REPORT
ZELTALABS PROBLEM STATEMENT

Introduction

Algorithmic trading, or algo-trading, involves using automated software to execute trades based on predefined criteria. In the context of the cryptocurrency market, specifically BTC/USDT (Bitcoin against US Dollar Tether), algorithmic trading provides an efficient way to capitalize on market fluctuations, high volatility, and liquidity.

This report details the development process of an algorithmic trading strategy for the BTC/USDT market. The focus is on the end-to-end process of strategy design, backtesting, and optimization of the strategy.

1.Market Overview: BTC/USDT

BTC/USDT represents one of the most liquid cryptocurrency pairs. The market is known for its:

- **High volatility:** Prices can change rapidly, making it an ideal candidate for short-term trading strategies.

- **24/7 operation:** Unlike traditional stock markets, crypto markets never close.

- **Liquidity:** With significant trading volumes, especially on large exchanges, BTC/USDT provides ample opportunity for both institutional and retail traders.

Given these characteristics, algorithmic strategies can be particularly effective in exploiting short-term market inefficiencies and trends.

2.Strategy Development Process

The development process for an algorithmic trading strategy can be broken down into several stages:

2.1 Timeframe Selection

Cryptocurrency markets operate continuously, so selecting the appropriate timeframe is crucial. In this case, we target the **1-day timeframe** for the BTC/USDT pair.

2.2 Strategy Ideation

In finance, technical analysis is an analysis methodology for forecasting the direction of prices through the study of past market data, primarily price and volume. Behavioral economics and quantitative analysis use many of the same tools of technical analysis, which, being an aspect of active management, stands in contradiction to much of modern portfolio theory. The efficacy of both technical and fundamental analysis is disputed by the efficient-market hypothesis, which states that stock market prices are essentially unpredictable, and research on whether technical analysis offers any benefit has produced mixed results.
It is critical to determine the type of strategy that best fits the BTC/USDT market conditions. A few common technical strategies include:

- **Trend Following:** Buying or selling based on the assumption that the current price direction will continue.
- **Momentum Following :** Give us some insight about the momentum of the closing price along with time.
- **Volume Following ：** A technical tool to measure and analyze the amount of trading activity or volume of assets being traded over the specific period.
- **Volatility Following:** It measures the magnitude of price fluctuations in a financial market. High volatility indicates

larger price swings, while low volatility suggests more stable, less dramatic price movements.

2.3 Selecting Indicators

First we computed 50+ indicators using the OLHC data and grouped them based on their type (Volatility,Momentum,Trend ,Volume) .Then we made four different heat maps on each category based on the mutual correlation coefficients of each pair of indicators.

We have chosen those indicators which are either oscillators or ratios or indices.

We did not consider those indicators that are highly correlated with target as they become autocorrelated with the prices which makes the model overfit.

We have taken indicators whose correlation coefficient is less than 0.8 between each type of indicators. The shortlisted indicators included are:

Volume

- Money Flow Index (MFI)
- Chaikin Money Flow (CMF)
- Force Index (FI)
- Ease of Movement (EoM, EMV)

Volatility

- Ulcer Index (UI)
- Coefficient of Variance

Trend

- Moving Average Convergence Divergence (MACD)
- Average Directional Movement Index (ADX)
- Trix (TRIX)
- Mass Index (MI)

- Detrended Price Oscillator (DPO)
- Schaff Trend Cycle (STC)

Momentum

- Relative Strength Index (RSI)
- True strength index (TSI)
- Ultimate Oscillator (UO)
- Awesome Oscillator (AO)
- Percentage Price Oscillator (PPO)
- Percentage Volume Oscillator (PVO)

Dealing with missing values(NaN): The missing values were set to 0 since XGBoost has the ability to handle missing values when inserted as zero.

2.4 XGBoost Model

XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm that is widely used for classification and regression tasks due to its performance, speed, and flexibility. It belongs to the family of **boosting algorithms**, where multiple weak learners (usually decision trees) are combined to create a strong predictive model. XGBoost is based on the principle of **gradient boosting**, but it offers additional enhancements like regularization, parallel processing, and handling missing data efficiently.

Key concepts in XGBoost :

- Boosting
- Gradient Boosting
- Regularization
- Tree Pruning
- Handling missing data
- Weighted quantile sketch
- Parellelization

- **Implementation of XGBoost**:

We have used the XGBClassifier model in our approach. We took all shortlisted indicators as features while the target was set to either 0 or 1 based on whether the daily return was negative or positive . The model was optimised using hyper parameter tuning.

- **Hyper-parameter Tuning** - We used the Bayes Search CV to find the ideal hyper parameter for our model , some of the tuned hyper parameters include max_depth of regression tree, learning rate ,L1 and L2 regularisation coefficients , no of regression trees trained .

- **Train-test Split :** Here we used the time series split since the chronological order of the data must be maintained to avoid look-ahead bias and resulting overfitting due to it.We also used a validation set primarily for model evaluation purpose .

- **Training the model:** The model was trained on the training data with eval_set set as validation data and evaluation metric set to 'logloss'.

- **Finding train and test accuracy**: We tested the model based on train and test accuracy , the model slightly overfitted on train set , but XGBoost has the ability to give good results despite overfitting.

3.Backtesting Strategy

3.1 Data collection

We have taken the historical data from 1st Jan 2022 to current of BTC/USD for backtesting our model. Data include the following :

- Price(Open , High , Low , Close)
- Volume
- Market Cap(Valuation)

3.2 Implementation

While backtesting we have computed the cumulative compounded returns of the strategy and compared it with cumulative returns of the stock in that timeframe.

3.3 Evaluating Criteria

When backtesting, several performance metrics are evaluated:

- **Sharpe-Ratio:**Measures risk-adjusted returns. A higher ratio indicates better risk-adjusted performance. It is the ratio of the excess expected monthly returns of the portfolio when compared with risk free rate of return divided by the standard deviation of the monthly portfolio returns , this value is annualised.

- **Maximum Drawdown:**The largest percentage drop in equity during the backtesting period, indicating the strategy's risk.

- **Win ratio:**A win ratio is a measure that shows the number of wins relative to total trades.

4.Enhanced Model for Risk Management

In the original model , the maximum drawdown is 32% which is repulsive for risk averse investor therefore we have devised an alternate approach for such investor assuming that log returns follow student t- distribution

Here instead of choosing a pure long short trade investor chooses a long short split which decreases returns but also reduces related risk.

Students t-distribution

We have proved that a random sample of size 38 follows t-distribution more closely than normal distribution.

The p_value is the ratio of the number of random samples of optimal sample size (found using ks-test) that follows t-distribution more closely than normal distribution

P_value =0.94 for our data.

- We use an XGBoost regressor model with tuned hyper-parameter, with indicators as features and the next day's returns as target.

- Such a model outputs a value instead of a trade call, based on this value we make a long-short split investment.

- We divide the predicted log-returns value with the probability distribution function to get the expected returns.

- Such an approach will help us to penalise the noises and help us to exploit the chances when the predicted log returns are very high showing that our model is highly confident about the price trend thus we can exploit those market conditions to book high profits while a lower expected return i.e the one having higher P.D.F indicates that the model is not confident and reacting to arbitrary noises.

5. Performance Evaluation

The backtesting results on historical data indicated the following:
- **Annualised Sharpe ratio:** 1.94
- **Maximum drawdown:** 32%
- **Cumulative Returns:** 320.8% in 649 days
- **Annual Return:** 74%
- **Win ratio:** 53%

6. Future Improvements

We have taken the future data of BTC/USDT from 1st Jan 2023 to the present date and compare it with our future results. Based on results and insights from backtesting , the following improvements can be considered:

- **Minimize drawdown:** By using t-distribution, we shrink the maximum drawdown to 22.9% with lower cumulative returns of

70% in 649 days which is less volatile for future results but also less rewarding.

- **Returns Improvement:** By using machine learning model, we improve the returns of the data .

Conclusion

Overall, the report offers a good level of completeness and provides solid insights into the BTC/USDT market. With a more robust risk management section and comparative analysis, future reports could provide even greater value to traders and investors.