Report

Bored From Ludo

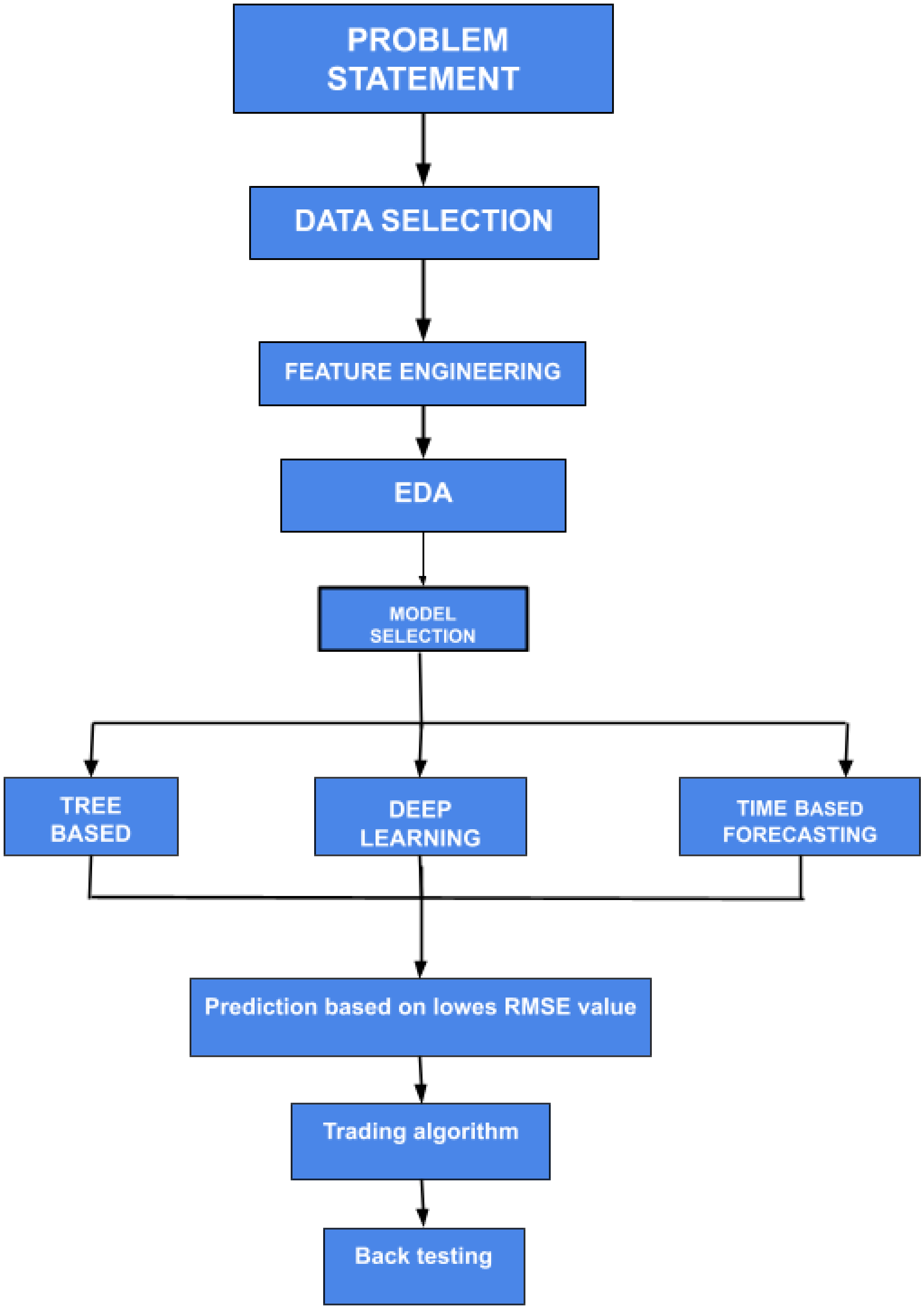
PROBLEM STATEMENT: ALGORITHMIC TRADING STRATEGY DEVELOPMENT FOR THE BTC/USDT CRYPTO MARKET

Table of contents

Introduction

The problem involves developing algorithmic trading strategies for the BTC/USDT cryptocurrency market, aiming to outperform benchmark(buy and hold) returns. Participants are required to create trading algorithms that can generate returns while managing risk effectively in the specific BTC/USDT market.

Approach



PROBLEM STATEMENT
DATA SELECTION
FEATURE ENGINEERING
EDA
MODEL SELECTION
TREE BASED
DEEP LEARNING
TIME BASED FORECASTING
Prediction based on lowes RMSE value
Trading algorithm
Back testing

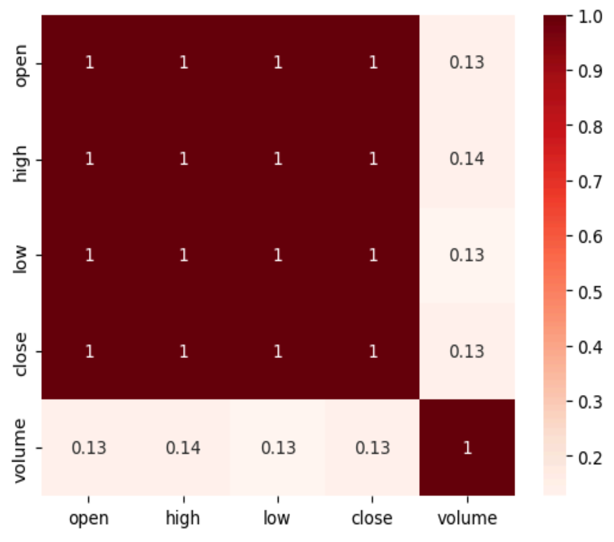
Data selection and preprocessing

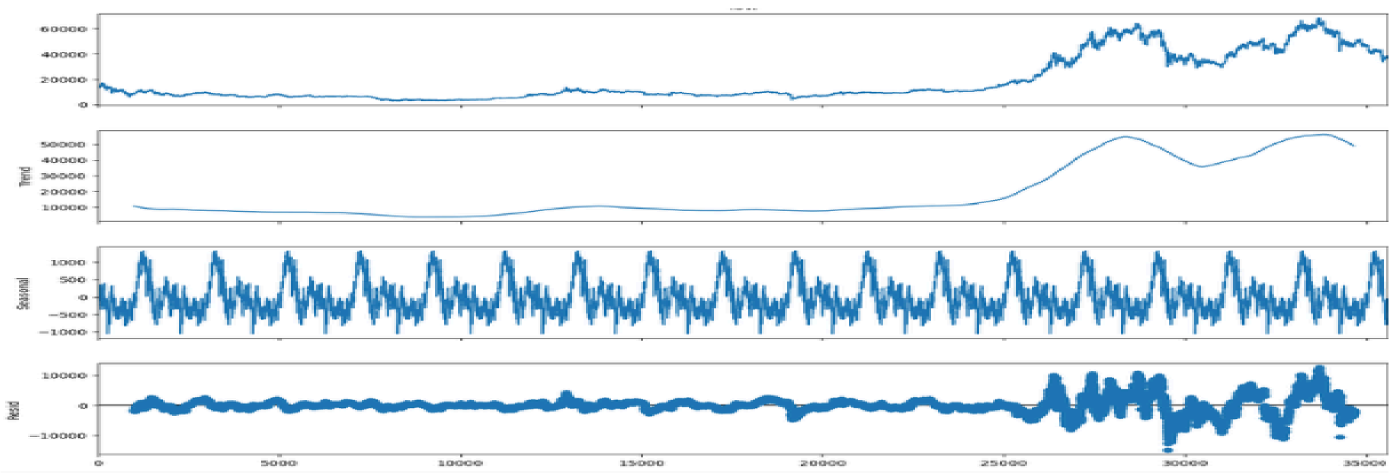
Data selection

In the realm of trading, the selection of an appropriate time frame is a delicate consideration. Through an exploration of temporal intervals, the 6-hour increments proved too broad, fleeting moments (3, 5 minutes) lacked depth, and higher frequencies (2, 4 hours) presented a scattered view of the market.The optimal compromise emerged at the 1-hour mark, striking a balance between granularity and data richness. This time frame allowed for a nuanced understanding of market dynamics, facilitating the formulation of effective trading strategies. For traders, the choice of timeframe is an ongoing quest to capture the essence of market movements without losing sight of the broader context—a meticulous calibration of temporal perspective to navigate the intricacies of financial landscapes.

EDA

We started with calculation of the correlation with the features of the data and then we moved on to plotting the heatmap between the values obtained. We observed very high correlation among Open, High, Low and Close data.





Seasonal Decomposition of the OHLC dataset revealed that the prices follow a seasonal pattern and thus an approach which utilised seasonality was considered.

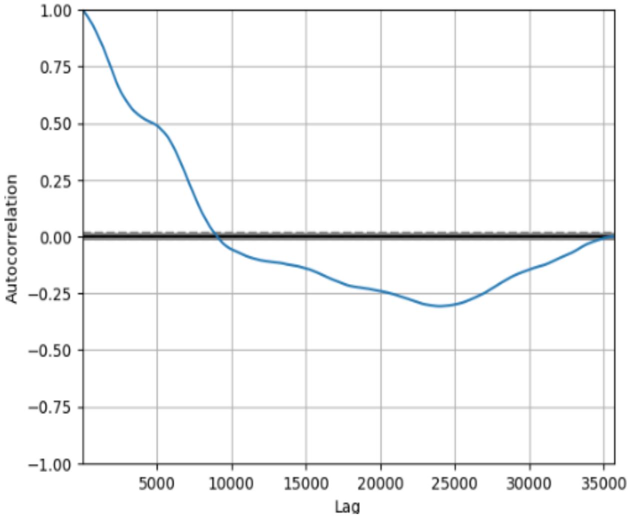
A seasonal decomposition plot slices time series data into trend, seasonality, and noise:-

1. Trend: This represents the long-term, underlying direction of the data, capturing sustained increases or decreases over time. Imagine it as the smooth, overarching curve guiding the data's overall trajectory.

2. Seasonality: This refers to recurring patterns within specific periods, like daily, weekly, monthly, or yearly cycles. Think of it as predictable fluctuations that repeat at regular intervals, influencing the data's behaviour within these cycles.

3. Noise: These are the remaining "wiggles" in the data, encompassing irregular variations not explained by the trend or seasonality. They can be due to random events, measurement errors, or other unpredictable factors.

Feature Engineering

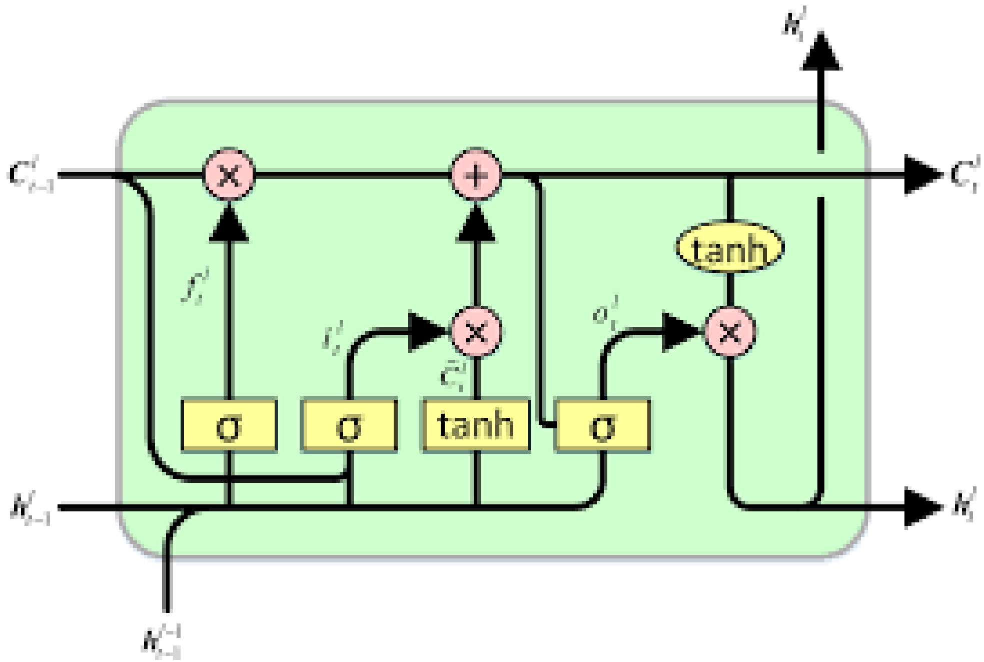


For training the Machine Learning Models we created lag features starting from the lag 1 till lag 24. Since we trained the models on 1 Hour timeframe data, we found that the last 1 day of values were highly correlated and thus we choose to use 24 period lag

ML model training

DL model: LSTM

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture. Unlike standard feedforward neural networks, LSTM has feedback connections.A standard LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.

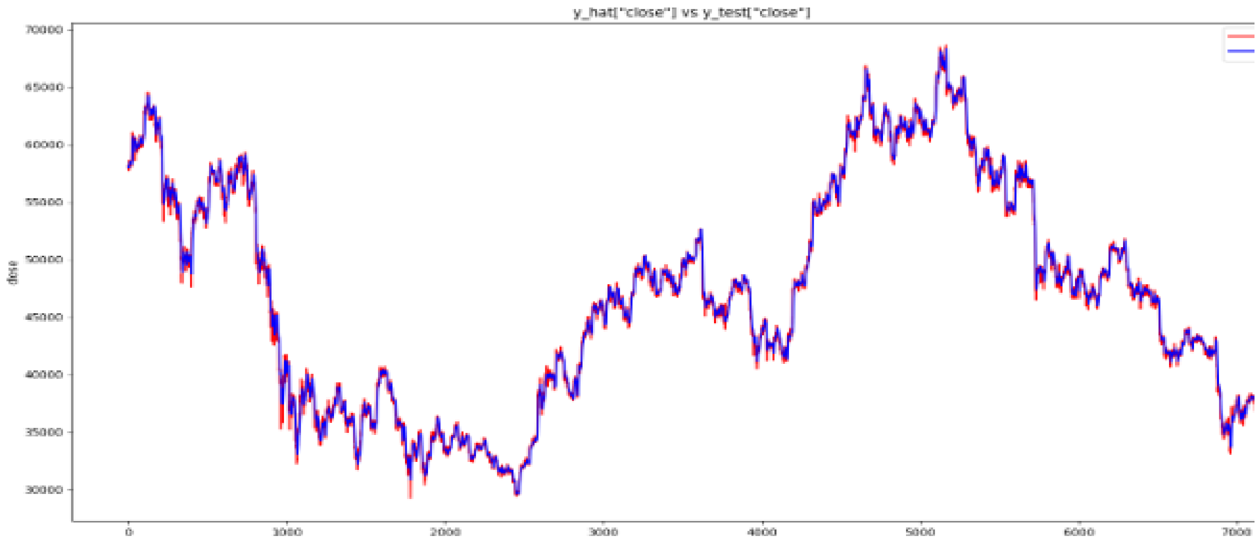


Why LSTM

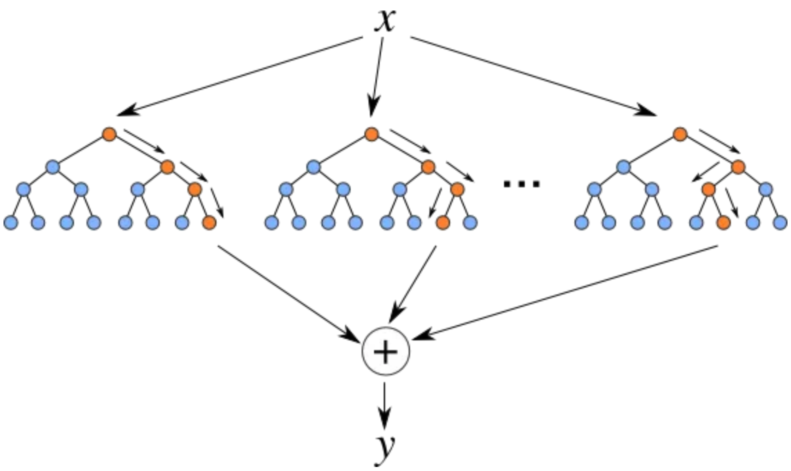
Having feedback connections in the neural network makes it viable for handling of time series data as each subsequent feature is informed by the previous one.Also, since there can be lags of unknown duration between important events in a time series, LSTM hold an advantage over traditional forecasting methods like Markov models or RNNs due to it relative insensitivity to gap lengths.

RMSE achieved : 618.8062

Tree-Based Model: Random Forest

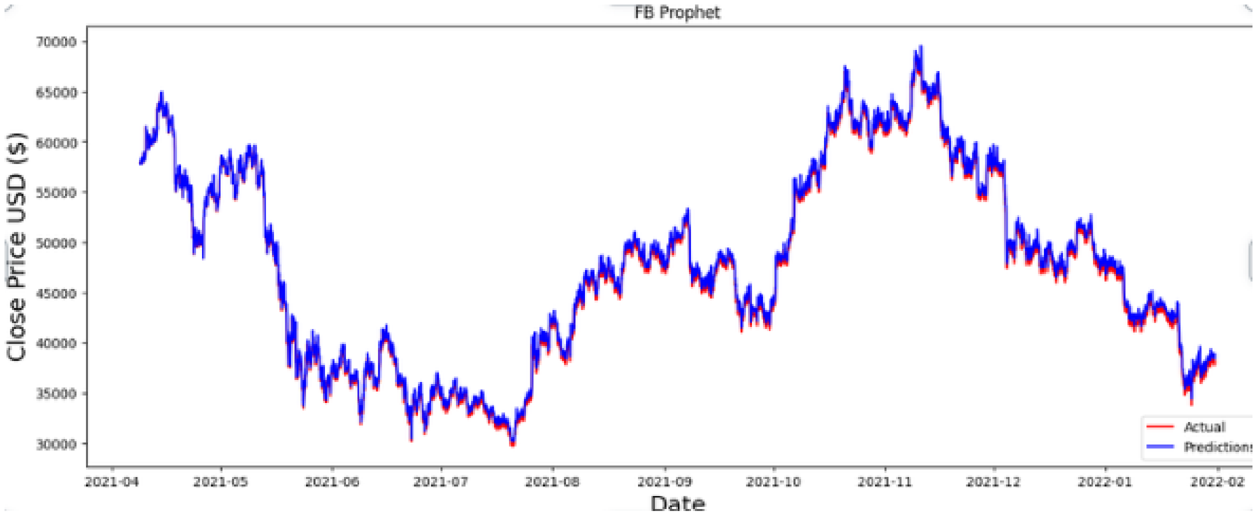


Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.



RMSE achieved : 1107.1095

Time Series Forecasting Model: FB Prophet



FBProphet is to model time series data as a combination of trend, seasonality, and noise components. By decomposing the data into these components, the algorithm can generate accurate forecasts that capture the underlying patterns in the data.

- The trend component captures the overall direction of the time series, whether it is increasing or decreasing over time. This component is modelled using a piecewise linear regression model, which allows for flexibility in fitting the trend to the data.
- The seasonality component captures the periodic patterns in the data, such as weekly or monthly trends. This component is modelled using the Fourier Series allowing for flexible modelling of different seasonal patterns.

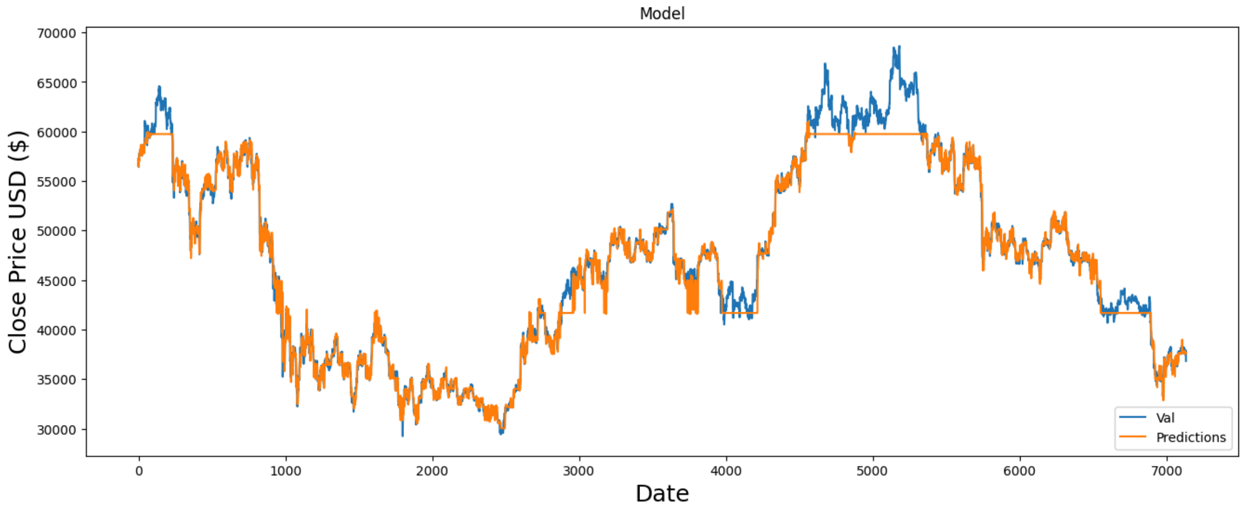
RMSE achieved : 599.8186

Other Models we tried

XG Boost:

It is an ensemble learning method that combines the predictions of multiple weak models to produce a stronger prediction.
One of the key features of XGBoost is its efficient handling of missing values, which allows it to handle real-world data with missing values without requiring significant pre-processing. Additionally, XGBoost has built-in support for parallel processing, making it possible to train models on large datasets in a reasonable amount of time.

RMSE achieved - 1450.66



Why it failed :

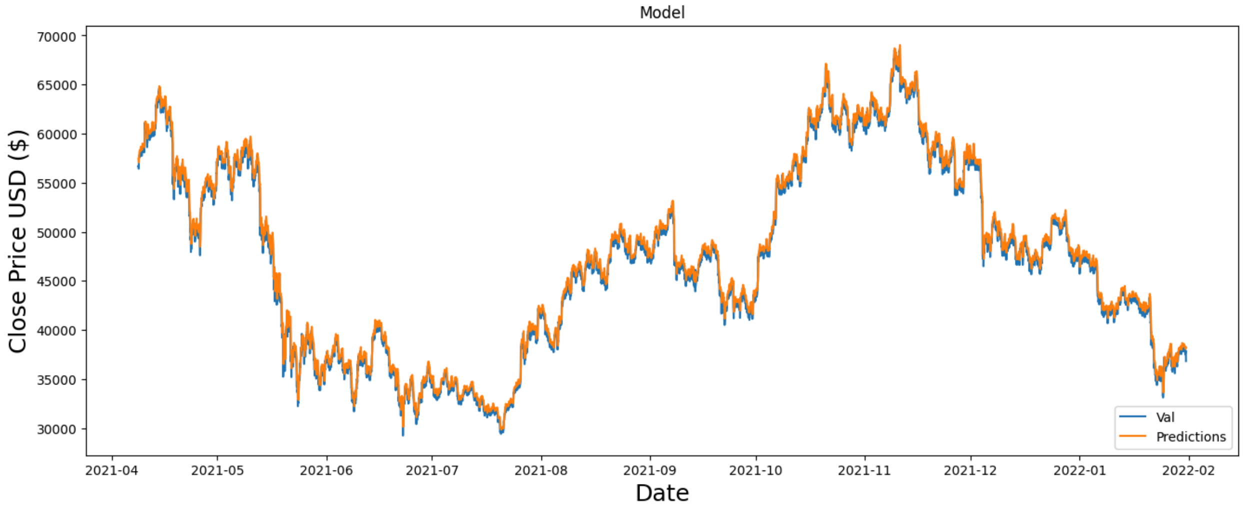
- Hyperparameter Tuning Challenge: XGBoost requires expertise and patience for effective hyperparameter tuning.
- overfitting Trap: Without careful tuning, XGBoost may overly adhere to specific training data. This can limit its applicability to new or unseen datasets.

GRU:

GRUs, or Gated Recurrent Units, are memory masters in the ML world. Imagine navigating a maze blindfolded - GRUs are your guide, remembering twists and turns, and choosing the best path forward. Like LSTMs, their cousins, they excel at tasks with sequences, like speech recognition or predicting stock prices.

Here's how they work: think of two gates - a reset gate decides what past info to forget, and an update gate selects the new info to keep. This keeps their memory lean and focused, making them faster and easier to train than some other models.

RMSE achieved - 618.89



Why it failed :

- Short memory: GRUs struggle with long-term dependencies due to their combined hidden state, potentially reducing accuracy in some tasks.
- Lower accuracy potential: Compared to LSTMs, GRUs may underperform on complex tasks requiring advanced memory management.

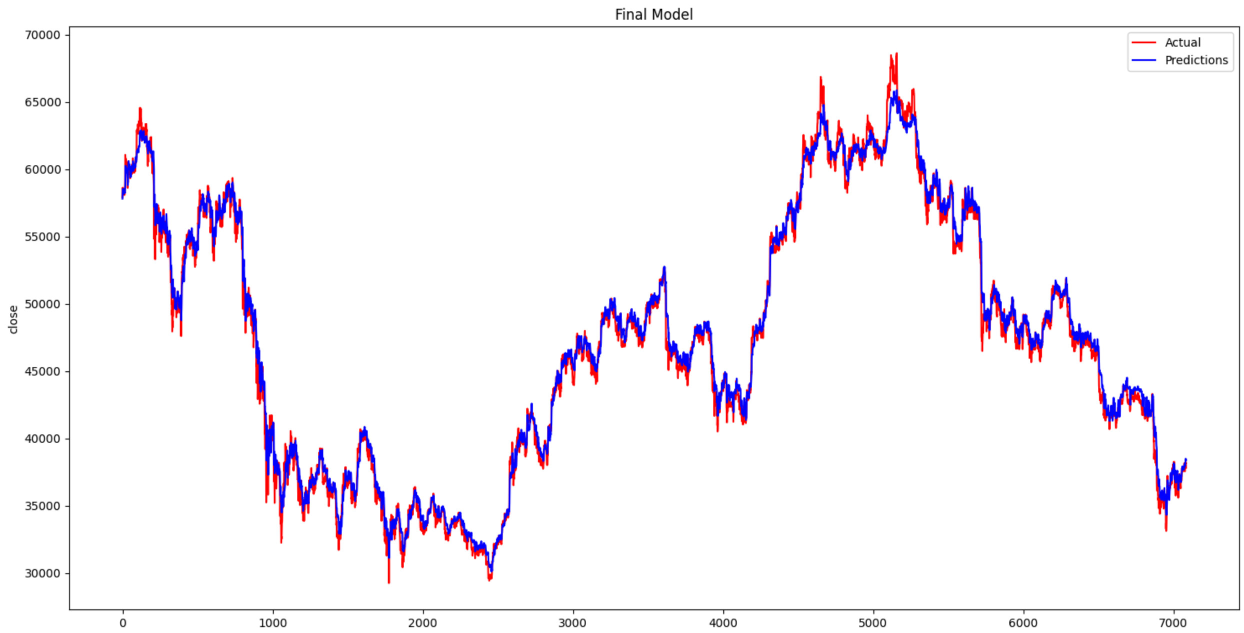
Model selection

We got the best results from LSTM, FB Prophet, and Random Forest. Each of these models excels in different areas: LSTM captured temporal dependencies, Prophet identified long-term trends, and Random Forest unravelled hidden data relationships.

In order to create a robust model which can perform with greater accuracy on the data, we combined their forecasts via averaging, achieving a significant RMSE reduction to . This success stems from:

- Diversity: Each model's unique approach mitigates individual biases, creating a more robust picture.
- Ensemble averaging: It smooths individual predictions, yielding a more reliable overall forecast.

Final RMSE: 807.628



Trading Strategy

This strategy combines two popular indicators (EMA crossover and RSI/MACD) for a more robust entry and exit decision-making process.

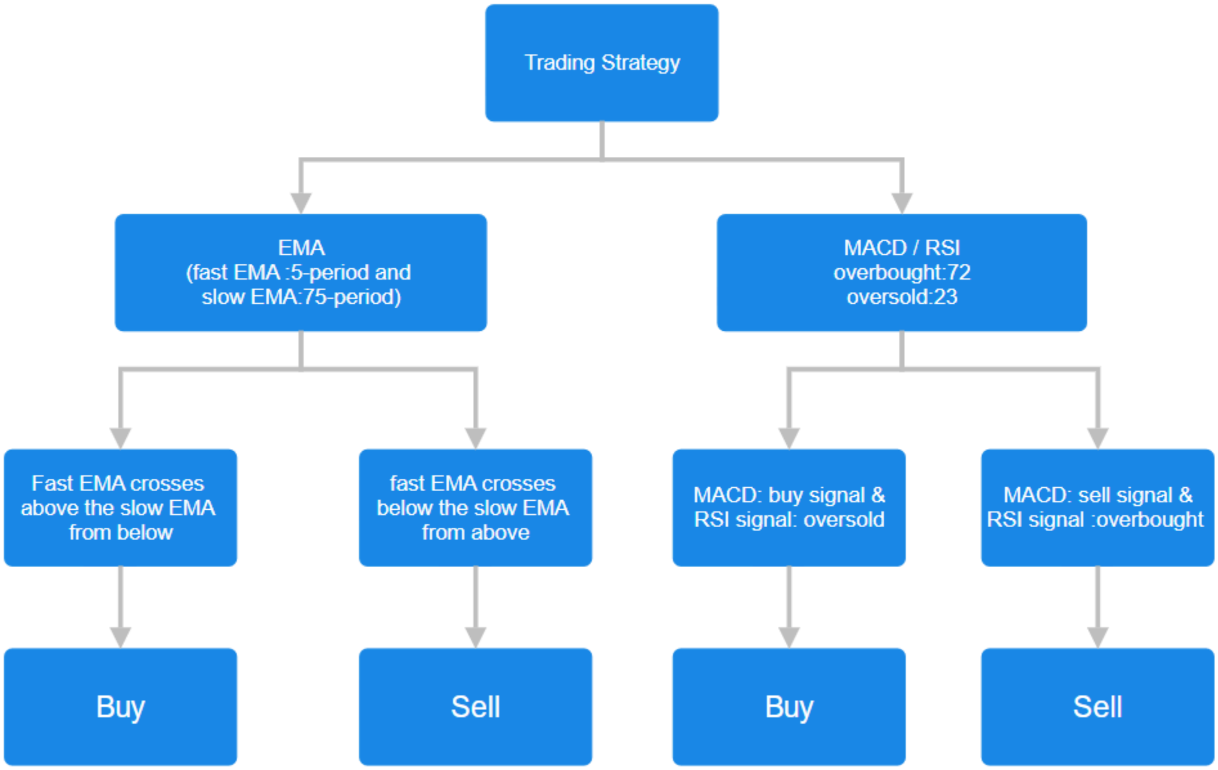
1. EMA Crossover:

- We utilise two Exponential Moving Averages (EMAs) with different timeframes.
- A fast EMA (5-period) reacts quickly to price changes, indicating short-term momentum.
- A slow EMA (75-period) acts as a trend indicator, offering bigger picture direction.
- Buy Signal: When the fast EMA crosses above the slow EMA from below, it suggests a potential trend reversal to the upside.
- Sell Signal: Conversely, when the fast EMA crosses below the slow EMA, it indicates a possible downtrend.

2. MACD and RSI Confirmation:

- The Moving Average Convergence Divergence (MACD) assesses trend strength and potential reversals.
- The Moving Average Convergence Divergence (MACD) assesses trend strength and potential reversals
- Buy Confirmation:the MACD line crosses above the signal line, suggesting stronger bullish momentum.
- Sell Confirmation: MACD line crossing below the signal line adds confirmation to the bearish trend.
- The Relative Strength Index (RSI) measures price momentum and identifies overbought/oversold conditions.
- We define overbought and oversold thresholds (72 and 23, respectively).
- A neutral zone around 50 signifies balanced momentum.

- Buy Confirmation: If the fast MACD generates a buy signal, we only confirm it if the RSI is in the oversold zone. A value below 23 gives confidence in the upward trend.
- Sell Confirmation: Similarly, for a sell signal from the MACD , we require the RSI to be above 72 ;ie. Overbought to confirm the downtrend.



Risk Management

Stop loss

A stop-loss order is a crucial tool for risk management in trading. It automatically closes your position when the price moves against you to a predetermined level, limiting your potential losses.

Long Positions:

- Calculate the Risk Percentage: Decide on your acceptable risk percentage per trade. For example, if you're willing to risk 2% on a $100 stock purchase, your risk amount would be $2.
- Convert to Price Difference: Divide the risk amount by the number of shares purchased. In this case, $2 / 100 shares = $0.02 per share.
- Subtract from Entry Price: Subtract the price difference from your entry price to find the stop-loss trigger point. With a $50 entry price, the stop-loss would be set at $50 - $0.02 = $49.98.

For Short Positions:

- Calculate the Risk Percentage: Same as for long positions.
- Convert to Price Difference: Again, divide the risk amount by the number of shares sold short.
- Add to Entry Price: Add the price difference to your entry price (which is negative for short positions). For example, a $50 short position with a 2% risk would trigger a buy-to-close order at $50 + $0.02 = $50.02.

Pros:

- Captures profits during trends
- protects against pullbacks.

Cons:

- Requires accurate identification of support/resistance
- susceptible to false breakouts.

By utilising stop-loss strategies effectively, you can protect your capital and navigate the market with greater confidence

Results

Backtesting: with optimization

Gross Profit	22866949.74
Net Profit	22766949.74
Total Closed Trades	906
Win Rate (Profitability %)	27.04
Max Drawdown	-35.51
Risk Reward Ratio	6.67
Average Winning Trade	0.60
Profit(%)	22766.949
Buy and Hold Return of BTC	172.22
Largest Losing Trade(%)	-3.16
Largest Winning Trade(%)	20.09
Sharpe Ratio	1.09
Sortino Ratio	8.36

Backtesting: without optimization

Gross Profit	471178.839
Net Profit	371178.83
Total Closed Trades	1688
Return(%)	371.17
Max Drawdown(%)	-49.66
Profit Factor	1.162
Buy and Hold Return of BTC	172.2246
Largest Losing Trade(%)	-5.12
Largest Winning Trade(%)	28.06
Sharpe Ratio	0.40819
Sortino Ratio	1.066785
Average Holding Duration per Trade	8 Days 11:00:00