Università degli Studi di Bari Aldo Moro

Big Data Course

A.A. 2023-2024

Multi-View Time Series Classification of Cryptocurrencies: An Exploration with Bagging Classifier

Authors:
Katya Trufanova
Alberto G. Valerio

Supervisors:
Professor Michelangelo Ceci
PhD Student Antonio Pellicani

September, 2024

Contents

1 Introduction 3

2 Business Understanding 3

2.1 Business Objectives . . . 3

2.2 Assessing the Situation: Cryptocurrency Market . . . 3

2.3 Data Sources . . . 5

2.4 Data Mining Goals . . . 5

3 Data Understanding 5

3.1 Initial Data Collection . . . 5

3.2 Dataset Description . . . 6

3.3 Data Quality Verification . . . 7

4 Data Preparation 8

4.1 Data Selection . . . 8

4.1.1 AVAX . . . 8

4.1.2 BNB . . . 9

4.1.3 DOGE . . . 10

4.1.4 DOT . . . 10

4.1.5 NEO . . . 11

4.1.6 SOL . . . 11

4.2 Data Cleaning . . . 12

4.3 Data Construction . . . 12

4.3.1 Labeling . . . 12

4.3.2 Indicator Addition . . . 13

4.3.3 Normalization . . . 14

5 Modeling 15

5.1 Modeling Technique Selection . . . 15

5.2 Experimental Design . . . 15

5.3 Model Building . . . 17

5.3.1 Hyperparameter Settings . . . 17

5.3.2 Model Expectations . . . 17

5.4 Model Assessment . . . 18

6 Conclusion 24

1 Introduction

The rapid evolution of the cryptocurrency market has motivated an increase in data-driven research aimed at understanding and predicting these digital assets' behavior. This report delves into the application of data mining techniques to analyze the cryptocurrency market. This case study focuses on the examination of cryptocurrency price trends using Multi-View Time Series Classification (MVTSC), an approach that provides deeper insights by exploring different data view combinations. The volatile nature of the cryptocurrency market poses significant challenges, which this project aims to address by employing the Bagging Classifier model in a multi-view analysis context. The objective is to create a model that improves the accuracy and robustness of time series data classification for various cryptocurrencies. The ultimate business objectives are to enhance predictive accuracy in cryptocurrency market analysis and to advance the field of financial data mining and cryptocurrency research.

2 Business Understanding

2.1 Business Objectives

The goal of this case study is to apply data mining techniques to the analysis of the cryptocurrency market. The project focuses on examining cryptocurrency price trends using Multi-View Time Series Classification (MVTSC) to provide deeper insights by exploring different data view combinations. The study addresses the challenges of the volatile cryptocurrency market by employing the Bagging Classifier model in a multi-view analysis context. The objective is to create a model that improves the accuracy and robustness of time series data classification for various cryptocurrencies. Ultimately, the business objectives are to improve predictive accuracy in cryptocurrency market analysis and to advance the field of financial data mining and cryptocurrency research.

2.2 Assessing the Situation: Cryptocurrency Market

The cryptocurrency market is a rapidly evolving digital landscape that has revolutionized the concept of currency [1] [2]. It operates 24/7, unlike traditional financial markets, and is decentralized, meaning it's not governed by any central authority like a government or financial institution [3]. Cryptocurrencies, the digital assets that operate within this market, leverage blockchain technology to gain transparency, immutability, and distributed nature [3]. The most well-known cryptocurrency is Bitcoin, but there are thousands of other 'altcoins' available, each with their own unique features and uses [1] [2]. The market is highly volatile, with prices capable of significant fluctuations in a short period. This volatility, while potentially risky, also provides opportunities for high returns, making it an attractive market for traders and investors [4].

In the context of our project, understanding the dynamics of the cryptocurrency market is crucial. The goal is to classify time series data of different cryptocurrencies, which involves predicting the category or class of a time series sequence. Given the market's volatility, this is a challenging but potentially rewarding task [5]. Our focus will be on the price movements of different cryptocurrencies. Price is influenced by various factors, including supply and demand, market sentiment, regulatory news, technological advancements, and macroeconomic trends [3] [6] [7].

The cryptocurrency market consists of various participants, including retail investors, institutional players, miners, and developers. Retail investors are individuals who buy and sell cryptocurrencies for personal accounts. Institutional investors, such as hedge funds and family offices, have increasingly entered the market, attracted by its potential for high returns and diversification benefits. Miners play a crucial role in maintaining the blockchain network, and developers contribute to the evolution of the technology. The influx of institutional investors has led to a more mature market structure, influencing the volatility and liquidity of cryptocurrencies. It has also brought about increased market capitalization and trading volumes [8].

Technological advancements are at the heart of the cryptocurrency market. Innovations like the Lightning Network for Bitcoin have significantly improved transaction speeds, enabling faster and more efficient transactions. Similarly, Ethereum's transition to a proof-of-stake (PoS) consensus mechanism is a significant development. This transition not only reduces energy consumption but also improves the scalability of the network, allowing it to handle more transactions and accommodate more users. These technological advancements have a substantial impact on the market dynamics, influencing factors such as transaction fees, block times, and network security [9] [10].

Regulation is another key factor influencing the market. Countries around the world have varied stances on cryptocurrencies, ranging from outright bans to embracing them with open regulations. For instance, some countries have banned cryptocurrencies due to concerns about money laundering and financial stability, while others have embraced them, recognizing their potential for innovation and economic growth. Regulatory news can significantly impact market prices, as seen in the past with announcements from major economies like the United States, China, or the European Union. This regulatory landscape presents both challenges and opportunities for traders and investors in the space, influencing market sentiment and price volatility [11].

The emergence of the cryptocurrency derivatives market has been a game-changer for the industry. Platforms offering futures, options, and other derivative products have enabled traders to hedge their positions, speculate on price movements, and manage risk more effectively. These instruments provide traders with the ability to profit from price movements without owning the underlying asset. The growth of this market segment reflects the increasing sophistication and institutionalization of the crypto space, attracting more professional and institutional traders. It also contributes to the overall liquidity and depth of the market [12].

Market sentiment, driven by news, social media, and other public information, significantly impacts cryptocurrency prices. Tools like sentiment analysis algorithms and social media monitoring have become crucial for traders to gauge the market mood. These tools analyze data from various sources, including news articles, social media posts, and market data, to determine the overall sentiment towards a particular cryptocurrency. This sentiment can shift rapidly, often leading to abrupt price changes. Understanding market sentiment can provide traders with valuable insights, helping them make informed trading decisions [13].

There is a growing integration of cryptocurrencies with traditional financial markets. Financial institutions, including banks and brokerage firms, are increasingly offering crypto-related services, such as trading, custody, and asset management. The launch of products like Bitcoin ETFs has further bridged the gap between traditional finance and crypto markets, providing mainstream investors with easier access to cryptocurrencies. This integration is important for our project, as it affects the liquidity and stability of cryptocurrencies, making time series analysis more complex but also potentially more rewarding. It also signifies the growing acceptance and legitimacy of cryptocurrencies in the broader financial landscape [14].

2.3 Data Sources

Building on our understanding of the cryptocurrency market, it is clear that the data sources used for market analysis must be robust and comprehensive. For the purpose of this case study, data was sourced from three different exchanges, specifically focusing on candles and order books. In the context of cryptocurrencies, an exchange is a digital marketplace where traders can buy and sell cryptocurrencies using different fiat currencies or altcoins. For this study, data was sourced from three major exchanges: Binance, Huobi, and OKX [15].

'Candles', or candlestick charts, are a key tool in financial analysis, especially in the volatile cryptocurrency markets. Each candles represents price movements within a chosen timeframe, typically a day, but this can be adjusted. The candle has a body and wicks (or shadows) that show the open, close, high, and low prices for that period. This way of presenting data is crucial for understanding market trends and behaviors. It helps analysts and traders spot patterns and predict future market movements. When combined with volume and other market indicators, candle patterns can provide valuable insights into potential market sentiments, whether bullish or bearish. For this study, data was collected for nine different cryptocurrencies (ADA, AVAX, BNB, BTC, DOGE, DOT, ETH, NEO, SOL) from three major exchanges [16].

An order book is a list of buy and sell orders for a specific cryptocurrency at various prices. These are pending orders waiting to be fulfilled. An order book provides detailed insight into the supply and demand for a particular cryptocurrency at different price levels. For this study, snapshot data of the order books was collected for the same nine cryptocurrencies across the three exchanges [16].

2.4 Data Mining Goals

The objective of this case study is to perform Multi-View Time Series Classification (MVTSC) of market data for six different cryptocurrencies. MVTSC is a technique that aims to improve the performance of time-series classification by fusing the distinctive temporal information from multiple views. In traditional time-series classification, a single view of the data is used. However, in many real-world scenarios, data can be represented in multiple views, each capturing different aspects or features of the data. By considering these multiple views, the classification model can potentially learn more robust and comprehensive patterns [17].

Time series classification (TSC) is a type of supervised machine learning task where the goal is to predict or classify the class that a new time series data set belongs to. Time series data is a sequence of data points, typically consisting of successive measurements made over a time interval. Time series classification involves analyzing these sequences and assigning them to one of the predefined classes [18].

3 Data Understanding

3.1 Initial Data Collection

The initial data collection for this study was performed using the Crypto Lake API. Crypto Lake is a service providing historical cryptocurrency market data in high detail, including order book data, tick trades, and 1-minute trade candles. It is tuned for convenient quant and machine-learning purposes and offers high performance, caching, and parallelization [19].

3.2 Dataset Description

The original dataset used for this case study is a comprehensive collection of time-series data from three major cryptocurrency exchanges: BINANCE, HUOBI, and OKX. The data is organized into two primary categories: candles and snapshot, each containing detailed information on nine cryptocurrencies: ADA, AVAX, BNB, BTC, DOGE, DOT, ETH, NEO, and SOL.

The candles data, sourced from each of the three exchanges, provides a detailed view of market dynamics. It includes the following columns for each cryptocurrency:

- **origin_time:** The timestamp marking the beginning of the candle period.
- **open, high, low, close:** These represent the opening, highest, lowest, and closing prices within the candle's time frame.
- **volume:** The total trading volume of the cryptocurrency within the period.
- **trades:** Number of trades executed.
- **start, stop:** Timeframe indicators for the candle.
- **exchange:** The name of the exchange (BINANCE, HUOBI, OKX).
- **symbol:** Cryptocurrency symbol (e.g., ADA, BTC).
- **diff_perc:** Percentage difference in price over the period.

The snapshot dataset offers a detailed view of the order book for each cryptocurrency across the three exchanges. Each record includes:

- **origin_time:** The timestamp for the snapshot.
- **sequence_number:** A unique identifier for the order book entry.
- **Bid information:** Including bid_0_price, bid_0_size up to bid_19_size, bid_19_size, representing the price and size of the top 20 bids.
- **Ask information:** Including ask_0_price, ask_0_size up to ask_19_size, ask_19_size, representing the price and size of the top 20 bids.
- **exchange:** The name of the exchange.
- **symbol:** Cryptocurrency symbol.

Table 3.2 presents the class distribution of different cryptocurrencies across various exchanges and view types. The classes, represented as 'Down', 'Neutral', and 'Up', indicate the trend of each cryptocurrency. It is evident that the 'Neutral' class dominates the distribution for all cryptocurrencies across all exchanges and view types, with percentages mostly in the 70s.

VIEW	EXCHANGE	CRYPTO	DOWN	NEUTRAL	UP
CANDLES	BINANCE	AVAX	68408 (14.2%)	340097 (70.7%)	72395 (15.1%)
CANDLES	BINANCE	BNB	69482 (14.5%)	343118 (71.8%)	65420 (13.7%)
CANDLES	BINANCE	DOGE	71291 (14.9%)	336601 (70.1%)	72006 (15.0%)
CANDLES	BINANCE	DOT	64452 (13.4%)	349548 (72.7%)	66900 (13.9%)
CANDLES	BINANCE	NEO	50782 (10.6%)	381230 (79.8%)	45998 (9.6%)
CANDLES	BINANCE	SOL	73210 (15.2%)	337995 (70.3%)	69695 (14.5%)
CANDLES	HOUBI	AVAX	44438 (11.3%)	301665 (76.7%)	46956 (11.9%)
CANDLES	HOUBI	BNB	47649 (12.1%)	299989 (76.3%)	45419 (11.6%)
CANDLES	HOUBI	DOGE	50949 (13.0%)	290424 (73.9%)	51687 (13.1%)
CANDLES	HOUBI	DOT	47303 (12.0%)	297164 (75.6%)	48593 (12.4%)
CANDLES	HOUBI	NEO	35137 (9.5%)	303830 (82.1%)	31060 (8.4%)
CANDLES	HOUBI	SOL	54600 (13.9%)	285927 (72.7%)	52533 (13.4%)
CANDLES	OKX	AVAX	44438 (11.3%)	301664 (76.7%)	46956 (11.9%)
CANDLES	OKX	BNB	47649 (12.1%)	299990 (76.3%)	45420 (11.6%)
CANDLES	OKX	DOGE	50949 (13.0%)	290422 (73.9%)	51687 (13.1%)
CANDLES	OKX	DOT	47303 (12.0%)	297163 (75.6%)	48593 (12.4%)
CANDLES	OKX	NEO	35187 (9.5%)	305146 (82.2%)	31114 (8.4%)
CANDLES	OKX	SOL	54600 (13.9%)	285927 (72.7%)	52533 (13.4%)
SNAPSHOT	BINANCE	AVAX	68408 (14.2%)	340097 (70.7%)	72395 (15.1%)
SNAPSHOT	BINANCE	BNB	69482 (14.5%)	343118 (71.8%)	65420 (13.7%)
SNAPSHOT	BINANCE	DOGE	71291 (14.9%)	336601 (70.1%)	72006 (15.0%)
SNAPSHOT	BINANCE	DOT	64452 (13.4%)	349548 (72.7%)	66900 (13.9%)
SNAPSHOT	BINANCE	NEO	50782 (10.6%)	381230 (79.8%)	45998 (9.6%)
SNAPSHOT	BINANCE	SOL	73210 (15.2%)	337995 (70.3%)	69695 (14.5%)
SNAPSHOT	HOUBI	AVAX	44438 (11.3%)	301665 (76.7%)	46956 (11.9%)
SNAPSHOT	HOUBI	BNB	47649 (12.1%)	299989 (76.3%)	45419 (11.6%)
SNAPSHOT	HOUBI	DOGE	50949 (13.0%)	290424 (73.9%)	51687 (13.1%)
SNAPSHOT	HOUBI	DOT	47303 (12.0%)	297164 (75.6%)	48593 (12.4%)
SNAPSHOT	HOUBI	NEO	35137 (9.5%)	303830 (82.1%)	31060 (8.4%)
SNAPSHOT	HOUBI	SOL	54600 (13.9%)	285927 (72.7%)	52533 (13.4%)
SNAPSHOT	OKX	AVAX	44438 (11.3%)	301664 (76.7%)	46956 (11.9%)
SNAPSHOT	OKX	BNB	47649 (12.1%)	299990 (76.3%)	45420 (11.6%)
SNAPSHOT	OKX	DOGE	50949 (13.0%)	290422 (73.9%)	51687 (13.1%)
SNAPSHOT	OKX	DOT	47303 (12.0%)	297163 (75.6%)	48593 (12.4%)
SNAPSHOT	OKX	NEO	35187 (9.5%)	305146 (82.2%)	31114 (8.4%)
SNAPSHOT	OKX	SOL	54600 (13.9%)	285927 (72.7%)	52533 (13.4%)

Table 1: Class distribution of different cryptocurrencies.

3.3 Data Quality Verification

The dataset time intervals were analyzed, to ensure that they were as expected and consistent across the dataset. Additionally, a verification process was conducted to ensure that there were no duplicate values in the dataset.

A primary focus of our data quality verification involved the identification and handling of null values within our dataset. Our initial assumption was that the dataset was free from null values. However, a detailed examination revealed the presence of unexpected null values in various segments of the data. We observed a recurring pattern of null values, particularly in the initial and final rows of the "candles" files across all cryptocurrencies.

For instance, in the case of the BINANCE exchange, we noted null values in specific rows for several cryptocurrencies, such as AVAX, BNB, DOGE, DOT, NEO, and SOL. The occurrence of these null values was not uniform across the dataset; some cryptocurrencies exhibited null values at the beginning and end of their respective files, while others had sporadic occurrences throughout. For example, with AVAX, null values were found in the first 58 rows, last 11 rows, and at various other positions totaling 114 rows.

Following the discovery of these null values, the indicators were recalculated and the corrected version of the dataset was provided to us. The corrected dataset underwent another round of verification to confirm the resolution of previous issues. During this phase, we found no further issues, thus successfully confirming the quality of the data.

4 Data Preparation

4.1 Data Selection

While the candle data retained its original one-minute granularity as provided by the Crypto-Lake API, some data selection had to be performed for the order book data. While Crypto-Lake allows for the collection of order book data at millisecond granularity, for the purposes of this case study, it was required to align this data with the granularity of the candle data. Therefore, a decision was made to only retain the last order book entry for each minute. This approach was adopted because the final entry of each minute most accurately reflects the market conditions at the close of that minute, thereby providing a more precise and representative snapshot of the market dynamics at each time point.

For this case study, six cryptocurrencies were selected for experimentation from the original dataset. Details for each of these cryptocurrencies are provided in the sections below.

4.1.1 AVAX

Avalanche, also known as AVAX, launched in 2020, is a type of blockchain platform that hosts decentralized applications and custom blockchain networks. It was created by Ava Labs, a company founded by Cornell University professor Emin Gün Sirer, and Cornell University computer science PhD's Kevin Sekniqi and Maofan "Ted" Yin. Unlike other blockchains like Bitcoin and Ethereum, Avalanche is made up of three separate blockchains: the X-Chain, C-Chain, and P-Chain. Each of these chains serves a specific role and uses different consensus mechanisms based on their functions. The X-Chain is used for creating and trading digital smart assets. The C-Chain, compatible with Ethereum's EVM, enables users to create powerful, scalable applications. The P-Chain coordinates validators, keeps track of active subnets, and allows the creation of new subnets [20].

Figure 1: Official Avalanche logo [21].

One of the technical goals of Avalanche is to achieve a high transaction throughput, targeting up to 6,500 transactions per second. This is facilitated by the platform's architecture. Since its mainnet was launched in 2020, Avalanche has been building its own ecosystem of decentralized applications (DApps) and decentralized finance (DeFi). Decentralized applications, or DApps, are applications that run on a P2P network of computers rather than a single computer. Decentralized finance (DeFi) is a form of finance that does not rely on traditional financial intermediaries such as brokerages, exchanges, or banks, and instead utilizes blockchain technology. Avalanche has also integrated with several projects that were originally based on Ethereum, such as SushiSwap and TrueUSD [20].

The AVAX token is the main token of the Avalanche blockchain. It is used for transaction processing fees, securing the network, and as a basic unit of account within the Avalanche network. According to the most recent data, the live market cap of AVAX is $13,029,321,767 USD, and the current price of Avalanche is $35.54 USD [20].

While Avalanche is a competitor to Ethereum that prioritizes scalability and transaction processing speed, it's important to note that transaction fees and the rate of AVAX coin creation are determined using a governance model. This could lead to potential fluctuations in the value of AVAX [22].

4.1.2 BNB

Binance Coin, or BNB, is the main cryptocurrency coin that powers the BNB Chain ecosystem. BNB was created by Binance, the largest cryptocurrency exchange globally. The company was founded by Changpeng Zhao, who is the CEO, and He Yi, who is the chief marketing officer. Binance started as a crypto exchange in 2017, but it has since expanded its services to include a variety of functionalities. The Binance network is made up of the Binance Chain, Binance Smart Chain, Binance Academy, Trust Wallet, and Research projects. The Binance Smart Chain is a blockchain network built for running smart contract-based applications, which allows users to get the best of both worlds: the high transaction capacity of BC and the smart contract functionality of BSC. BNB can be used to pay for goods and services, settle transaction fees on Binance Smart Chain, participate in exclusive token sales, and more. Originally, BNB was introduced as a utility token on the Binance exchange, allowing users to receive discounts on trading fees [23].

Figure 2: Official BNB logo [24].

The Binance network employs a specific model to determine transaction fees and the rate of BNB coin creation. Each trade on the network incurs a standard fee of 0.1% for regular users, but users can avail a 25% discount if they use BNB to pay for their trading fees. The withdrawal fees are dynamic and are automatically adjusted based on the status of the blockchain network. The creation of BNB coins is governed by an Auto-Burn system, which aims to reduce the total supply of BNB to 100,000,000. The number of BNB coins to be burned is adjusted based on BNB's price and the number of blocks generated on the BNB Smart Chain during each quarter. This model helps maintain the value of BNB by controlling its supply. As of the most recent data, the live market cap of BNB is $48,103,512,673 USD, and the current price of Binance Coin is $317.12 USD [25].

4.1.3 DOGE

Dogecoin, often referred to as DOGE, is a digital currency that was initially created as a fun and light-hearted cryptocurrency. It was developed by Billy Markus from Portland, Oregon, and Jackson Palmer from Sydney, Australia, and it was forked from Litecoin in December 2013. The logo of Dogecoin features a Shiba Inu dog from the "Doge" meme, which is where the name of the cryptocurrency comes from [26].

Figure 3: Official Dogecoin logo [27].

Dogecoin uses Scrypt technology and differs from Bitcoin's proof-of-work protocol in several ways. It has a block time of 1 minute, and there is no limit to the number of Dogecoin that can be mined. This means that Dogecoin can be mined either solo or by joining a mining pool. Dogecoin has been primarily used as a tipping system on Reddit and Twitter to reward the creation or sharing of quality content. It can be bought or sold at any exchange that offers the digital currency, and it can be stored on an exchange or in a Dogecoin wallet. The live market cap of DOGE is $11,628,019,079 USD, and the current price of Dogecoin is $0.081508 USD [26].

4.1.4 DOT

Polkadot, commonly known as DOT, is a digital currency that operates as an open-source multichain protocol. Its primary function is to facilitate the connection and security of a network of specialized blockchains. The Web3 Foundation, a Swiss Foundation dedicated to enabling a fully functional and user-friendly decentralized web, developed Polkadot. The founders of Polkadot include Dr. Gavin Wood, Robert Habermeier, and Peter Czaban, with Wood also serving as the president of the Web3 Foundation. Wood is recognized for his work as a co-founder of Ethereum and as the creator of the Solidity smart contract coding language [28].

Figure 4: Official Polkadot logo [29].

The Polkadot network is structured differently from other blockchains. It consists of four core components: the Relay Chain, Parachains, Parathreads, and Bridges.The Relay Chain is Polkadot's main blockchain where transactions are permanently recorded. Parachains are individual blockchains that run in parallel within the Polkadot ecosystem, each serving a unique function. Parathreads are similar to parachains but operate on a pay-as-you-go basis, making them more economical for blockchains that don't require continuous connectivity. Bridges are special blockchains that allow parachains and parathreads to connect and communicate with external blockchains, such as Ethereum [28].

DOT is the native token of the Polkadot blockchain. It is used for staking for operations and security, facilitating network governance, and bonding tokens to connect parachains. As of the most recent data, the live market cap of DOT is $9,886,918,423 USD, and the current price of Polkadot is $7.63 USD. Real-world applications of Polkadot include Acala, a decentralized finance hub and stablecoin platform, and Moonbeam, an Ethereum-compatible smart contract platform. [28].

4.1.5 NEO

NEO, often referred to as the "Chinese Ethereum", is a blockchain platform that enables the development of digital assets and smart contracts. It was created by Da Hongfei and Erik Zhang in 2014 and was originally known as Antshares. The platform was rebranded as NEO in 2017 [30].

Figure 5: Official NEO logo [31].

NEO's architecture is unique in that it uses a Delegated Byzantine Fault Tolerance (dBFT) consensus mechanism, which is designed to support a large number of transactions simultaneously. The Delegated Byzantine Fault Tolerance (dBFT) consensus mechanism is like a voting system where certain nodes, known as delegates, are voted in to validate transactions. This system ensures that even if some nodes aren't functioning correctly or are acting maliciously, the majority can reach a consensus, ensuring the network's reliability. This mechanism also ensures that the system maintains consensus even if some nodes fail or act maliciously. One of the key features of NEO is its dual-token system, which includes the NEO token and the GAS token. The NEO token represents ownership in the NEO blockchain and is used for network management and changes, while the GAS token is used to pay for operations and services on the network [30].

NEO has a strong focus on community development and has a global team of developers contributing to its growth. It also provides robust support for developers, including comprehensive documentation and a variety of programming languages for writing smart contracts. As of the most recent data, the live market cap of NEO is $850,447,486 USD, and the current price of NEO is $12.06 USD. Furthermore, NEO's dBFT consensus mechanism is environmentally friendly as it requires less energy than the Proof-of-Work systems used by many other blockchains, making NEO a more sustainable choice [30].

4.1.6 SOL

Solana, also known as SOL, is a blockchain platform that supports the development of decentralized applications and smart contracts. It was developed by the Web3 Foundation, which is based in Switzerland. The founders of Solana are Anatoly Yakovenko and Greg Fitzgerald, who were colleagues at Qualcomm before they started working on Solana. The platform was launched to the public in 2020 [32].

Figure 6: Official Solana logo [33].

The Solana network is unique in its structure. It uses a hybrid consensus model that combines proof-of-history (PoH) and proof-of-stake (PoS) consensus mechanisms. The Proof-of-History (PoH) mechanism in Solana is like a clock for the network, creating a historical record of events. It uses a cryptographic function to generate timestamps for each block in the blockchain, ensuring the order of events [34]. The Proof-of-Stake (PoS) mechanism, on the other hand, is a system where participants, known as validators, show their commitment by locking up a certain amount of their cryptocurrency. Validators are then randomly chosen to validate transactions and create new blocks, and they get rewarded for their work [35]. This model allows Solana to achieve consensus quickly while reducing the workload, making it attractive to both small-time and institutional traders [32].

SOL is the native token of the Solana blockchain. It is used for transaction fees, staking for operations and security, and facilitating network governance. As of the most recent data, the live market cap of SOL is $41,906,854,670 USD, and the current price of Solana is $96.85 USD [32].

4.2 Data Cleaning

As mentioned in the previous section, null values were removed from the dataset to ensure the integrity and accuracy of the analysis and modeling. This step was crucial to avoid any distortions or inaccuracies in the results that could arise from incomplete data. To handle null values, the "forward fill" strategy was employed. In this method, each null value in the data is replaced with the last observed non-null value. If there are consecutive null values, they are filled with the same last observed non-null value. However, if a null value is the first value of the series, it remains as a null since there is no previous value to forward fill with. The forward fill strategy is commonly used in time series data where the observations have a temporal order. It is based on the assumption that the value of a certain parameter does not change drastically within a short period of time. Therefore, it is reasonable to assume that the value at a missing timestamp is the same as the last observed value [36].

4.3 Data Construction

4.3.1 Labeling

The data that was provided to us for experimentation included labels that were added to categorize the hourly variations of each cryptocurrency into distinct classes based on their price. This is an important step for the subsequent phase of modeling, as it provides a target variable for the classification task.

The labeling approach was based on the statistical analysis of each cryptocurrency's hourly price variations.

The average increase and decrease for each cryptocurrency were calculated. These averages were used as benchmarks to categorize the hourly performance of the cryptocurrencies. The labeling criteria were as follows:

- An instance was labeled as 'up' if the hourly variation exceeded the average increase with reference to the average of each cryptocurrency.
- A 'down' label was assigned if the hourly variation was less than the average decrease with reference to the average of each cryptocurrency.
- If the variation did not meet either of the above criteria, it was labeled as 'neutral'.

4.3.2 Indicator Addition

During the data preparation phase, the dataset was enriched with additional indicators. This step was important for enhancing the quality of the analysis and providing deeper insight into the movements of cryptocurrency prices. The integrated indicators are widely recognized in financial analysis, particularly in cryptocurrency trading. A brief description of each is provided below.

- **Simple Moving Average (SMA_x):** The average price over specific time periods was calculated, simplifying the price data to understand trends better. Time windows of 5, 15, 30, and 60 minutes were chosen.
- **Exponential Moving Average (EMA_x):** Similar to SMA, this gives more weight to recent prices, aiding in capturing momentum and short-term trends. Time windows of 5, 15, 30, and 60 minutes were utilized.
- **Relative Strength Index (RSI):** This indicator, used for identifying overbought or oversold market conditions, was calculated over a 14-minute window with a standard deviation of 2.
- **Moving Average Convergence/Divergence (MACD):** Changes in momentum were spotted by comparing two moving averages, with a fast period of 12, a slow period of 26, and a signal period of 9.
- **Stochastic Oscillator (STOCH):** This compares the closing price to a range of prices over a certain period, aiding in identifying potential reversal points. Parameters were set with k at 14 and d at 3.
- **Bollinger Bands (BBL_x, BBM_x, BBU_x):** These indicators provide a view of how high or low the price is relative to the past. A length of 20 and a standard deviation of 2 were set.
- **Momentum (MOM):** The rate of rise or fall in prices was measured, with the length set at 10.
- **Chande Momentum Oscillator (CMO):** Providing a more refined view similar to Momentum, the length was set at 9.
- **Detrend Price Oscillator (DPO):** The underlying trend was identified by removing the direct influence of the long-term trend, with a length of 21.
- **Ultimate Oscillator (UO):** This integrates three different time frames for a comprehensive view of market momentum, with fast, medium, and slow periods set at 7, 14, and 28, respectively.

4.3.3 Normalization

Normalization was an important step during the data preparation phase of our project. The primary purpose of normalization in this context is to scale the numerical values of the cryptocurrency data, sourced from multiple exchanges, into a common range. This procedure is required for ensuring comparability and eliminating potential biases arising from the varying scales of the data.

For our project, min-max normalization was employed. This technique was applied individually to each cryptocurrency across all exchanges. The approach involved defining a universal minimum and maximum normalization value for each cryptocurrency, independent of the exchange. These normalization boundaries were determined based on the observed data ranges across all exchanges for each cryptocurrency.

To implement this, we first identified the minimum and maximum values recorded for each cryptocurrency on each exchange. Subsequently, a universal minimum normalization value that was lower than the lowest observed minimum across all exchanges for a given cryptocurrency was selected. Similarly, the maximum normalization value was chosen to be higher than the highest observed maximum for that cryptocurrency across all exchanges.

In Table 4.3.3, the normalization process is exemplified. The table showcases the observed minimum and maximum values for various cryptocurrencies across different exchanges, alongside the chosen normalization minimum and maximum values. This tabular representation provides a clear overview of the normalization parameters and their application across the dataset.

Cryptocurrency	Exchange	Observed Min.	Observed Max.	Norm. Min.	Norm. Max
ADA	BINANCE	0.22	0.4617	0	0.5
	HUOBI	0.221011	0.461766		
	OKX	0.21489	0.46211		
AVAX	BINANCE	9.23	22.79	9	25
	HUOBI	9.2401	22.8272		
	OKX	9.239	22.77		
BNB	BINANCE	203.4	398.3	200	400
	HUOBI	203.44	350.2		
	OKX	203.36	350.53		
BTC	BINANCE	15476.0	31804.2	15000	35000
	HUOBI	16500.0	31906.47		
	OKX	16502.0	31830.0		
DOGE	BINANCE	0.053	0.148	0	0.2
	HUOBI	0.056	0.104689		
	OKX	0.05281	0.10491		
DOT	BINANCE	4.103	7.9	4	8
	HUOBI	4.1041	7.8882		
	OKX	4.087	7.897		
ETH	BINANCE	1073.53	2141.54	1000	2500
	HUOBI	1191.0	2142.93		
	OKX	1191.0	1191.27		
NEO	BINANCE	5.94	15.8	5	16
	HUOBI	6.38	15.77		
	OKX	6.255	15.757		
SOL	BINANCE	8.0	38.79	8	40
	HUOBI	9.6909	32.2362		
	OKX	9.693	32.9		

Table 2: Normalization Parameters for Cryptocurrencies Across Exchanges

5 Modeling

5.1 Modeling Technique Selection

The modeling technique selected for this study is the Bagging Classifier model from the sklearn library. This model was chosen due to its ability to reduce the variance of a black-box estimator by introducing randomization into its construction procedure and then making an ensemble out of it [37] [38]. The Bagging Classifier model was trained using the HistGradientBoostingClassifier as the base estimator. The base estimator was fitted on random subsets of the original dataset, and their individual predictions were aggregated to form a final prediction. The model was trained on the training dataset and then used to predict the labels of the testing dataset.

In addition to the default parameters of the Bagging Classifier model, a grid search was performed to tune the hyperparameters of the HistGradientBoostingClassifier. The hyperparameters tuned included the learning rate, maximum number of iterations, and maximum leaf nodes. The selection of the Bagging Classifier model and the subsequent hyperparameter tuning aimed to create a model that improves the accuracy and robustness of time series data classification for various cryptocurrencies.

5.2 Experimental Design

To implement Multi-View Time Series Classification (MVTSC), specific combinations of data views were selected from the available set. Out of 63 possible view combinations, a subset was chosen based on their potential to provide diverse yet complementary perspectives on market dynamics. The goal of this selection was to maximize the informational value and ensure a broad coverage of market behaviors across different exchanges.

The selection process for the view combinations was guided by the decision to avoid irrelevant combinations between candles from one exchange and order books from another. For instance, combinations like 'Candles BINANCE - Orderbook HUOBI' were deliberately excluded. This approach was adopted to maintain the integrity and relevance of the data views, ensuring that each combination provided meaningful insights into market dynamics specific to a single exchange or a consistent set of exchanges.

The selected view combinations, as detailed in the list below, include various permutations of snapshot and candle data from the three exchanges (BINANCE, HUOBI, OKX). These combinations range from single-view scenarios, such as snapshot data from a single exchange, to complex multi-view scenarios integrating both snapshot and candle data from multiple exchanges.

1. snapshot BINANCE
2. snapshot HUOBI
3. snapshot OKX
4. candles BINANCE
5. candles HUOBI
6. candles OKX

7. snapshot BINANCE, snapshot HUOBI
8. snapshot BINANCE, snapshot OKX
9. snapshot BINANCE, candles BINANCE
10. snapshot HUOBI, snapshot OKX
11. snapshot HUOBI, candles HUOBI
12. snapshot OKX, candles OKX
13. candles BINANCE, candles HUOBI
14. candles BINANCE, candles OKX
15. candles HUOBI, candles OKX
16. snapshot BINANCE, snapshot HUOBI, snapshot OKX
17. candles BINANCE, candles HUOBI, candles OKX
18. snapshot BINANCE, snapshot HUOBI, candles BINANCE, candles HUOBI
19. snapshot BINANCE, snapshot OKX, candles BINANCE, candles OKX
20. snapshot HUOBI, snapshot OKX, candles HUOBI, candles OKX
21. snapshot BINANCE, snapshot HUOBI, snapshot OKX, candles BINANCE, candles HUOBI, candles OKX

For each selected view, the utilized features included typical candle data points (open, high, low, close, volume, trades) and technical indicators (SMA, EMA, RSI, MACD, STOCH, Bollinger Bands, Momentum, CMO, DPO, UO). In the order book data, we concentrated on the top 20 bid and ask prices and sizes.

Despite the initial plan to execute all the experiments listed above, we encountered a system limitation that prevented us from executing certain experiments. Specifically, an "out of memory" error occurred on both project team members' systems (MacOS 64GB RAM, Windows 16GB RAM). This error led to the inability to execute experiments involving the following view combinations:

- snapshot BINANCE, snapshot HUOBI, snapshot OKX
- snapshot BINANCE, snapshot HUOBI, candles BINANCE, candles HUOBI
- snapshot BINANCE, snapshot OKX, candles BINANCE, candles OKX
- snapshot HUOBI, snapshot OKX, candles HUOBI, candles OKX
- snapshot BINANCE, snapshot HUOBI, snapshot OKX, candles BINANCE, candles HUOBI, candles OKX

Upon encountering this issue, we attempted to mitigate the problem by adjusting some hyperparameters, hoping to avoid the out of memory error. However, the issue persisted. Consequently, it was agreed with the project supervisor that we would instead perform hyperparameter tuning for the experiments that we were able to execute without encountering the error.

In the experimental design, a window size of 10 was utilized for the experiments. This means that each input to the model consisted of the data from the 10 consecutive time steps preceding the prediction point. For the train-test split, the last month of the time series data was used for testing, while the rest of the data was used for training. This approach ensures that the model is evaluated on the most recent data, which is crucial in time series forecasting due to the temporal dependencies in the data. It also mimics a realistic scenario where the model is trained on historical data and used to make predictions about future events.

5.3 Model Building

5.3.1 Hyperparameter Settings

The hyperparameter tuning process was designed to optimize the performance of the Bagging Classifier model. The parameters tuned included the learning rate, maximum iterations, and maximum leaf nodes of the underlying HistGradientBoostingClassifier. The specific values tested for each parameter were as follows:

- **Learning rate:** 0.5, 0.001
- **Maximum iterations:** 50, 150
- **Maximum leaf nodes:** 10, 100

These parameters were selected based on their potential impact on the model's learning process and generalization ability. The learning rate controls the step size at each iteration while moving toward a minimum of a loss function. The maximum iterations parameter determines the number of boosting stages to perform, and the maximum leaf nodes parameter limits the number of leaves for each tree in the ensemble. The hyperparameter tuning was performed for each of the view combinations that did not encounter the "out of memory" error. For each combination, the model was trained and tested using each possible configuration of the hyperparameters. The performance metrics (accuracy, precision, recall, F1 score) were then recorded for each configuration. This comprehensive experimental design, although constrained by system limitations, allowed us to thoroughly explore the performance of the Multi-View Time Series Classification approach across a variety of scenarios and hyperparameter configurations.

5.3.2 Model Expectations

The Bagging Classifier model, with its underlying HistGradientBoostingClassifier, is expected to deliver a reasonable level of accuracy in predicting cryptocurrency market trends. The model's accuracy is dependent on the selected hyperparameters and the specific view combinations used for training. For instance, a higher learning rate might allow the model to learn faster, but it could also lead to overfitting, reducing the model's generalization ability. On the other hand, a lower learning rate might improve the model's robustness by preventing overfitting, but it could also slow down the learning process and potentially lead to underfitting.

The maximum iterations parameter also plays an important role in the model's performance. A higher number of iterations allows the model to learn more complex patterns in the data, potentially improving accuracy. However, this could also lead to overfitting if the model becomes too complex. On the other hand, a lower number of iterations might prevent overfitting but could also limit the model's ability to capture complex patterns, potentially reducing accuracy.

The maximum leaf nodes parameter controls the complexity of the trees in the ensemble. A higher number of leaf nodes allows for more complex trees, potentially improving the model's ability to capture intricate patterns in the data. However, this could also lead to overfitting. A lower number of leaf nodes results in simpler trees, which might improve the model's robustness by preventing overfitting but could also limit the model's predictive power.

Despite the potential for high accuracy, the model's robustness might be compromised if the hyperparameters are not appropriately tuned. Overfitting is a significant concern, especially given the volatile and complex nature of cryptocurrency markets. If the model becomes too complex, it might perform well on the training data but fail to generalize to new, unseen data. Furthermore, the model's performance might be limited by the quality and relevance of the data views used for training. The selection of view combinations was guided by the need to maintain the integrity and relevance of the data views. However, some potentially informative view combinations were excluded due to system limitations, which might have impacted the model's performance.

Lastly, the model's performance is inherently limited by the unpredictable nature of cryptocurrency markets. While the model can learn patterns from historical data, it might not always accurately predict future trends due to the influence of unforeseen events and market dynamics.

5.4 Model Assessment

In this section, we present an evaluation of the performance of our model based on several key metrics, including accuracy, precision, recall, and F1 score. These metrics provide insights into how well our model is performing in terms of its ability to correctly classify the state of cryptocurrencies (i.e., 'up', 'down', or 'neutral') based on multi-view time series data.

Figure 7 showcases a visual comparison between the performance obtained for different view combinations, including metrics such as accuracy, precision, recall, and F1 score across various view combinations, identified with different letters, according to Table 3.

ID	VIEW
A	candles BINANCE
B	candles HUOBI
C	candles OKX
D	snapshot BINANCE
E	snapshot HUOBI
F	snapshot OKX
G	candles BINANCE, candles HUOBI
H	candles BINANCE, candles OKX
I	candles HUOBI, candles OKX
J	snapshot BINANCE, snapshot HUOBI
K	snapshot BINANCE, snapshot OKX
L	snapshot HUOBI, snapshot OKX
M	snapshot BINANCE, candles BINANCE
N	snapshot HUOBI, candles HUOBI
O	snapshot OKX, candles OKX
P	candles BINANCE, candles HUOBI, candles OKX

Table 3: View combinations for Multi-View Time Series Classification

- **Average Accuracy by View:** The 'neutral' class (orange) consistently demonstrates high accuracy

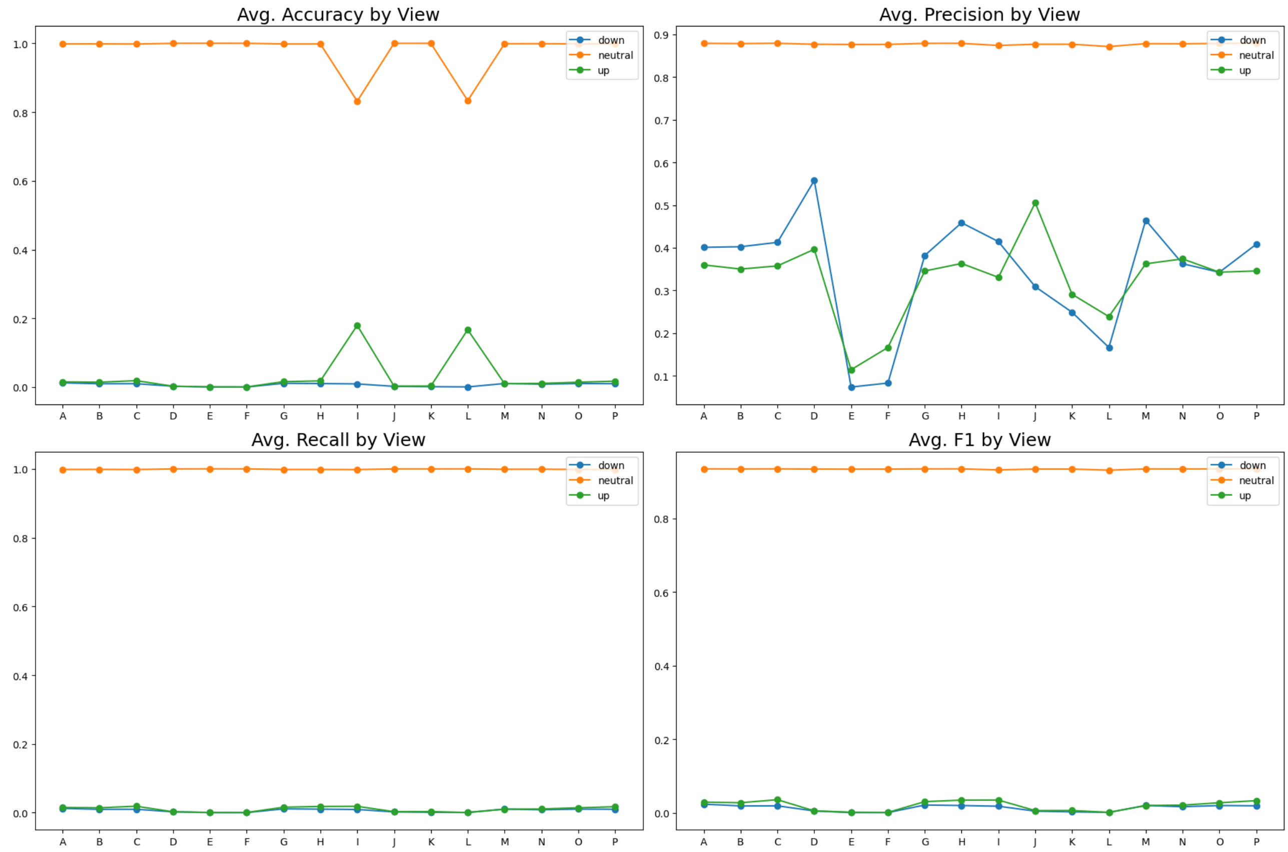

Figure 7: Average accuracy, precision, recall, and F1 score by view.

across all view combinations, maintaining a value close to 1.0, with dips to around 0.8 for the I (candles HOUBI, candles OKX) and L (snapshot HOUBI, snapshot OKX) view combinations. The 'down' (blue) and 'up' (green) classes exhibit significantly lower accuracy, with values hovering around the 0.0 mark and two peaks to about 0.2 for the I and L view combinations.

- **Average Precision by View:** Precision for the 'neutral' class (orange) is presents very little variance across different views, remaining close to 0.9, which, as was the case for accuracy, is the highest among the three classes. The precision for the 'down' (blue) and 'up' (green) classes fluctuates significantly. The best precision for the 'down' category is achieved by the D view combination (snapshot BINANCE, single view) with a value close to 0.6, while for the 'up' category the winning combination is J (snapshot BINANCE, snapshot HOUBI), with a precision close to 0.5.
- **Average Recall by View:** Recall for the 'neutral' category (orange) is consistently very high across all views, with values close to 1.0. Both 'down' (blue) and 'up' (green) categories exhibit extremely low recall rates, with values close to 0.
- **Average F1 Score by View:** Similar to recall and precision trends, F1 score for 'neutral' (orange) is consistently high, nearing a score of 1.0. Conversely, performance for both 'down' (blue) and 'up' (green) is very low, with D, E, F, J, K, and L view combinations performing especially poorly. In this case, slightly better performance for the 'up' class is encountered for view combinations C, G, H, and I.

The consistent high performance of the model in predicting the "neutral" category indicates that certain features or patterns associated with stable cryptocurrency prices are well-captured and recognized by the model. This high performance can be attributed to the fact that the ‘neutral’ class is the majority class in the dataset. As shown in Table 3.2 in the 3.2 section, the ‘neutral’ class has a significantly higher distribution compared to the ‘down’ and ‘up’ classes across all cryptocurrencies and exchanges. This imbalance in the class distribution means that the model has more examples from which to learn the characteristics of the ‘neutral’ class, leading to better performance in predicting this class. However, there’s room to improve classification performance for “down” and “up” categories as indicated by their lower metric scores. This could be due to the volatile nature of cryptocurrency prices which makes it challenging to predict drastic increases or decreases.

Figure 8 presents a comparison of model performance for each of the six analyzed cryptocurrencies, showcasing accuracy, precision, recall, and F1 score for each cryptocurrency for each of the three classes.

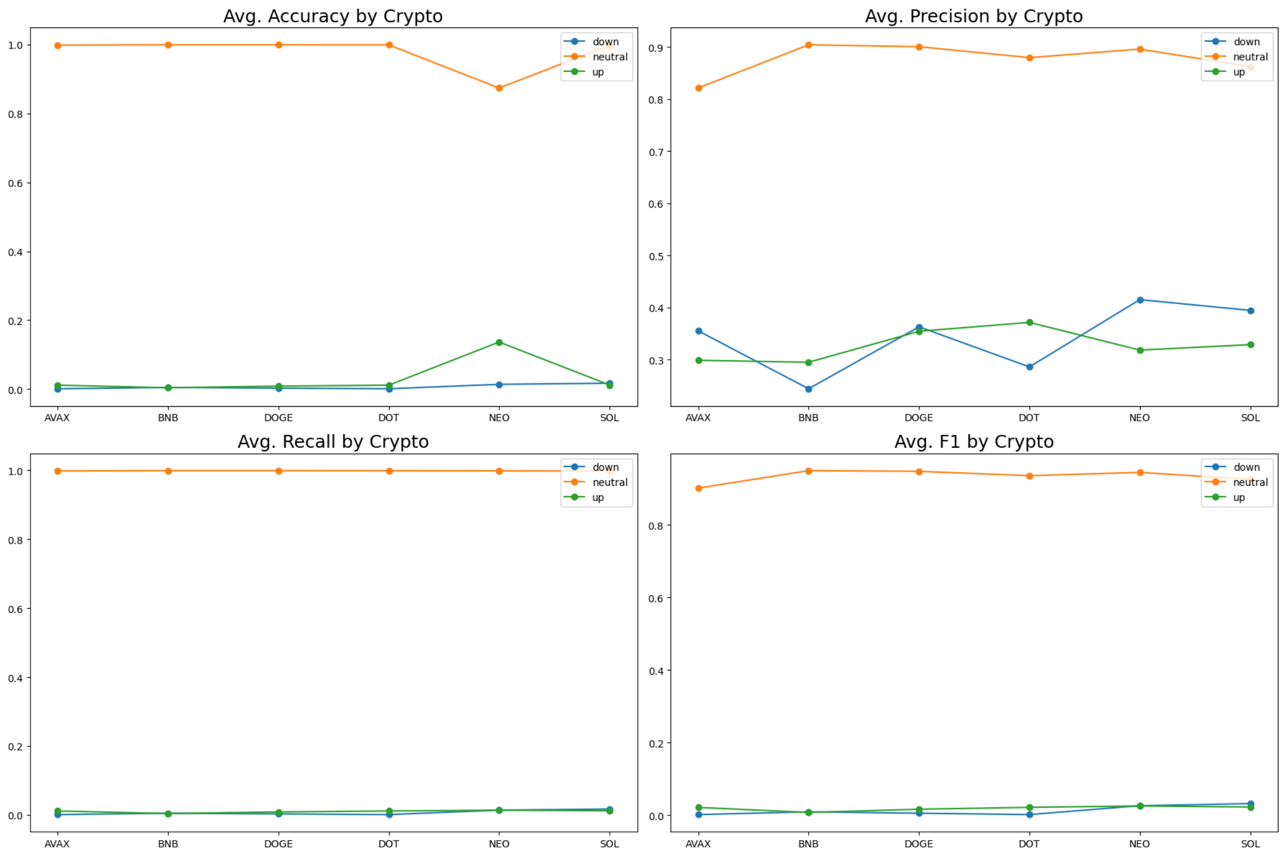

Figure 8: Model performance by cryptocurrency.

The model performed exceptionally well in classifying the ‘neutral’ state across all cryptocurrencies, as indicated by consistently high scores in all metrics, with slight dips in average precision and average F1 score for the AVAX currency and a slight dip in accuracy for the NEO cryptocurrency.

The model struggled to accurately predict the ‘up’ and ‘down’ classes for all cryptocurrencies, with accuracy, recall, and F1 score being extremely low for all currencies, with a small exception for the ‘up’ class for the NEO cryptocurrency, for which there was a slight increase. As for average precision, the results were more varied, but values did not exceed 0.5, with NEO having the best performance for the ‘down’ class and DOT showcasing the best result for the ‘up’ category.

Figure 9 presents the gains in average accuracy, precision, recall, and F1 score across different view combinations (A to P) for two scenarios: using default model parameters (blue line) and best model parameters after hyperparameter tuning (orange line).

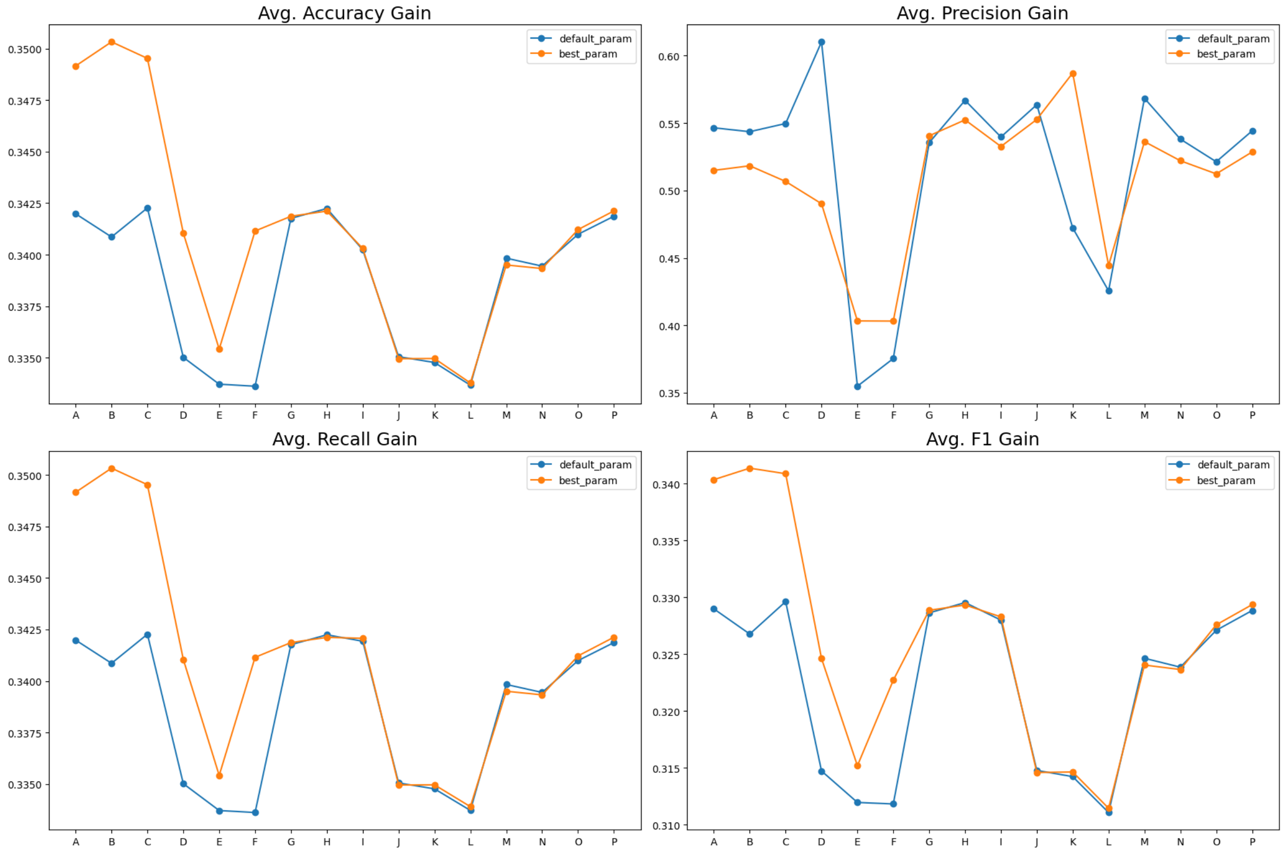

Figure 9: Gains in average accuracy, precision, recall, and F1 score across different view combinations.

- **Average Accuracy Gain:** The tuned parameters outperform the default parameters for view A to F (all single views), but have comparable performance for the remaining view combinations (G to P, multi-view combinations). The best results can be observed for the first three views, while views E, J, K, and L had the poorest performance.

- **Average Precision Gain:** The default parameters show a dramatic increase at view D, far exceeding the performance of the best parameters and showing the biggest precision gain overall, but also have the worst performance for view combination E. The best performance with the tuned parameters is encountered for view K.

- **Average Recall Gain:** Similar trends as seen in accuracy gain; the tuned parameters outperform the default parameters for single data views, but have comparable performance for multi-view combinations. Significant gains for the tuned parameters are observed at views A, B, and C, far exceeding the performance of the default parameters.

- **Average F1 Score Gain:** Here, the same pattern of accuracy gain and recall gain can be observed, with the best performance obtained for the first three single views for the tuned parameters. The words results are obtained for view combination L (snapshot HOUBI, snapshot OKX).

Overall, the tuning of model parameters has a clear positive impact on average accuracy gain, average recall gain and average F1 score gain for single data views. For multi-view combinations, the performance gain is less pronounced, indicating that the default parameters already provide a robust baseline for these more complex scenarios. However, it's worth noting that the best performance with the tuned parameters is encountered for average precision gain for view K (snapshot BINANCE, snapshot OKX), suggesting that some multi-view combinations can still benefit from hyperparameter tuning.

Figure 10 presents the performance metrics – accuracy, precision, recall, and F1 score – plotted against execution time (log base 2) for different view combinations (A to P).

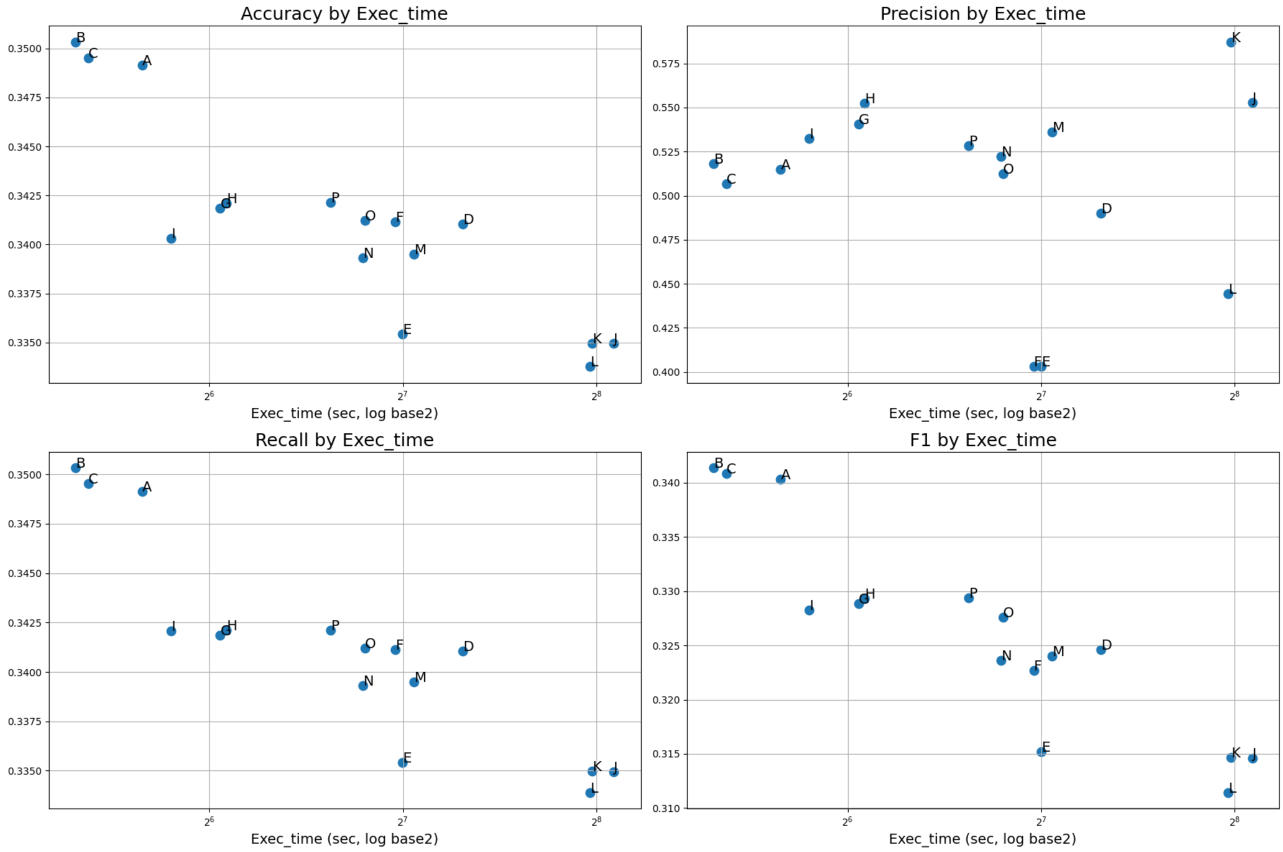

Figure 10: Performance metrics vs. execution time for different view combinations.

- **Accuracy by Execution Time:** The best performance is observed for view B (candles HUOBI), while also having the shortest execution time. Views A (candles BINANCE) and C (candles OKX) have similar accuracy levels and execution times. The lowest accuracy along with the longest execution time is encountered for views K, J, and L.

- **Precision by Execution Time:** The highest precision is observed for view combination K (snapshot BINANCE, snapshot OKX), but it also takes a high execution time. Conversely, the lowest precision is achieved for view E, while also having a significant execution time.

- **Recall by Execution Time:** View B (candles HUOBI) stands out with the highest recall and the shortest execution time. Views A (candles BINANCE) and C (candles OKX) follow closely in terms of recall but have slightly longer execution times. The recall drops significantly for views K, J, and L, which also have extended execution times.

- **F1 Score by Execution Time:** Consistent with accuracy and recall, view B achieves the highest F1 score at minimal execution time. Views A and C are comparable in their F1 scores but require slightly more time to execute. The performance dips for views K, J, and L as indicated by their lower F1 scores and increased execution times.

The results suggest that the choice of view combination has a significant impact on both the performance of the model and the execution time. For instance, view B (candles HUOBI) consistently achieves high performance across all metrics (accuracy, precision, recall, and F1 score) while maintaining a short execution time. This indicates that the candles HUOBI view is particularly effective for this classification task and can be processed efficiently. On the other hand, view K (snapshot BINANCE, snapshot OKX) shows the highest precision but also takes a longer execution time. This suggests that while this view combination can provide precise predictions, it may require more computational resources or time to process.

Views A (candles BINANCE) and C (candles OKX) offer a balance between performance and execution time, performing well across all metrics and having moderate execution times. However, view L shows very poor performance across all metrics and also takes longer to execute. This could be due to the increased complexity when combining snapshot data from different exchanges.

In Table 4 the results of the hyperparameter tuning are shown, presenting the best values for the learning rate, max iterations, and max leaf nodes hyperparameters for each data view combination.

ID	**learning_rate**	**max_iter**	**max_leaf_nodes**
A	0.5	50	10
B	0.5	50	100
C	0.5	50	100
D	0.5	50	100
E	0.5	50	100
F	0.5	50	100
G	0.5	50	100
H	0.5	50	10
I	0.5	50	10
J	0.5	50	10
K	0.5	50	10
L	0.5	50	10
M	0.01	100	10
N	0.01	100	31
O	0.01	50	31
P	0.5	100	10

Table 4: Best hyperparameter values for different view combinations.

The results in Table 4 highlight the importance of hyperparameter tuning in model performance. It's evident that different view combinations benefit from different hyperparameters. For instance, a learning rate of 0.5 is optimal for most view combinations, but for views M, N, and O, a significantly smaller learning rate of 0.01 yields the best results. This suggests that for these specific views, the model benefits from a more gradual learning process.

In terms of the maximum number of iterations (max_iter), most views achieve the best performance with 50 iterations. However, views M, N, and P require a higher number of iterations (100), indicating that these views may have more complex patterns that need a larger number of iterations to learn effectively. The maximum leaf nodes (max_leaf_nodes) also vary across views. While a smaller tree with 10 leaf nodes is sufficient for some views (A, H, I, J, K, L, M, P), others benefit from a larger tree with 100 leaf nodes (B, C, D, E, F, G) or an intermediate size of 31 leaf nodes (N, O).

6 Conclusion

This case study aimed to enhance the predictive accuracy in cryptocurrency market analysis and advance the field of financial data mining and cryptocurrency research. The study employed the Bagging Classifier model in a multi-view analysis context to improve the accuracy and robustness of time series data classification for various cryptocurrencies. The Bagging Classifier model demonstrated high performance in predicting the "neutral" category, indicating that certain features or patterns associated with stable cryptocurrency prices are well-captured and recognized by the model. However, there is room to improve classification performance for "down" and "up" categories, which could be due to the volatile nature of cryptocurrency prices. Hyperparameter tuning also played a significant role in model performance, as different view combinations benefited from different hyperparameters.

In conclusion, this study demonstrates the potential of using a Bagging Classifier model in a multi-view analysis context for cryptocurrency market analysis. While the results are promising, they also highlight the complexities and challenges inherent in predicting cryptocurrency prices. Future research could focus on improving the prediction of "down" and "up" categories by exploring even more data view combinations or different models.

References

[1] *Cryptocurrency prices, charts and market capitalizations*, Accessed: 2024-01-17. [Online]. Available: https://coinmarketcap.com/.

[2] *All crypto prices and market cap on one page*, Accessed: 2024-01-17. [Online]. Available: https://www.tradingview.com/markets/cryptocurrencies/prices-all/.

[3] *The 7 key factors influencing cryptocurrency value*, Accessed: 2024-01-17. [Online]. Available: https://www.makeuseof.com/factors-influencing-the-cryptocurrency-value/.

[4] *Cryptocurrency volatility markets*, Accessed: 2024-01-17. [Online]. Available: https://link.springer.com/article/10.1007/s42521-021-00037-3.

[5] *Time series analysis of cryptocurrency returns and volatilities*, Accessed: 2024-01-17. [Online]. Available: https://link.springer.com/article/10.1007/s12197-020-09526-4.

[6] *Factors influencing cryptocurrency prices: Evidence from bitcoin, ethereum, dash, litcoin, and monero*, Accessed: 2024-01-17. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3125347.

[7] *What influences the price of crypto?* Accessed: 2024-01-17. [Online]. Available: https://crypto.com/university/crypto-price.

[8] M. Shen, "More institutional investors jumping into bitcoin leaves less to go around, data shows," *CoinDesk*, 2021. [Online]. Available: https://www.coindesk.com/markets/2021/03/02/more-institutional-investors-jumping-into-bitcoin-leaves-less-to-go-around-data-shows/.

[9] "Lightning network — top news and stories by cointelegraph," *Cointelegraph*, 2023. [Online]. Available: https://cointelegraph.com/tags/lightning-network.

[10] Ethereum.org. "Proof-of-stake (pos)." (2023), [Online]. Available: https://ethereum.org/en/developers/docs/consensus-mechanisms/pos/.

[11] T. Reuters, "Cryptocurrency regulations by country," 2023. [Online]. Available: https://www.thomsonreuters.com/en-us/posts/wp-content/uploads/sites/20/2022/04/Cryptos-Report-Compendium-2022.pdf.

[12] S. P. Lee, "Crypto derivatives market size 2023," *CoinGecko*, 2023. [Online]. Available: https://www.coingecko.com/research/publications/crypto-derivatives-market.

[13] A. Zamuda, V. Crescimanna, J. C. Burguillo, *et al.*, "Forecasting cryptocurrency value by sentiment analysis: An hpc-oriented survey of the state-of-the-art in the cloud era," *Springer*, 2019. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-16272-6_12.

[14] C. Kim, "Wall street takes traditional finance to crypto," *Axios*, 2023. [Online]. Available: https://www.axios.com/2023/06/22/wall-street-traditional-finance-crypto-edx.

[15] "Binance vs. huobi vs. okex: Comparing international crypto buying/selling options." (2020), [Online]. Available: https://en.cryptonomist.ch/2020/10/21/binance-huobi-okex-comparing-crypto-buying-selling-options/.

[16] "How to read candlestick charts." (), [Online]. Available: https://www.coinbase.com/en-gb/learn/tips-and-tutorials/how-to-read-candlestick-charts.

[17] D. Zhan, S. Yi, D. Xu, *et al.* "Adaptive transfer learning of multi-view time series classification." arXiv: 1910.07632 [cs.LG]. (2019).

[18] T. R. Dinger, Y.-c. Chang, R. Pavuluri, and S. Subramanian, *What is time series classification?* https://developer.ibm.com/learningpaths/get-started-time-series-classification-api/what-is-time-series-classification/, 2022.

[19] *Crypto lake*, https://crypto-lake.com/, 2024.

[20] CoinMarketCap, *Avalanche price today, avax to usd live price ... - coinmarketcap*, 2024. [Online]. Available: https://coinmarketcap.com/currencies/avalanche/.

[21] Wikipedia, *Avalanche (blockchain platform) - wikipedia*, 2024. [Online]. Available: https://en.wikipedia.org/wiki/Avalanche_%28blockchain_platform%29.

[22] N. Reiff, *What is avalanche (avax), its pros, cons, and risks? - investopedia*, 2023. [Online]. Available: https://www.investopedia.com/avalanche-avax-definition-5217374.

[23] CoinMarketCap, *Bnb price today, bnb to usd live price, marketcap and chart — coinmarketcap*, 2024. [Online]. Available: https://coinmarketcap.com/currencies/bnb/.

[24] Wikipedia, *File:bnb, native cryptocurrency for the binance smart chain.svg - wikipedia*, 2024. [Online]. Available: https://en.m.wikipedia.org/wiki/File:BNB,_native_cryptocurrency_for_the_Binance_Smart_Chain.svg.

[25] Binance, *What is bnb? - binance*, 2024. [Online]. Available: https://www.binance.com/en/bnb.

[26] CoinMarketCap, *Dogecoin price today, doge to usd live price, marketcap and chart — coinmarketcap*, 2024. [Online]. Available: https://coinmarketcap.com/currencies/dogecoin/.

[27] Wikimedia, *File:dogecoin logo.png - wikimedia*, 2024. [Online]. Available: https://upload.wikimedia.org/wikipedia/en/d/d0/Dogecoin_Logo.png.

[28] CoinMarketCap, *Polkadot price today, dot to usd live price, marketcap and chart — coinmarketcap*, 2024. [Online]. Available: https://coinmarketcap.com/currencies/polkadot-new/.

[29] W. Commons, *File:polkadot logo.png - wikimedia commons*, 2024. [Online]. Available: https://en.wikipedia.org/wiki/File:Polkadot_Logo.png.

[30] CoinMarketCap, *Neo price today, neo to usd live price, marketcap and chart — coinmarketcap*, 2024. [Online]. Available: https://coinmarketcap.com/currencies/neo/.

[31] W. Commons, *File:neo symbol.svg - wikimedia commons*, 2024. [Online]. Available: https://en.wikipedia.org/wiki/File:Neo_symbol.svg.

[32] CoinMarketCap, *Solana price today, sol to usd live price, marketcap and chart — coinmarketcap*, 2024. [Online]. Available: https://coinmarketcap.com/currencies/solana/.

[33] W. Commons, *File:solana logo.png - wikimedia commons*, 2024. [Online]. Available: https://en.wikipedia.org/wiki/File:Solana_logo.png.

[34] LCX, *Proof of history explained*, https://www.lcx.com/proof-of-history-explained/, 2024.

[35] Investopedia, *Proof of stake (pos)*, https://www.investopedia.com/terms/p/proof-stake-pos.asp, 2024.

[36] J. Paton, *Forward fill in pyspark*, https://johnpaton.net/posts/forward-fill-spark/, 2022.

[37] S.-l. developers, *Sklearn.ensemble.baggingclassifier*, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingClassifier.html, 2022.

[38] W3cubDocs, *Ensemble.baggingclassifier() - scikit-learn - w3cubdocs*, https://docs.w3cub.com/scikit_learn/modules/generated/sklearn.ensemble.baggingclassifier.html, 2022.