

Bitcoin Trend Classifier

Machine Learning-Based Crypto Market Predictor

Team 4

Project by :

- Haywar Zhu
- Vishnu Vijay
- Nariman Hosseini
- Ramees Thanikkal



Introduction: Project Core Message

The core message of this project is to investigate whether machine learning models can effectively predict 'Trendy' or 'Sideway' states in the cryptocurrency market, specifically the Bitcoin market.

We aim to provide a data-driven approach to support trading decisions, focusing on various technical indicators like ADX, RSI, EMA and Bollinger Bands.



Why do we need our model?

Traders know different market conditions often favor different trading strategies.

For Trendy Market:

- Momentum Trading: Traders aim to catch stocks or cryptos in the act of accelerating, buying high and selling higher.
- Breakout Trading: In a trendy market, levels of support and resistance are often tested. A breakout strategy involves entering a position as the asset breaks through these levels, anticipating that the trend will continue.

For Sideway Market:

- Mean Reversion: In a sideways market, prices tend to revert to the mean or average level. Traders aim to buy low and sell high within the range.



Data Preparation

- Data Source: Historical daily price data for Bitcoin from 2010 to 2023.
- Technical Indicators Used:
 - Average Directional Index (ADX)
 - Relative Strength Index (RSI)
 - Bollinger Bands
 - Exponential Moving Average (EMA)
- Feature Scaling: Used Standard Scaler to normalize all the feature values.
- Data Splitting: Data is divided into training, testing, and validation sets.
- Model Training: SVM, GB, RF, LG, LSTM, NN



Automated Labeling

Context: In machine learning, having a target label (y label) is crucial for supervised learning models.

Challenge: Manually labeling years of Bitcoin market data for trends is impractical and extremely time-consuming.

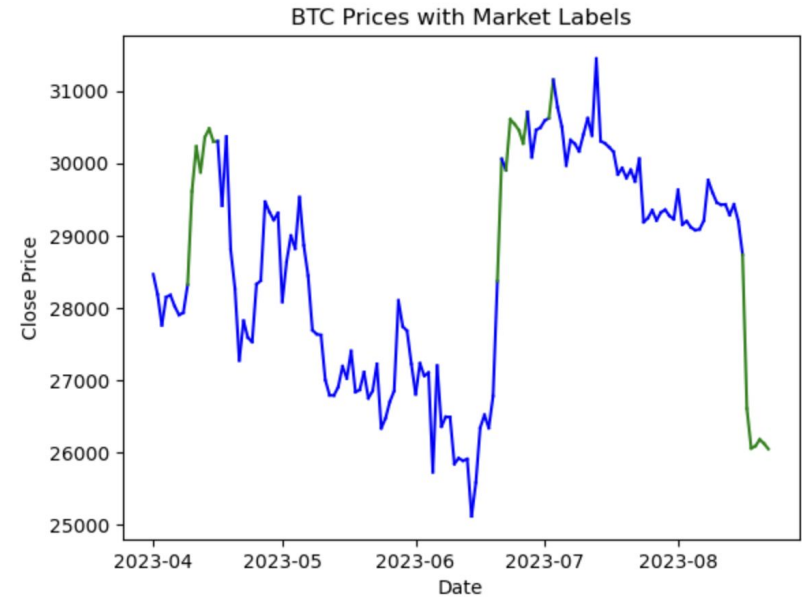
Solution: Automated Labeling using Technical Indicators. They are widely used in trading and market analysis to understand market trends and momentum.

label_market() Function:

- Takes the dataframe with the technical indicators.
- Adds a 'MarketLabel' column, labeling each row as 'Trendy' (1) or 'Sideway' (0).
- Two out of Three Rule: A market is considered 'Trendy' if at least two of the three selected indicators suggest a trend.

Backtesting

- Backtesting is essential for validating the model's predictions. We've set aside a validation dataset specifically for backtesting.
- Different colors denote the market conditions. This plot allows for human inspection.
- Human inspection adds another layer of validation. It's crucial, especially for volatile markets like cryptocurrencies, to cross-verify machine predictions.



SVM backtesting plot



Machine Learnings Models

Value Counts:

- Sideways Label - 2922
- Trending Label - 1837
- Total Labels - 4759



SVM Model

Test Set Report

Precision :

Sideways - 80 %

Trending - 91 %

Recall :

Sideways - 96 %

Trending - 62 %

Accuracy - 83 %

Validation Set Report

Precision :

Sideways - 79 %

Trending - 95 %

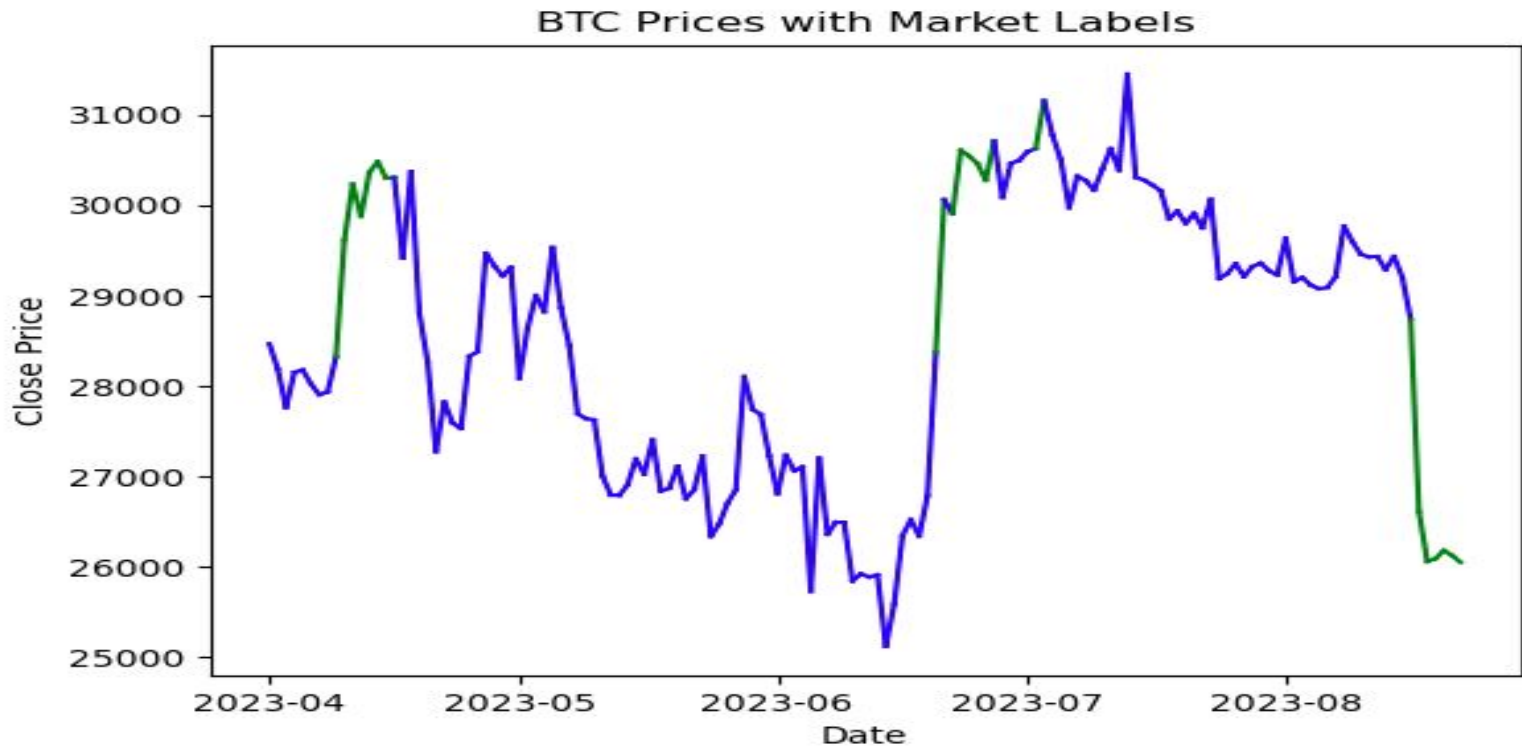
Recall :

Sideways - 99 %

Trending - 42 %

Accuracy - 81 %

Backtesting plot for SVM model.





Gradient Boosting Model

Test Set Report

Precision :

Sideways - 86 %

Trending - 94 %

Recall :

Sideways - 97 %

Trending - 75 %

Accuracy - 88 %

Validation Set Report

Precision :

Sideways - 80 %

Trending - 81 %

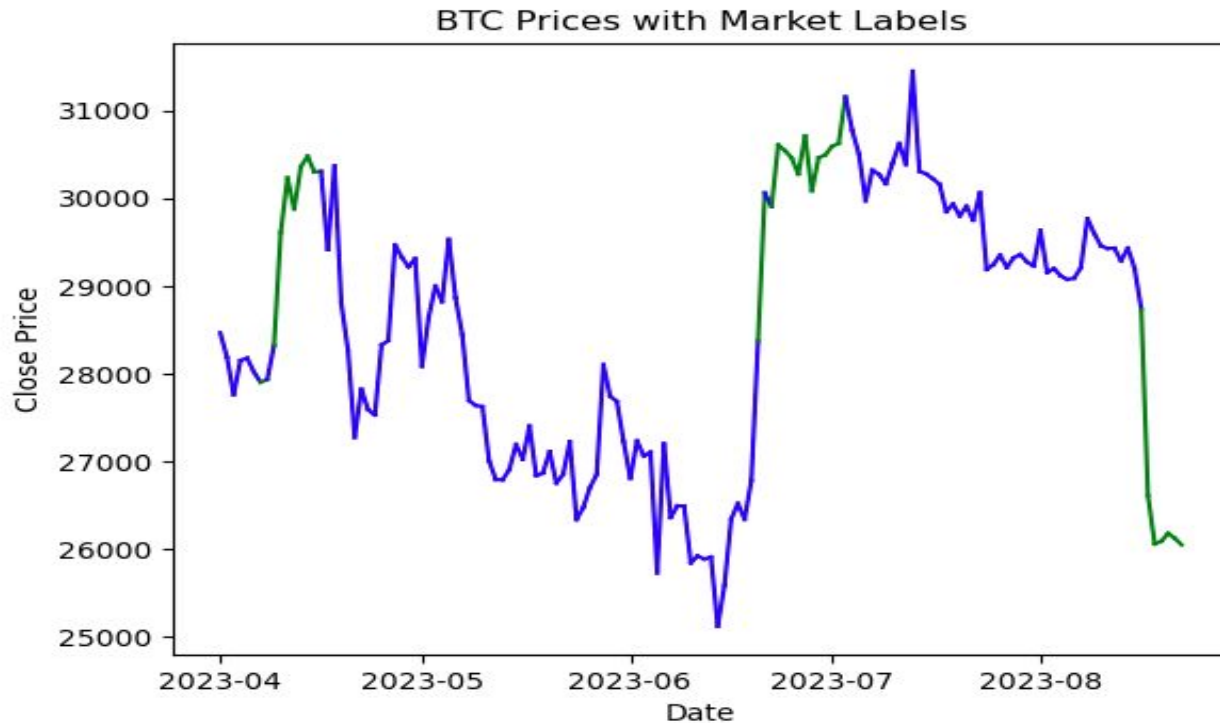
Recall :

Sideways - 95 %

Trending - 49 %

Accuracy - 81 %

Backtesting plot for Gradient Boosting Model





Random Forest Model

Test Set Report

Precision :

Sideways - 81 %

Trending - 93 %

Recall :

Sideways - 97 %

Trending - 62 %

Accuracy - 91 %

Validation Set Report

Precision :

Sideways - 77 %

Trending - 94 %

Recall :

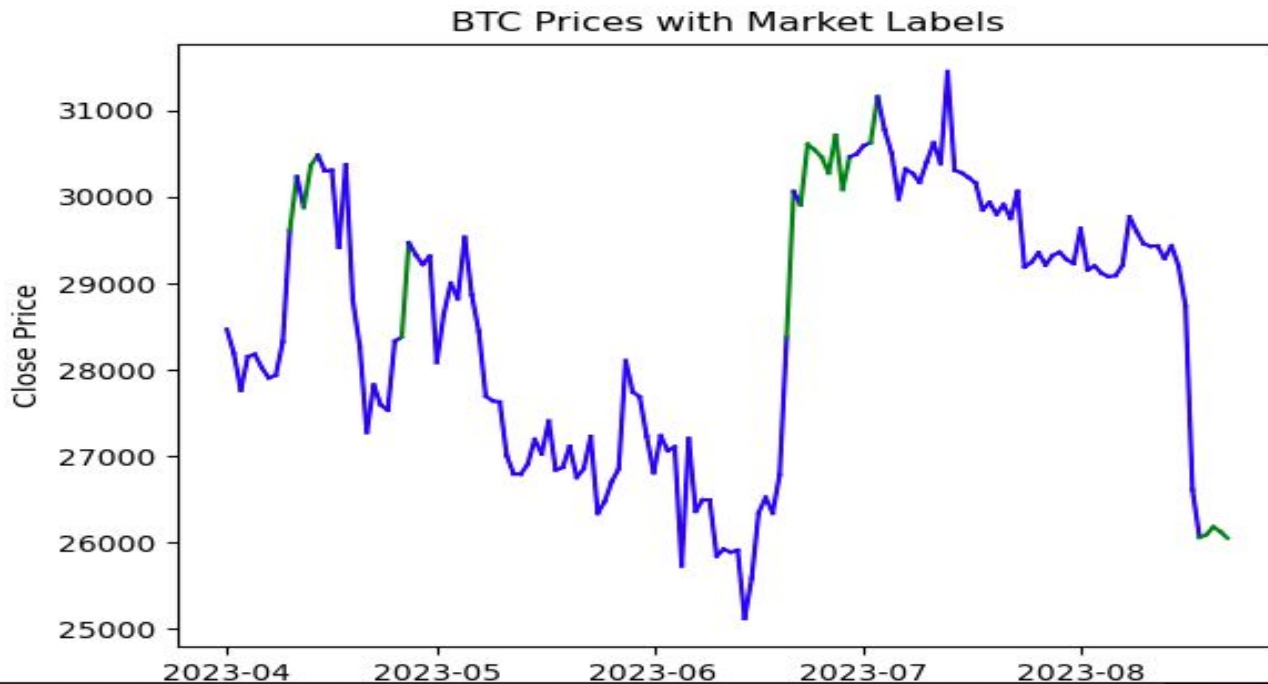
Sideways - 99 %

Trending - 36 %

Accuracy - 79 %



Backtesting plot for Random Forest Model





Neural Network

Test Set Report

Precision :

Sideways - 85 %

Trending - 91 %

Recall :

Sideways - 96 %

Trending - 72%

Accuracy - 87 %

Validation Set Report

Precision :

Sideways - 92 %

Trending - 69 %

Recall :

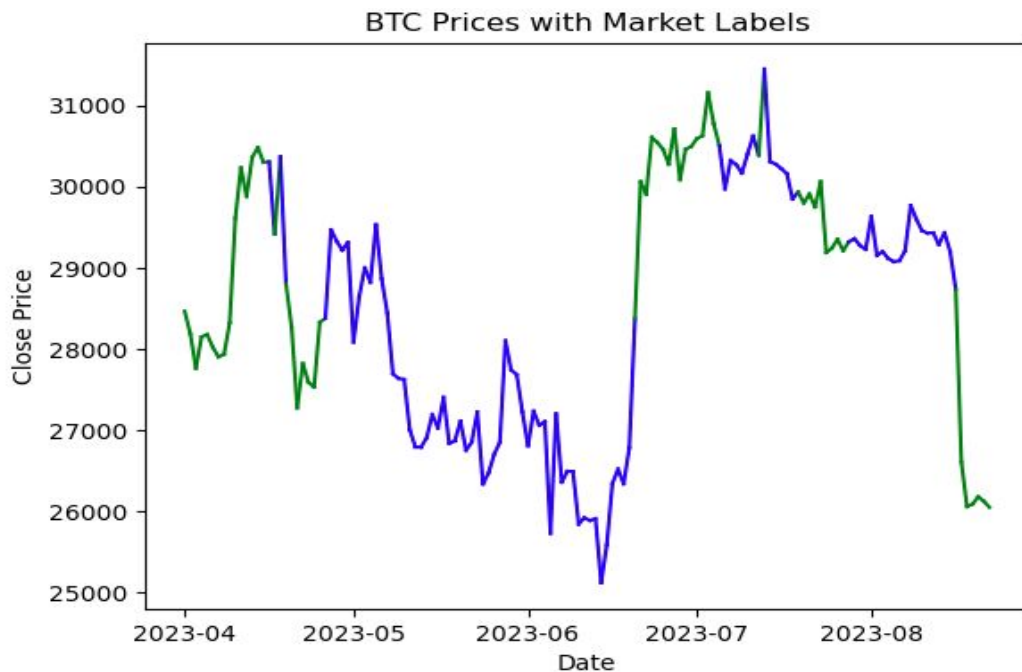
Sideways - 83 %

Trending - 84 %

Accuracy - 83 %

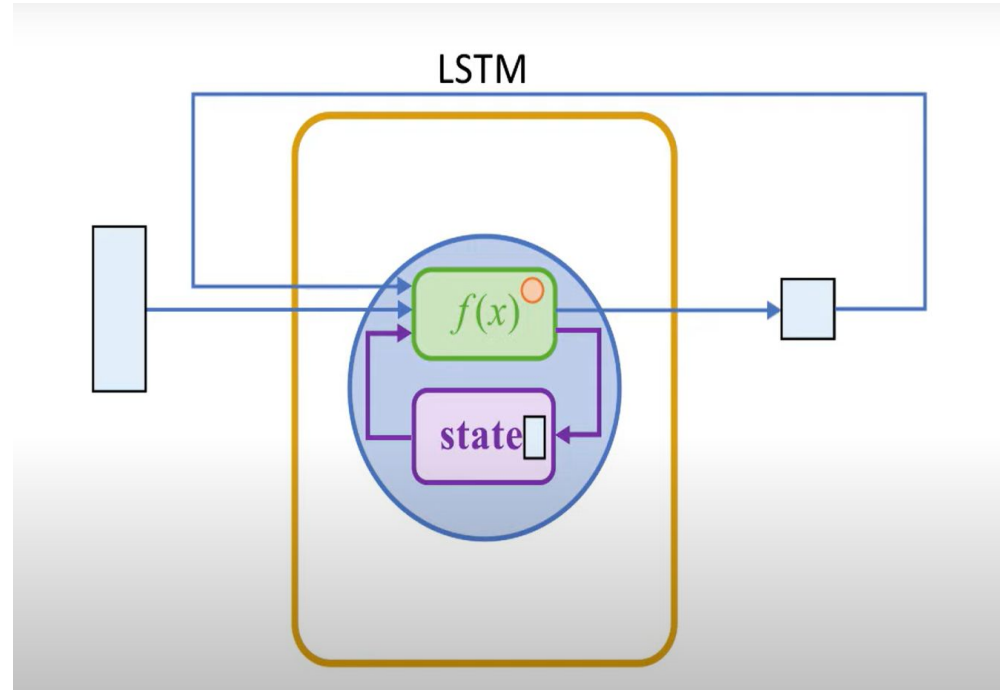


Backtesting plot for Neural Network Model



LSTM (Long Short Term Memory)

- A type of recurrent neural network.
- Recurrent Neural Network : Neural networks with the output fed back in as input.
- Useful to process sequential data for predictions.
- LSTMs have a working memory space called 'state' to avoid the effect of vanishing gradient in long term data.
- There are gates used in the node to control the flow of data interaction with the state.





LSTM

Test Set Report

Precision :

Sideways - 80 %

Trending - 90 %

Recall :

Sideways - 96 %

Trending - 62 %

Accuracy - 83 %

Validation Set Report

Precision :

Sideways - 84 %

Trending - 64 %

Recall :

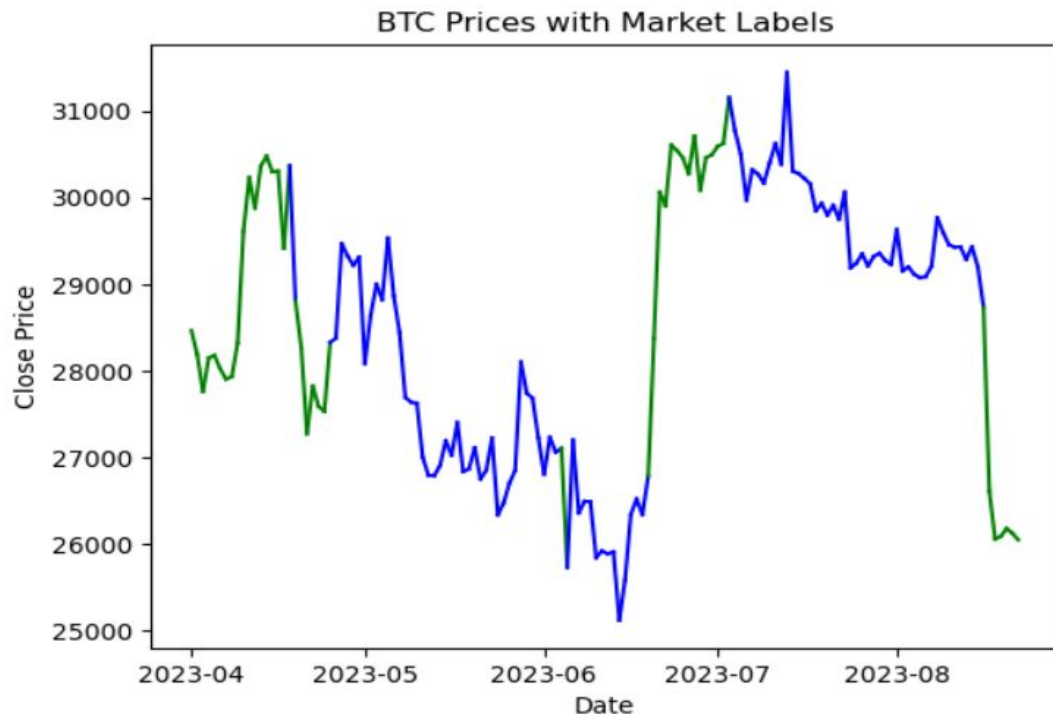
Sideways - 84 %

Trending - 64 %

Accuracy - 78 %



Backtesting plot for LSTM Neural Network Model





Challenges faced

- Labelling
- Backtesting
- Optimizing the model
- AWS sagemaker



Conclusion

- The models that performed the best are Random Forest with the accuracy of 91% and Gradient Boosting with 88% accuracy.
- These models would fit good for the chatbot that advise if the bitcoin is trending.



THANK YOU