Developing a Crypto Trading Strategy Leveraging On-Chain Data

Togay Atmaca – 12268957

Jordan Cassella – 12452926

Matthew Haimes – 12452937

Kunj Shah – 12258505

Yan Song Zhao – 12448802

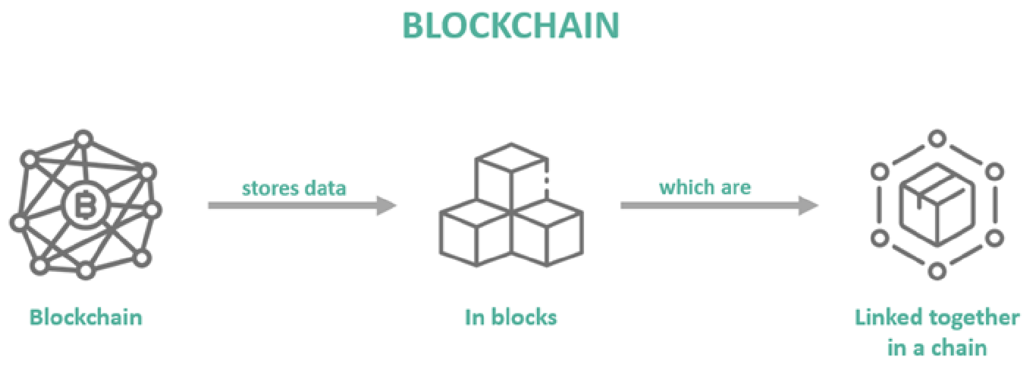
Agenda

Overview	Strategy Overview	MLP Model	Backtesting
• Blockchain Discussion • Proof of Work (PoW) vs. Proof of Stake (PoS) • Correlation of Token Returns	• Motivation • Investment Universe • Feature Space • Strategy Discussion	• Parameterization • Segmentation • Model Construction • Model Pipeline	• Assumptions • Parameter Tuning • Risk Management & Portfolio Allocation

Blockchain

- Ledger records transactions securely across multiple computers.
- Distributed storage keeps copies across a network of nodes.
- Blocks bundle transactions and update the ledger.
- Blockchain ensures decentralization, security, and immutability.

• **Proof of Stake (PoS)** is part of the **consensus mechanism**, which determines how new blocks are added to the blockchain



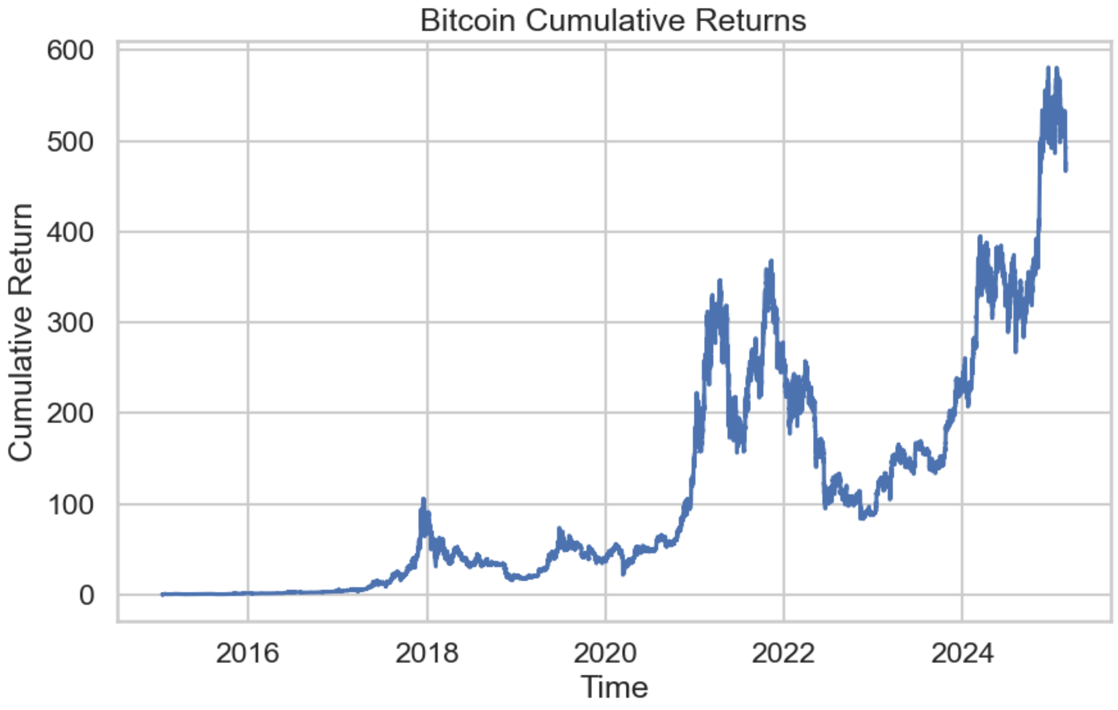
Proof of Work (PoW)

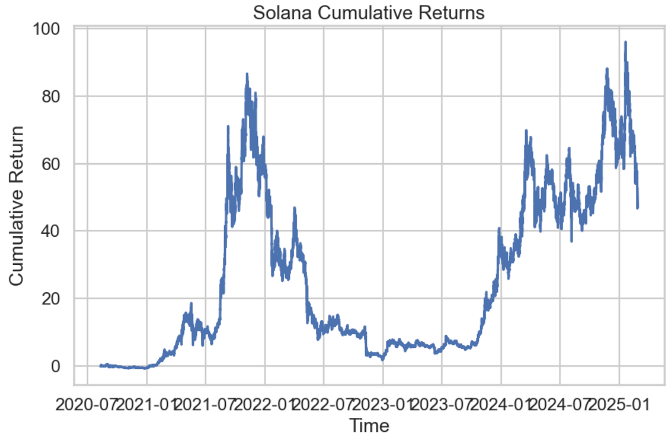
- PoW is a consensus mechanism in the blockchain network to validate transactions, by miners solving complex problems
  - This takes significant computation power. The winner adds a new block on to blockchain. This is very energy intensive

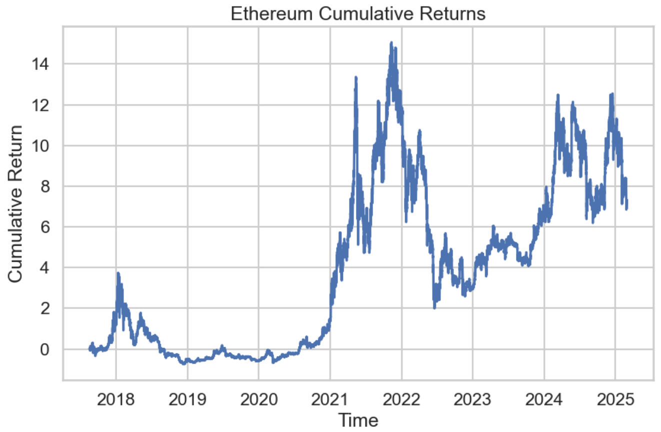
Proof of Stake (PoS)

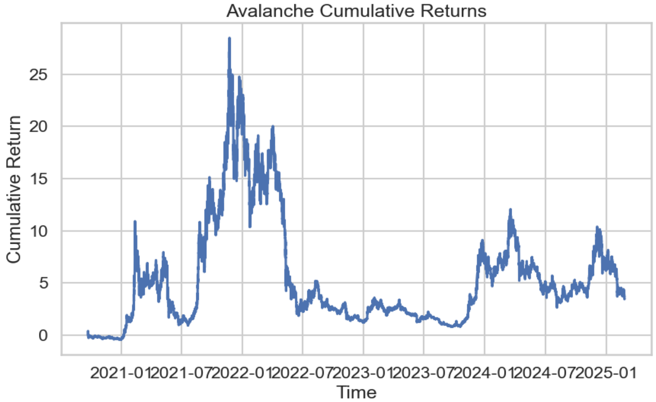
- PoS accomplishes the same without requiring high computational power.
  - Instead of miners, the PoS uses validators to lock up a certain amount of cryptocurrency as a stake
  - The validators create new blocks based on the amount staked and other factors like staking duration
  - The more stake the validator has increases the chance of being selected
  - More energy efficient

Proof of Work (PoW) vs Proof of Stake (PoS) Tokens



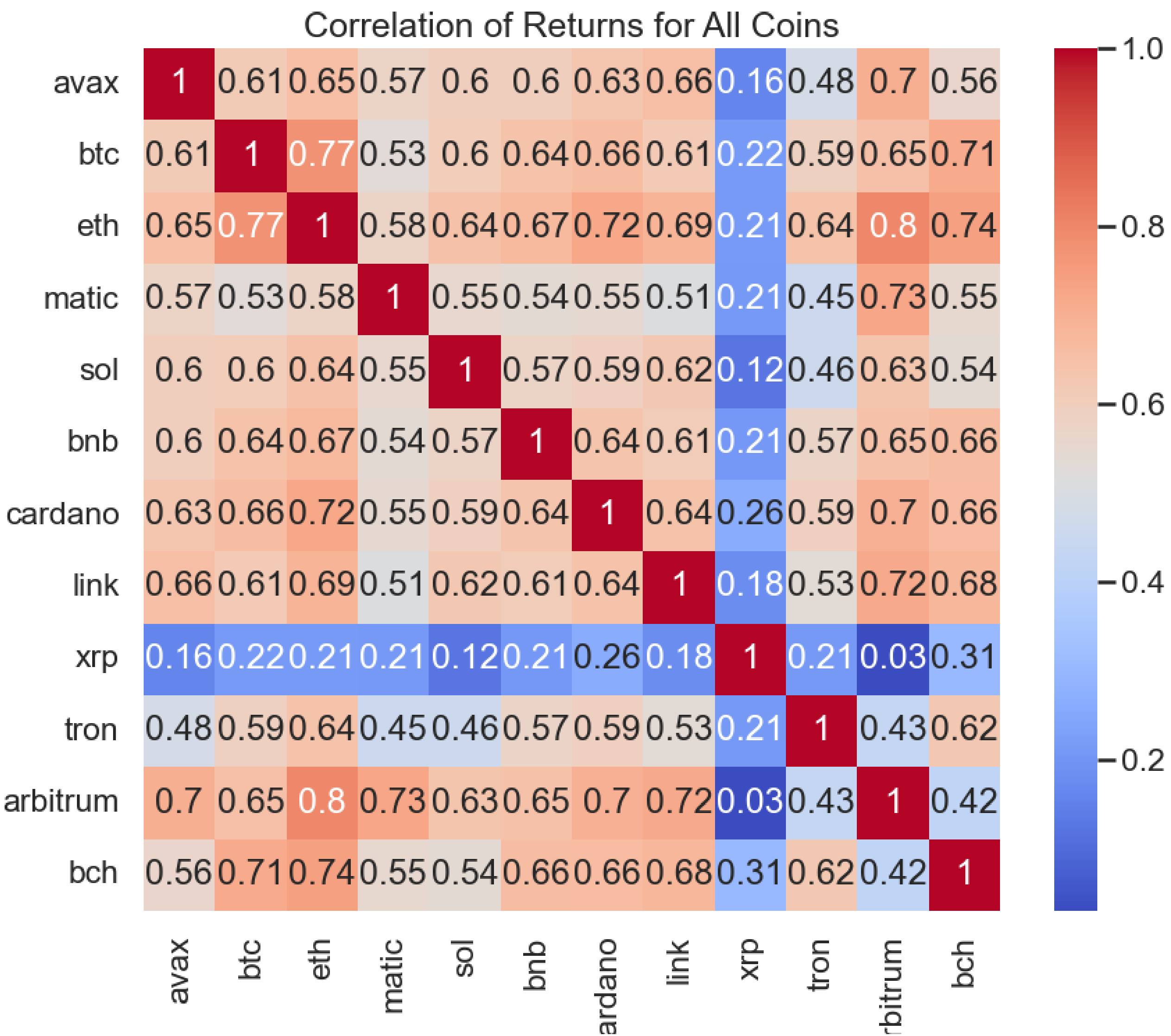






Bitcoin is one of the last remaining PoW tokens in the market, with many others (e.g. Ethereum) switching to PoS in the past 5 years. Thus, Bitcoin will be our main benchmark for PoW token performance.

Correlation of Token Returns



We observe instances of uncorrelated returns between our tokens of interest. We interpret this as a positive signal for diversification in our strategy's ultimate portfolio allocation.

Motivation

- Each asset class relies on fundamental data to determine proper valuation
- Traditional prediction models for cryptocurrency are limited by focusing on either:
  - Technical indicators from price history (Parente et al., 2024)
  - On-chain fundamental data (Kim et al., 2022)
- Our model bridges this gap by combining both data types and various methodologies from both papers
- This integrated approach delivers more accurate buy/hold/sell signals for enhanced risk management and portfolio allocation

Fundamental Data Types			
Equities	**Fixed Income**	**Commodities**	**Crypto**
- Financial Statements	- Interest Rates	- Supply and Demand	- On Chain Data

Hypothesis and Economic Rationale

- On-chain data metrics are fundamentally related to the price of PoW coins due to the nature of the block processing system
- Processing blocks in PoW systems is fundamentally related to energy consumption, making these metrics more useful in explaining the price of the digital asset
- Processing blocks in PoS systems is fundamentally related to how much validators are staking
- We hypothesize that because PoW on-chain metrics have a fundamental, contemporaneous relationship to asset price, there would not be as much predictive power looking into the future
- Therefore, we believe PoS on-chain metrics have a lead-lag relationship to digital asset price, where the information contained in the on-chain metrics are not instantaneously reflected in digital asset price, but instead, it's leading the asset price
- As a result, this lead-lag relationship can leave us a window to predict price directions in a profitable quantitative trading strategy

Investment Universe

	Investment Securities	Data
PoS	- Tron (TRON) - Arbitrum (ARB) - Avalanche (AVAX) - Polygon (MATIC) - Solana (SOL)	- Technical Time Series - On Chain Block Data - Sourced from Dune API
PoW	- Bitcoin (BTC) - Ethereum (ETH) - Binance (BNB)	

Feature Space for On-Chain Data

Feature Name	Feature Description
Block Difficulty	Unitless, relative measure of computational power required to mine a new block on the blockchain.
Block Mint Reward	The reward (typically in the respective cryptocurrency) for successfully mining a new block.
Block Stripped Size	The size of a given block on the blockchain (in bytes) stripped of signature bytes (included for validation and do not include relevant metadata).
Total (Gas) Fees	Fees collected by miners to host transactions on a given block.
Block Transaction Count	The number of transactions which have occurred on a given block.
Block Weight	A measure of a given block's size and complexity, weighing in favor of blocks with more transactions on them.

Data Sources

- Hourly (aggregated) On-Chain data was pulled from Dune
- Hourly Price data was pulled from CoinAPI

Strategy Overview

- Our strategy aims to leverage untransformed crypto data alongside constructed technical indicators (Fibonacci retracement, VWAP, RSI, etc.) to forecast strategy moves (buy, hold, sell).
- These indicators will form the data inputs for a Neural Network (Multi-layer Perceptron; MLP) model, aiming to identify profitable patterns and weed out irrelevant features.

Parameterization

The model itself features the following parameters:

- b := lookback window size
- f := prediction horizon
- x := lower prediction threshold
- y := upper prediction threshold

Segmentation

- Since crypto price data has significant time-varying volatility and dynamics that can negatively affect model training, we use a change point detection algorithm called PELT to mitigate these issues
- PELT allows us to identify "change" points in the time-series data, which creates several segments with their own unique statistical dynamics
- Using segmentation, we utilize a more robust methodology for normalization of non-stationary time-series data

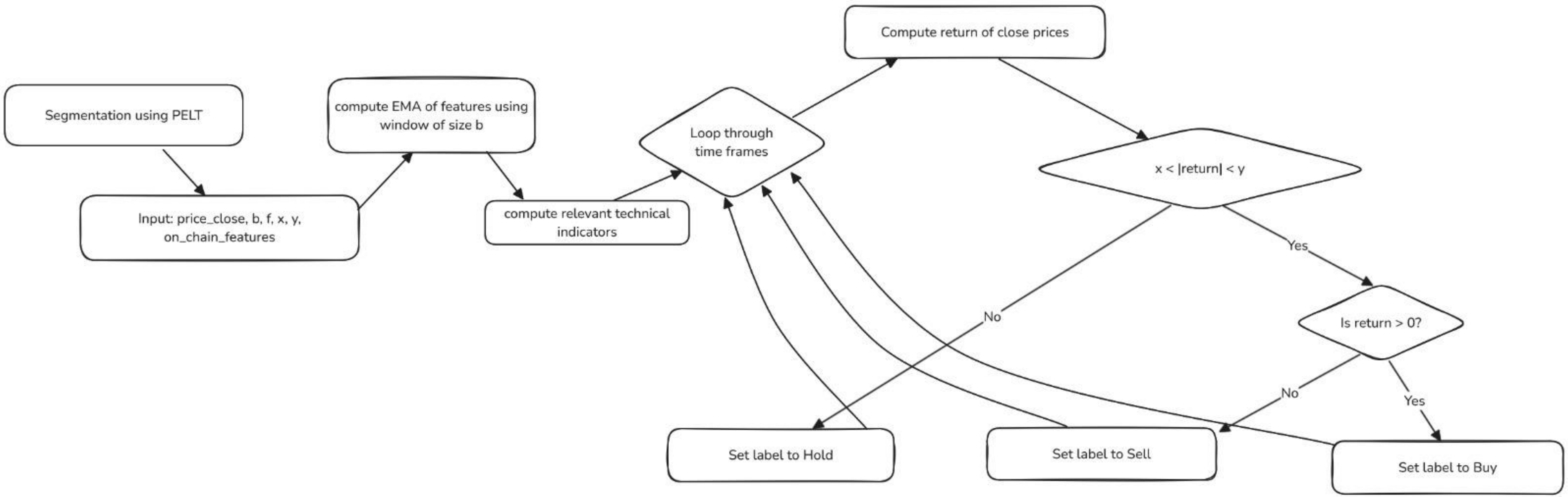
Model Construction

Our model is a 4-layer Multi-layer Perceptron (MLP).

- The input layer contains all the data discussed in our feature space
- The output layer consists of three nodes (buy / sell / hold) which will be interpreted as our position actions.

The two hidden middle layers at present have an unspecified number of nodes, but this will be a parameter that we can fine-tune as we test our strategy's performance.

Model Pipeline



Backtesting - Assumptions

For our backtesting, we plan to include the following parameters:

- f := trading fees
- l := stop loss threshold
- K := initial portfolio capital

This gives us the flexibility to closely replicate actual trading conditions and implement safeguards that may optimize performance in out-of-sample performance.

We also make the assumption that our movements have no market impact (prices are not retroactively changed due to our trades).

Risk Management & Portfolio Allocation

At present, our proposed model runs independently on each token, providing position actions (buy / sell / hold) at each time step. In order to combine these actions into a portfolio featuring all of the tokens, we choose to size our actions with two considerations:

- We first assign an "accuracy score" to each token based on its performance during training. When entering a new position with any given token, its sizing will be proportional to its accuracy score.
- When encountering successive "buy" classifications, we will increase the size of a tokens position in an exponentially decaying fashion. This allows us to take advantage of bullish predictions without banking our entire portfolio on a "hot streak" of any one token.

Our strategy also features a stop-loss condition, which is parameterized as discussed in the previous slide.

Sources

- Kim, G., Shin, D.-H., Choi, J. G., & Lim, S. (2022). A deep learning-*based cryptocurrency price prediction model that uses on-chain data*. IEEE Access, 10, 56232–56247. https://doi.org/10.1109/ACCESS.2022.3177888
- Parente, M., Rizzuti, L., & Trerotola, M. (2024). A profitable trading algorithm for cryptocurrencies using a neural network model. Expert Systems with Applications, 238, 121806. https://doi.org/10.1016/j.eswa.2023.121806