market.*

Participants are encouraged to leverage their ML expertise to develop explainable models that operate with soundness, enabling traders to adapt to market dynamics and seize opportunities with precision and agility. *This challenge not only recognizes the transformative impact of ML in cryptocurrency trading but to navigate market fluctuations effectively and capitalize on opportunities with data-driven ML based solutions to provide alpha opportunities.*

PROBLEM DESCRIPTION

The problem consists of developing algorithmic trading models for the BTC/USDT cryptocurrency market, aiming to outperform benchmark returns. Participants are required to create trading algorithms that can generate returns while managing risk effectively in the specific BTC/USDT market. The problem can be divided into the following tasks:

1. Data Acquisition: Participants are advised to use historical price and trading volume data for the BTC/USDT trading pair from January 1, 2018, to January 31, 2022, for model development and testing.

2. Data Preprocessing: Clean and preprocess the acquired data, with a specific focus on the provided timeframes (January 1, 2018, to January 31, 2022), to ensure its quality and compatibility for analysis. Handle missing data, adjust for splits, and perform any necessary data transformations within this defined period.

3. Model Design:- Design algorithmic trading models that utilize statistical and mathematical models tailored to the BTC/USDT market. *This may include trend-following, mean-reversion, momentum centric, machine learning, or other quantitative/statistical approaches.*

4. Backtesting:- Implement the models and conduct extensive backtesting using historical data of the BTC/USDT market to assess their performance. *Participants should take into account transaction costs and slippage in their simulations. These costs are to be considered at a rate of 0.15 percent per transaction.*

5. Risk Management:- Develop risk management rules and mechanisms specific to the BTC/USDT market to protect capital and reduce drawdowns.

6. Optimization:- Fine-tune the models to maximize returns while maintaining acceptable risk levels in the BTC/USDT market. Participants may need to adjust parameters and rules based on the backtesting results.

DATA AND RESOURCES

Participants can use publicly available cryptocurrency market data sources, API services, or simulated data for BTC/USDT. Access to historical data for the BTC/USDT trading pair from January 1, 2018, to January 31, 2022, will be provided or suggested as a resource.

DELIVERABLES

Participants should submit the following deliverables:-

1. Algorithm Code:- The source code of the algorithmic trading models developed for the BTC/USDT market, along with documentation explaining the logic and parameters used. *Code submissions must be made using Google Colab Notebook, Jupyter Notebook, or any other web-based interactive computing platform on Zelta Labs' GitHub repository by a specified deadline.*

2. Backtesting Results:- Detailed backtesting results specific to the BTC/USDT market, using the four-year historical data, *including performance metrics (e.g., Sharpe ratio, annualized returns, maximum drawdown) and visualizations (e.g., equity curve, trade history). The objective is not just to generate profits but to do so with statistical soundness and robustness.*

3. Risk Management Plan:- A clear description of the risk management mechanisms specific to the BTC/USDT market, *including stop-loss rules, and risk-reward ratios.*

4. Presentation:- A presentation summarizing the model, its rationale, backtesting results, and risk management approach, all tailored to the BTC/USDT market.

5. Report:- A comprehensive report detailing the development process, model optimization, and any insights gained during the project, with a focus on the BTC/USDT market.

APPROACHES TO GO AHEAD WITH

Some other common approaches in Algorithmic Trading that can be employed for this competition:-

1. Time series models like ARIMA , STL , LSTM networks etc.
2. Machine learning models like regression models, decision trees, random forests, support vector machines etc.
3. Deep learning models, such as recurrent neural networks and convolutional neural networks etc.
4. Reinforcement learning techniques, such as Q-learning or deep reinforcement learning etc.
5. Technical analysis using various indicators or combinations of indicators to generate noise free signals etc.
6. Various miscellaneous approaches like bagging and boosting for accurate prediction , monte carlo methods or stochastic models like Markov chains etc.

JUDGING CRITERIA

The judging criteria will be based on the following aspects, with a focus on the BTC/USDT market:-

1. Performance (35 points):- Evaluation of the model's historical performance in the BTC/USDT market, including risk-adjusted returns, consistency, and drawdowns

ASSET CLASS	AVERAGE ANNUAL RETURNS	MAXIMUM DRAWDOWNS
BTC/USDT	100%	1.5%
BTC/USDT	150%	2%
BTC/USDT	300%	4%
BTC/USDT	450%	6%
BTC/USDT	600%	8%

In this competition, crafting algorithmic trading models closely aligned with given parameters—annual returns and drawdowns—is key.
Aim for the sweet spot: impressive returns and prudent risk management. *Models inherently robust, explainable and interpretable, not exceeding 7-8 percent drawdowns, define excellence for us and participants devising these models will have an inherent advantage in the competition.*

2. Model Logic (20 points):- Assessment of the soundness and effectiveness of the mathematical and statistical models used in the model, with a focus on BTC/USDT market dynamics.

3. Risk Management (20 points):- Evaluation of the risk management mechanisms specific to the BTC/USDT market and their effectiveness in preserving capital.

4. Code Quality and Data Visualization (10 points):- Review of the clarity, organization, and efficiency of the code tailored to the BTC/USDT market. We will also assess the implementation of Object-Oriented Programming (OOP) principles, which promote code modularity and maintainability.

5. Presentation (10 points):- Assessment of the quality and clarity of the presentation materials in English, focusing on the BTC/USDT market at offline events.

6. Report (5 points):- Evaluation of the completeness and depth of the written report, with a specific emphasis on insights and analysis related to the BTC/USDT market.

GENERAL CONCERNS

1. MACHINE LEARNING CONSIDERATIONS:-

Explainability and Interpretability:

- Machine learning models, while potentially profitable on training data, often lack explainability and interpretability. This can lead to challenges in live trading like failure of producing same results on out-of-sample scenarios.

Data Sets and Noble Suggestions:

- Teams are provided with specific data sets and noble suggestions to address these challenges:

1. Training, Validation, and Test Data (Jan 1, 2018, to Jan 31, 2022) - Teams should use this period for training, validation, and testing their models.

2. Out-of-Sample 1 (Feb 1, 2022, to Dec 31, 2022)
   - A separate dataset is provided for teams to assess their model's performance in an out-of-sample scenario.

3. Out-of-Sample 2 (Jan 1, 2023, to Dec 31, 2023)
   - Another dataset is supplied for teams to evaluate their model's generalization capabilities further.

2. ITERATION AND MODEL OPTIMIZATION:-

Retry or Re-structuring Allowed:

- Teams are encouraged to iterate and optimize their models after each attempt. This flexibility allows refinement to avoid submitting an overfit model in the final submission.

3. MONITORING CODEBASE AND MODEL EVALUATION:-

Codebase Monitoring:

- All codebases will be closely monitored. Teams are advised to focus on solutions that prioritize the "readability and modularity" of their code.

Strategic Evaluation Point:

- The ability of models to offer explanations for their decisions becomes a strategic evaluation point, emphasizing practicality in live trading environments.

4. OUT SAMPLES FOR GENERALIZATION TESTING:-

Outsamples 1 and 2:

- These datasets (Out-of-Sample 1 and Out-of-Sample 2) will not be involved in any training, validation, or testing processes. Instead, the final model, post-training, will be tested on these out samples to assess its generalization capabilities on unseen data.

5. AVOIDING DATA PEEK FOR HYPERPARAMETER TUNING(DATA LEAKAGE):-

Peeking at the test data/out of samples for hyperparameter tuning is discouraged, as it can provide an unrealistic estimate of generalization performance. Teams are challenged to stimulate intellectual depth, ensuring their machine learning models generalize to unseen data.

6. DATA FREQUENCY LIMITATION:-

Teams are restricted to working with dataset frequencies up to 6 hours. The provided dataset spans from January 1, 2018, to January 31, 2022, with time intervals ranging from 3 minutes to 6 hours. This limitation is in place to align the competition with a realistic trading scenario and to emphasize the importance of models that can adapt to lower frequency data.

7. PRIORITIZE INTEGRATED APPROACHES AND XAI:-

Integrated Approach;-

In this competition, we encourage teams to leverage both statistical models and machine learning models *synergistically to develop sophisticated, noise-free alpha-producing data-driven solutions for algorithmic trading.*

The objective is to harness the strengths of statistical methods in capturing market dynamics and the predictive power of machine learning to enhance the robustness and adaptability of trading algorithms.
We look forward to witnessing how teams integrate these methodologies to deliver sophisticated, noise-free, data-driven alpha models.

Explainable AI / Explainable Machine Learning (XML)
XAI becomes a mechanism for participants to retain intellectual oversight over AI-driven trading models. This empowers users to comprehend the decision-making processes, fostering a collaborative environment where human expertise synergizes with AI capabilities. Trust is paramount in algorithmic trading, and *XAI plays a crucial role in fostering it. When participants can follow the reasoning of the trading algorithm—understanding what has been done, what is happening, and what will happen next—they are more likely to trust the system's capacity to make sound decisions.*

8. TRY STRIKING A BALANCE BETWEEN ACCURACY AND EXPLAINABILITY:-

While innovation is encouraged, teams are advised to exercise caution with machine learning approaches. *If teams opt for machine learning models, they should prioritize the "explainability and interpretability" of their models while making profitable models.* This is crucial for real-world implementation as overly complex models might not be practically scalable in live trading environments.

In the pursuit of implementing profitable ML models for algorithmic trading, *use approaches like state-of-the-art explainability techniques, including SHAP, model-agnostic methods such as LIME, Partial Dependence Plots (PDP) etc. We emphasize the significance of these techniques in not only enhancing model explainability and transparency but also contribute to the continual advancement of explainable AI in the finance domain, facilitating informed decision-making for algorithmic trading strategies.*

Explore strategies to improve generalization including reducing capacity, early stopping, weight decay, ensembles, input transformations, and stochastic regularization.

KHARAGPUR DATA SCIENCE HACKATHON

REQUESTED METRICS FOR MODEL:-

We require participants to organize and assess their trading models based on a set of comprehensive metrics.
These metrics provide a multifaceted view of each model's performance and risk characteristics, simplifying the evaluation of its robustness and effectiveness.
This structured approach allows us for a thorough analysis and comparison of trading models.

Certainly, here are the metrics presented in numerical order:

1. Gross Profit
2. Net Profit
3. Total Closed Trades
4. Win Rate (Profitability %)
5. Max Drawdown
6. Gross Loss
7. Average Winning Trade (in USDT)
8. Average Losing Trade (in USDT)
9. Buy and Hold Return of BTC
10. Largest Losing Trade (in USDT)
11. Largest Winning Trade (in USDT)
12. Sharpe Ratio
13. Sortino Ratio
14. Average Holding Duration per Trade
15. Max Dip and Average Dip in Running Trade

HYPOTHETICAL EXAMPLE OF ABOVE COMPREHENSIVE METRICS:-

Model Description - A mean-reversion model aims to profit from the tendency of BTC/USDT prices to revert to their mean or average price. When the price moves significantly away from the mean, the model opens a trade with the expectation that the price will return to the mean.

Certainly, let's analyze the mean-reversion trading model for BTC/USDT using the provided metrics, considering a starting capital of $100,000.

1. Gross Profit: $50,000
   - This represents the total profit generated from all closed trades. It indicates that the model was successful in capturing profits from price fluctuations.

2. Net Profit: $47,000
   - After accounting for fees, commissions, and brokerage costs, the model generated a net profit of $47,000. This is the actual amount you would have made.

3. Gross Loss: -$10,000
   - The model also incurred losses, totaling $10,000. This is a normal part of trading, but the model managed to outweigh losses with profits.

4. Max Drawdown: -$7,000
   - The maximum drawdown represents the largest loss experienced during the trading period. In this case, the model's worst drawdown was -$7,000.

5. Buy and Hold Return of BTC: 20%
 - The model's performance is compared to a simple buy and hold approach. If you had invested your initial $100,000 in BTC i first trade and held it throughout the test period, you would have achieved a 20% return.

6. Sharpe Ratio: 1.2
   - The Sharpe ratio is a measure of risk-adjusted return. A ratio of 1.2 indicates that the model is generating a good return for the level of risk taken.

7. Sortino Ratio: 1.5
   - The Sortino ratio, which focuses on downside risk, is also favorable at 1.5. This suggests that the model is efficient in managing losses and generating returns

8. Total Closed Trades: 100
   - The model executed 100 trades during the test period.

9. Number of Winning Trades: 75
 - Out of the 100 trades, 75 were profitable. This gives a win rate (profitability percentage) of 75%.

10. Number of Losing Trades: 25
    - The remaining 25 trades resulted in losses.

11. Average Winning Trade (in USDT): $2,000
    - On average, each profitable trade yielded a profit of $2,000.

12. Average Losing Trade (in USDT): -$400
    - On average, each losing trade resulted in a loss of $400.

13. Largest Winning Trade (in USDT): $5,000
    - The best single trade generated a profit of $5,000.

14. Largest Losing Trade (in USDT): -$1,500
    - The largest loss in a single trade was $1,500.

15. Average Holding Duration per Trade: 2 days
    - On average, each trade was held for 2 days before being closed.

16. Max Dip and Average Dip
 - - The maximum dip percentage in a trade is a measure of the largest percentage decline from the entry price to the lowest price observed during the duration of a running trade. We have attached a calculation for this in the tech deliverables doc.

Given a starting capital of **$100,000**, the model was able to generate a net profit of **$47,000**, resulting in a **47% return on investment (ROI) during the test period**. It outperformed a simple buy and hold model, which would have yielded a **20% return**, and achieved a favorable risk-adjusted return with a **Sharpe ratio of 1.2 and a Sortino ratio of 1.5** .

PERIODIC CHECK-IN SYSTEM:-

To ensure the smooth progression of the project, we will allocate a dedicated team for periodic check-ins. This team will be available to address any queries or concerns faced by participating teams in their algorithmic trading model development.

ADDITIONAL REFERENCE POINTS:-

1. Triumph of Jim Simons

Jim Simons, a mathematician who achieved legendary status as a hedge fund manager. He harnessed mathematics and statistics to consistently beat the financial markets. His algorithms analyzed data, identified patterns, and consistently outperformed traditional strategies/models, making him the most successful manager on Wall Street.

Simons' story illustrates the transformative power of maths and statistics in trading, serving as inspiration for participants aiming to craft innovative and profitable models for the BTC/USDT market.

2. Trading Insights

TradingView, a widely used online platform, provides a wealth of resources for traders and investors. It offers real-time market data, advanced charting tools and participants are encouraged to explore TradingView for insights, chart analysis, and model development. It can be an invaluable reference to stay informed about market trends.

3. Interdisciplinary Insights

Zelta Tech recognizes the multidisciplinary nature of algorithmic trading. Teams are encouraged to explore insights from various domains, including mathematics, statistics, finance, and economics. The goal is to foster a holistic approach that combines quantitative methods with a deep understanding of market dynamics.

Thank You and Good Luck!

Thank you for joining Zelta Tech's Algorithmic Trading Model Development competition for BTC/USDT. Your expertise and innovation are vital to this challenge.

Best of luck in crafting models that redefine success. It's not just about the profits but the journey of discovery and learning.