XGBoost for Prediction of Ethereum Short-term Returns Based on Technical Factor

Miss Wipawee Nayam

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2022

เอกซ์จีบูสต์สำหรับการทำนายผลตอบแทนระยะสั้นของอีเทอเรียมบนพื้นฐานปัจจัยเชิงเทคนิค

น.ส.วิภาวี นาแหยม

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565

Thesis Title	XGBoost for Prediction of Ethereum Short-term Returns Based on Technical Factor
By	Miss Wipawee Nayam
Field of Study	Computer Science
Thesis Advisor	Associate Professor Yachai Limpiyakorn, Ph.D.

Accepted by the FACULTY OF ENGINEERING, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

.... Dean of the FACULTY OF ENGINEERING

(Professor SUPOT TEACHAVORASINSKUN, D.Eng.)

THESIS COMMITTEE

.... Chairman

(Assistant Professor SUKREE SINTHUPINYO, Ph.D.)

.... Thesis Advisor

(Associate Professor Yachai Limpiyakorn, Ph.D.)

.... External Examiner

(Paskorn Apirukvorapinit, Ph.D.)

วิภาวี นาแหยม : เอกซ์จีบูสต์สำหรับการทำนายผลตอบแทนระยะสั้นของอีเทอเรียมบนพื้นฐานปัจจัยเชิงเทคนิค. (XGBoost for Prediction of Ethereum Short-term Returns Based on Technical Factor) อ.ที่ปรึกษาหลัก : รศ. ดร.ญาใจ ลิ่มปิยะกรณ์

ปัจจุบัน คริปโทเคอร์เรนซีได้รับความนิยมอย่างแพร่หลาย เนื่องจากเป็นการทำธุรกรรมแบบไม่รวมศูนย์ ซึ่งแตกต่างจากระบบเงินตราแบบดั้งเดิมที่ต้องผ่านตัวกลาง ได้แก่ ธนาคาร หรือรัฐบาล ข้อดีของการไม่รวมศูนย์ที่ไม่ต้องการตัวกลางคือ สามารถลดค่าธรรมเนียมการทำธุรกรรมและเวลาที่ใช้ดำเนินการ อย่างไรก็ตาม การลงทุนในคริปโทเคอร์เรนซีมีความเสี่ยงและความไม่แน่นอน อันเนื่องมาจากความผันผวนของราคาและการเปลี่ยนแปลงที่รวดเร็ว ความจริงที่ว่า การคาดการณ์ราคาของสินทรัพย์นั้นมีความซับซ้อน เนื่องจากมีปัจจัยหลายอย่างที่ส่งผลต่อความเคลื่อนไหวของราคา งานวิจัยนี้ได้ศึกษาปัจจัยทางเทคนิคเพื่อวิเคราะห์ผลตอบแทนระยะสั้นของอีเทอเรียมในช่วง 1-10 วัน ข้อมูลราคาปิดอีเทอเรียมย้อนหลังถูกรวบรวมจากคอยน์เก็กโก ตัวชี้วัดทั้งหมด 22 ตัวได้รับการเลือกเพื่อให้เข้าใจแนวโน้มตลาดอย่างลึกซึ้ง โดยแบ่งตามกลุ่มต่างๆ ประกอบด้วย ปัจจัยโมเมนตัม ปัจจัยความผันผวน และปัจจัยแนวโน้มของตลาด ค่าตัวชี้วัดต่างๆ เหล่านี้ถูกคำนวณบนพื้นฐานค่าราคาปิดอีเทอเรียมในอดีต และใช้สำหรับการเรียนรู้ของเอกซ์จีบูสต์ เพื่อค้นหาแพตเทิร์นการซื้อขายก่อนหน้า การประเมินสมรรถนะแบบจำลองใช้เมตริก AUC-ROC แบบพหุคลาส สำหรับวัดความแม่นยำในการทำนายผลตอบแทนอีเทอเรียม 3 ประเภท ได้แก่ แนวโน้มขาลง แนวราบ และขาขึ้น ผลลัพธ์การทดลองรายงานช่วงค่าเฉลี่ยไมโครกราฟอาร์โอซีระหว่าง 0.65 ถึง 0.67 นอกจากนี้ จากการศึกษาเน้นให้ความสำคัญกับการพิจารณาตัวชี้วัดโมเมนตัมเมื่อตัดสินใจลงทุนในอีเทอเรียม

สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์	ลายมือชื่อนิสิต ....
ปีการศึกษา	2565	ลายมือชื่อ อ.ที่ปรึกษาหลัก ....

6470276021 : MAJOR COMPUTER SCIENCE
KEYWORD: cryptocurrency investment, technical factor, Ethereum, XGBoost, machine learning

Wipawee Nayam : XGBoost for Prediction of Ethereum Short-term Returns Based on Technical Factor. Advisor: Assoc. Prof. Yachai Limpiyakorn, Ph.D.

Unlike traditional currencies that rely on centralized such as banks or governments, cryptocurrencies today have become popular due to its decentralized transactions. Decentralization takes advantage of no requirement for intermediaries, thus reducing transaction fees and processing time. However, investing in cryptocurrencies incurs risks and uncertainties due to price volatility and rapid changes. The fact that prediction of asset prices is complex due to the influence of multiple factors on price movements. This paper studied the technical factor to analyze the short-term returns of Ethereum in the periods of 1-10 days. The historical data containing Ethereum closing price are collected from CoinGecko. The twenty-two indicators are chosen from Momentum, Volatility, and Sentiment factors as candidates to provide valuable insights in market trends. The values of these indicators are calculated based on past Ethereum closing prices and then used for XGBoost learning to discover patterns in previous trading. The model performance is evaluated using the multi-class AUC-ROC metric, which measures the accuracy of predicting three types of Ethereum returns: Downtrend, Sideway, and Uptrend. The experimental results reported that the models achieved the values of micro-average ROC curve ranging from 0.65 to 0.67. Moreover, the study emphasizes the importance of considering momentum indicators when making investment decisions in Ethereum.

Field of Study:	Computer Science	Student's Signature	....
Academic Year:	2022	Advisor's Signature	....

ACKNOWLEDGEMENTS

The completion of this thesis could be possible with the encouragement and invaluable from my advisor, Assoc. Prof. Yachai Limpiyakorn for their encouragement and invaluable guidance throughout my thesis. Their support has made a significant difference in my progress and helped me reach this point.

I am also grateful to Asst. Prof. Dr. Sukree Sinthupinyo and Dr. Paskorn Apirukvorapinit for their time and excellent advice during the presentations. Their feedback and insights have been incredibly helpful in improving my work.

Lastly, I want to express my appreciation to my family and friends for support and belief in my motivated.I am truly grateful to everyone mentioned above for their contributions and support, which have been the successful completion of my thesis.

Wipawee Nayam

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

Introduction

1.1 Statement of problems

The advent of cryptocurrencies in recent years has presented a significant phenomenon to traditional currency and payment systems in the financial market. Bitcoin (BTC) as the first cryptocurrency invented in 2009, marked the beginning of a new era in decentralized finance. It gained widespread acceptance and laid the foundation for the development of other cryptocurrencies such as Ethereum (ETH), which has emerged a prominent player, second to Bitcoin [1]. Unlike traditional currencies that rely on centralized such as banks or governments, cryptocurrencies, which are secured by blockchain technology, have become popular due to its decentralized transactions. The decentralized software platform enables the execution of smart contracts, which are self-executing agreements with predefined terms and conditions. However, investment in digital currencies involves risks and uncertainties, and these assets have high volatility, like traditional investment. The volatile of cryptocurrency markets requires necessary analysis, understanding of market trends, and evaluation of potential risks before committing any funds. Forecasting the prices of assets is complicated due to various factors affecting the prices. Investors must be careful and thoughtful when investing and trading.

This research utilizes the Technical factor in analysis. The Technical factor can be divided into sub-categories consisting of Momentum, Volatility, and Sentiment. These factors provide information about market trends. A set of indicators belonging to these factors can be calculated based on historical close prices to help predict prices in the future. A highly efficient gradient-boosted decision tree, XGBoost, is selected to discover patterns in past trading activity. The study focuses on ETH short-term returns in the period of 1-10 days based on the historical data of Ethereum close price from CoinGecko [2]. The model performance was evaluated using the multi-class AUC-ROC metric, which measures accuracy of predicting the three classes of ETH returns: Downtrend, Sideway and Uptrend.

1.2 Objective

To study and develop models with artificial intelligence technology to predict the short-term returns of Ethereum in the period of 1-10 days based on technical factors.

1.3 Scope of Study

1. Use the historical data of the end-of-day ETH close prices on August 7, 2015 to January 31, 2023 from CoinGecko website.
2. Develop a model for predicting the short-term returns of Ethereum with the XGBoost algorithm using the Python language.
3. Use the technical factors consisting of 22 indicators.

1.4 Research Methodology

1. Study theory and related research of work.
2. Prepare the dataset for research.
3. Design and model development
4. Evaluation of research results
5. Publish of academic works
6. Summarize and compile the thesis.

1.5 Contribution

Develop models to predict and compare the short-term returns of Ethereum to assist in making investment decisions for cryptocurrency trading.

1.6 Publication

1. W. Nayam and Y.Limpiyakorn, "XGBoost for Classifying Ethereum Short-term Return Based on Technical Factor", 9th International Conference on Computer Technology Applications (ICCTA 2023), May 10-12, 2023, Vienna, Austria

2. W. Nayam and Y.Limpiyakorn, "Prediction of Ethereum Short-term Returns Using XGboost Model", International Journal of Emerging Technology and Advanced Engineering, 13 (8), August 2023.

CHAPTER 2
Literature Review

2.1 Related Theory

2.1.1 Ethereum

Ethereum was conceived in 2013 by Vitalik Buterin, a Russian-Canadian programmer [3]. Ethereum is a global and decentralized platform that enables various applications and financial transactions. It operates on a blockchain network, which is a distributed and decentralized public ledger for verifying and recording transactions. The blockchain network is distributed among all participants, ensuring that everyone holds an identical copy of the ledger and can access past transactions. It is not controlled by any centralized entity but is collectively managed by the distributed ledger holders. Transactions on the Ethereum blockchain are secured through cryptography to maintain network integrity and transaction verification. Ether, the native cryptocurrency of Ethereum, can be used for purchasing goods and services, like Bitcoin. However, what sets Ethereum apart is its capability for users to develop and run applications on the blockchain, much like software running on a computer. These applications can store and transfer personal data and handle complex financial transactions. Ethereum serves as the underlying blockchain network where Ether is stored and exchanged. In addition to its primary function as a currency, Ethereum offers a range of other functionalities [4]. Ether is designed to extend the functionality of cryptocurrencies beyond just sending and receiving value. It offers a range of unique features, including the ability to create decentralized applications (dapps). These dapps are created using smart contracts, which are self-executing contracts with the terms of the agreement written into code. This allows for automation of processes and reduces the need for intermediaries [5].

Smart contracts are revolutionizing how traditional contracts work. A smart contract is a simple computer program that facilitates the exchange of any asset between two parties, for example, money, shares, property, or any other digital asset. Anyone on the Ethereum network can create these contracts. The contract consists primarily of the terms and conditions mutually agreed on between the parties (peers). Once the smart contract has been executed, it cannot be altered, and any transaction done on top of a smart contract is registered permanently. Even if the smart contract is modified in the future, the transactions correlated with the original contract will not get altered. The verification process for the smart contracts is carried out by anonymous parties in the network without the need for a centralized authority, and that is what makes any smart contract execution on Ethereum a decentralized execution. The transfer of any asset or currency is done in a transparent and trustworthy manner. The identities of the two entities are secure on the Ethereum network. Once the transaction is successfully committed, the accounts of the sender and receiver are updated accordingly. In this way, it generates trust between the parties [6]. Figure 1 illustrates an example of Smart Contracts Work for buying a house on Ethereum.

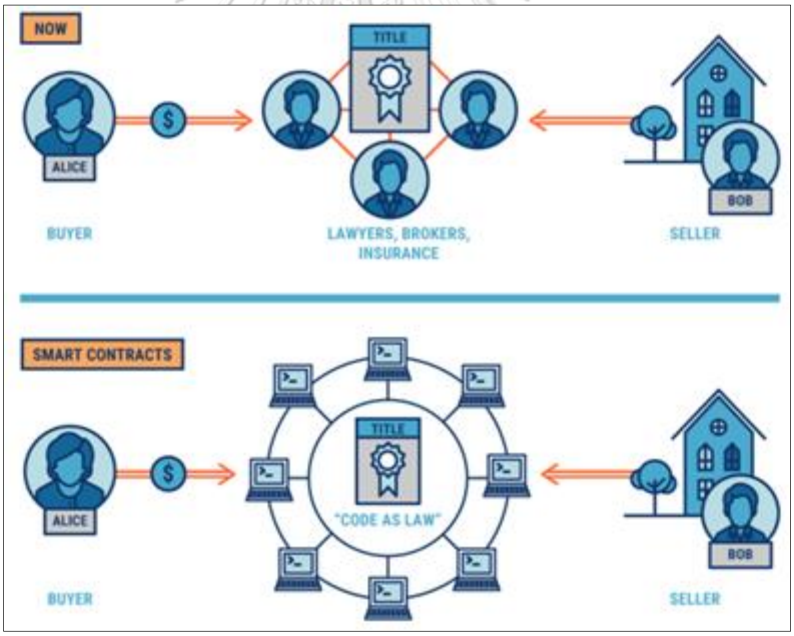

Figure 1 Smart contracts work for buying a house on Ethereum [7]

In August 2014, Ethereum launched its native token, Ether, through an initial coin offering (ICO) that sold 50 million ETHs at a price of $0.31 per coin, raising over $16 million for the project. Unlike many other cryptocurrencies, Ethereum cryptocurrency has an unbounded supply, and its annual inflation rate is about 4.5%.

Since its launch, Ether's price remained rangebound between \$0.70 and \$21 until May 2017, when it went above \$100 for the first time. From there, Ether skyrocketed to a peak of \$414 in June 2017. By January 2018, ETH's price peaked at \$1,418 before falling sharply. Between February and May 2021, ETH's price more than tripled to set a new all-time high of \$4,379. Overall, Ether has a unique supply mechanism and has seen significant price fluctuations throughout its history [8].

2.1.2 Technical Factor

Technical factor is a key component of technical analysis which is a method of analyzing statistical trends and past trading activity by means of mathematical calculations based on past price and volume data that are used to generate signals about the future direction of a security's price [9]. Technical factors can be divided into sub-factors including momentum, volatility, and sentiment. The momentum factor contains a set of indicators such as Moving Average (MA), Relative Strength Index (RSI), and Rate of Change (ROC). Example of volatility indicators is Average True Range (ATR), while Rate of Change of BTC price (BTCROC) is an example of sentiment indicators.

Moving Average (MA) is used to identify trends in the market over a given period. The equation of Moving Average is shown in equation (1). Each Moving Average is calculated by taking into consideration a window of previous values. This window can be of any length, but common values are 100, 200, and 250 for the Simple Moving Average (SMA) and 5, 10, and 20 for the Exponential Moving Average (EMA). These values help smooth out pricing trends, in addition to be used to help identify price reversals and emerging trends.

$$MA_k = \frac{\sum_{i=n-k+1}^{n} p_i}{k} \qquad (1)$$

Figure 2 demonstrates how to calculate a 5-day Simple Moving Average. The 5-Day SMA for Day 6 is calculated by adding the price history of the previous 5 days and dividing by 5: (\$10.25 + \$12.75 + \$13.17 + \$10.88 + \$14.63) resulting in an SMA of \$12.34 which is lower than the current day price of \$13.92, indicating an uptrend in price.

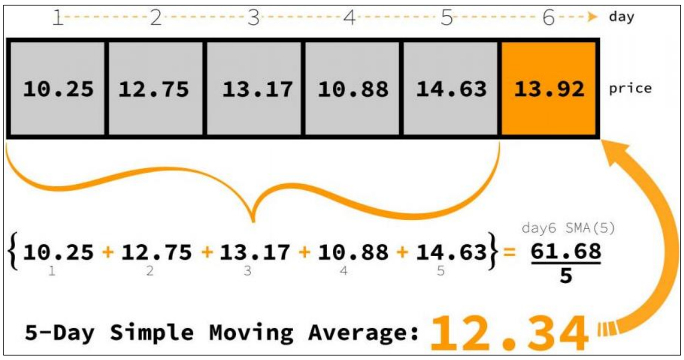

Figure 2 Example of calculating 5-day Simple Moving Average [10]

Figure 3 visualizes the 10-day and 20-day periods of Moving Average. Here, Simple Moving Average (SMA) is chosen to smooth out the data prices over a specific period. The closing prices of the last ten days are added up and divided by ten. This gives the average price for that period. By connecting these average values, it gets a line on the chart, often shown in green. It is found that the 10-day Moving Average congruous with the prices, but it is smoother and slightly delayed. When adding a 20-day Moving Average to the chart, it becomes even smoother and lags further behind the price. This is because the 20-day Moving Average considers a longer time, i.e., the last 20 days instead of just the last ten. Because of this, technical analysts say that the 20-day Moving Average is slower than the 10-day Moving Average, that is, it provides a smoother line but reacts more slowly to price changes [11].

Figure 3 Graph of 10-day and 20-day periods of Moving Average [11]

Relative Strength Index (RSI) is used to measure market momentum and determine if an asset or cryptocurrency is overbought or oversold. Typically, it is represented with a line chart ranging from 0 to 100, where the values below 30 indicate oversold conditions, and the values above 70 indicate overbought conditions. However, these values should not be treated as absolutes, but interpreted with caution. The RSI provides information on the rate of change of prices for an asset, but it can give false signals during volatile market conditions. Therefore, it should be used carefully and in conjunction with other technical indicators and fundamental analysis [12]. Figure 4 shows the graph indicating oversold and overbought of RSI indicator.

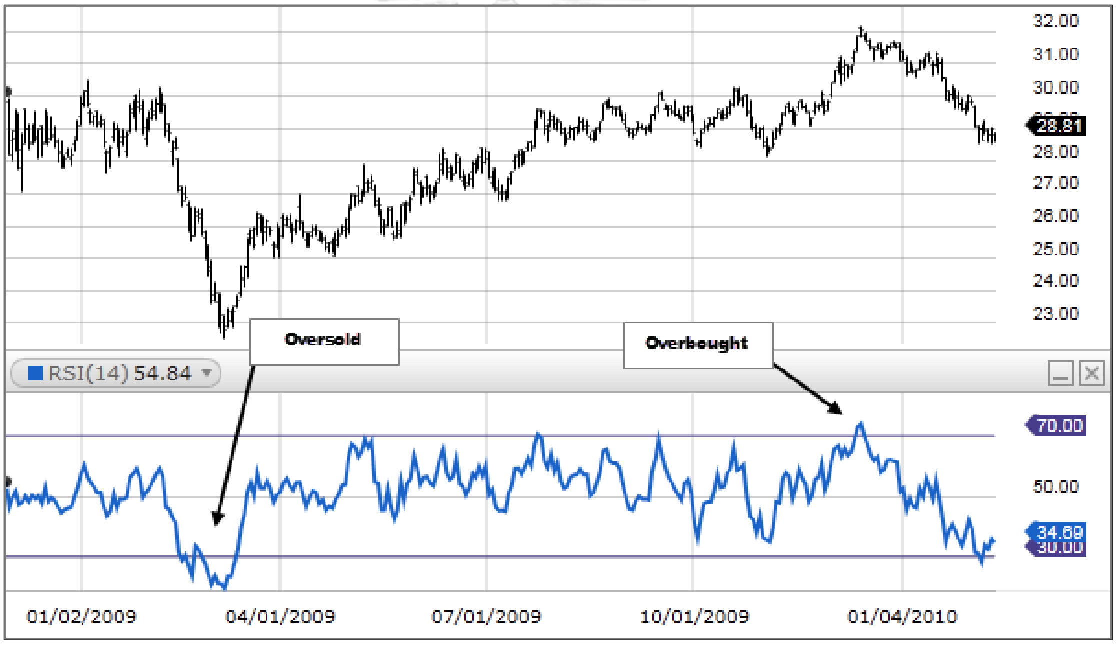

Figure 4 Graph indicates oversold and overbought of RSI indicator [13]

Rate of Change (ROC) is used in trading to measure the percentage change between the current price and a previous price from a specific number of periods ago. Its primary purpose is to identify potential trading opportunities by observing divergences and shifts in positive and negative momentum [14]. The equation of Rate of Change is shown in (2).

$$ROC = \left(\frac{Close\ Price[p]}{Close\ Price\ [p-n]} - 1\right) * 100 \qquad (2)$$

Average True Range (ATR) is a measurement of price volatility. The range is directly proportional to volatility, and that range is high and low of a stock for a given period, be it intraday, daily, weekly, or monthly was indicative of a trend. If the volatility of a stock increased, it was entering a trend, and if it slowed down, it suggested a reversal [15]. The concept called *True Range (TR)*, which is defined as the greatest of the following:

- Method 1: Current High less the current Low
- Method 2: Current High less the previous Close (absolute value)
- Method 3: Current Low less the previous Close (absolute value)

Absolute values are used to ensure positive numbers. After all, Wilder was interested in measuring the distance between two points, not the direction. If the current period's high is above the prior period's high and the low is below the prior period's low, then the current period's high-low range will be used as the True Range. This is an outside day that would use Method 1 to calculate the TR. Figure 5 shows the formula calculating of True Average.

$$TR = \mathrm{Max}[(H - L), \mathrm{Abs}(H - C_P), \mathrm{Abs}(L - C_P)]$$

$$ATR = \left(\frac{1}{n}\right) \sum_{(i=1)}^{(n)} TR_i$$

where:

TR_i = A particular true range

n = The time period employed

Figure 5 Formula for calculating Average True Range (ATR) [16]

Example A: A small high/low range formed after a gap up. The TR equals the absolute value of the difference between the current high and the previous close.

Example B: A small high/low range formed after a gap down. The TR equals the absolute value of the difference between the current low and the previous close.

Example C: Even though the current close is within the previous high/low range, the current high/low range is quite small. In fact, it is smaller than the absolute value of the difference between the current high and the previous close, which is used to value the TR. Figure 6 visualizes the three examples of concept of True Rage (TR) described above, and Figure 7 shows the graph Average True Range indicator [17].

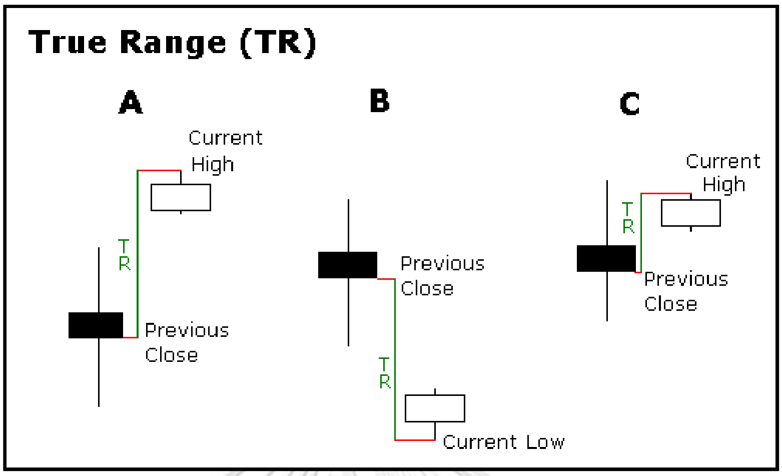

Figure 6 Concept of Average True Range [17]

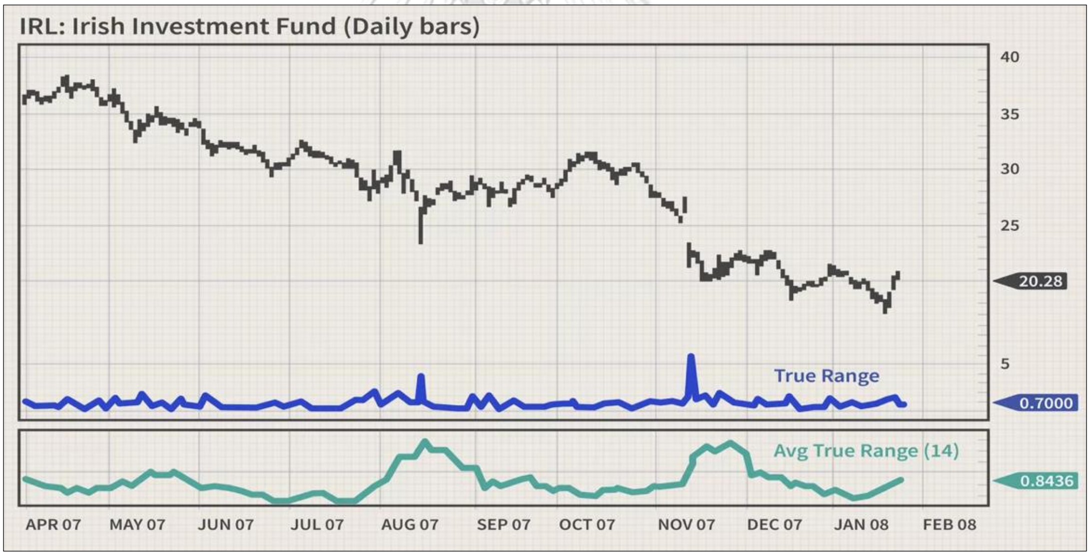

Figure 7 Graph of Average True Range indicator [18]

Rate of Change of BTC price (BTCROC) is the rate of change of the Bitcoin price. It refers the price of BTC changing over a given period, measuring the speed or velocity at which the price is increasing or decreasing. A positive rate of change indicates that the price is increasing, while a negative rate of change indicates that the price is decreasing.

2.1.3 XGBoost

In 2016, Chen et al. [19] developed XGBoost (eXtreme Gradient Boosting), an open-source software library, for building classification and regression models. XGBoost utilizes a tree ensemble learning technique, which involves the iterative addition of decision trees to the ensemble model. As illustrated in Figure 8, the algorithm calculates the loss function based on the predictions made by the existing ensemble model and the target values of the training data. This loss function serves as the basis for training a new decision tree that aims to reduce the overall error of the model. The algorithm continues to iterate until it meets the stopping conditions, such as the maximum number of rounds or the minimum improvement in performance. XGBoost has become one of the most popular machine learning algorithms due to its high performance and scalability and provides a feature importance ranking that help to understand the patterns and relationships in the data by showing which factors have the biggest impact on the model's predictions. XGBoost has been used in many different areas such as in the field of finance, it has been used for credit risk modeling, stock market prediction, and fraud detection.

Figure 8 Visualization of XGBoost algorithm [20]

2.2 Related Research

2.2.1 Prediction bitcoin returns using high-dimensional technical indicators

In 2018, Huang et al. [21] studied technical indicators to forecast the returns of Bitcoin investment. The study investigates the relationship between various technical indicators and the future performance of Bitcoin, aiming to develop a predictive model. The research focused on a predictive model using a tree-based classification algorithm. The model was trained on price data spanning a five-year period from 2012 to 2017. The objective of the study was to predict the next day return of Bitcoin which was divided ranges of returns. Ten groups represented positive changes in Bitcoin's value, ten groups represented negative changes, and one group represented no change at all. By categorizing the returns into these ranges, the researchers aimed to capture the various levels and directions of Bitcoin's performance. And the predictive model consists of 124 technical indicators which are derived from analyzing historical Bitcoin price data. These indicators were grouped into five categories, including overlap study indicators, cycle indicators, momentum indicators, volatility indicators, and pattern recognition indicators. Each category represented a different aspect of market analysis and provided unique insights into Bitcoin's price dynamics.

2.2.2 Short term return prediction of cryptocurrency based on XGBoost algorithm

In 2022, Wu et al. [22] studied predicting the short-term return of digital currencies using various machine learning algorithms. The researchers collected and pre-processed data from 14 different types of digital currencies obtained from the Kaggle platform. To evaluate the performance of the models, the researchers compared the predicted results with the actual values using the Person correlation coefficient. The Gradient Boosting, SVM, Linear Regression, and XGBoost algorithms achieved correlation coefficient values of 0.0312, 0.0301, 0.0245, and 0.0351. Among these models, The XGBoost algorithm is the best in predicting in predicting short-term returns of digital currencies. It outperformed the other models, achieving performance improvements of 12.5% compared to Gradient Boosting, 16.6% compared to SVM, and 43.4% compared to Linear Regression.

2.2.3 Predicting Trends of Bitcoin Prices Based on Machine learning Methods

In 2020, Li et al. [23] studied predicting the price trends of Bitcoin using machine learning techniques. The researchers collected daily historical price data from 2015 to 2019 and employed various technical indicators such as CCI, AROON, MA, PSA, among others. To predict the price trends and utilize machine learning models, specifically Ridge Regression and XGBoost and evaluated the performance of these models using metrics are Accuracy, Precision, Recall, and F1 Score. The research findings indicate that the XGBoost classifier outperforms the logistic regression models, demonstrating higher scores in terms of performance. Specifically, the XGBoost method achieved a positive return of 163% and a maximum retracement of 32% during the evaluation period. And The results of this research offer insights for making better buying and selling decisions in the Bitcoin market, potentially leading to higher positive returns for investors.

CHAPTER 3

Research Methodology

An individual predictive model, XGBoost, is constructed following the major steps as illustrated in Figure 9.

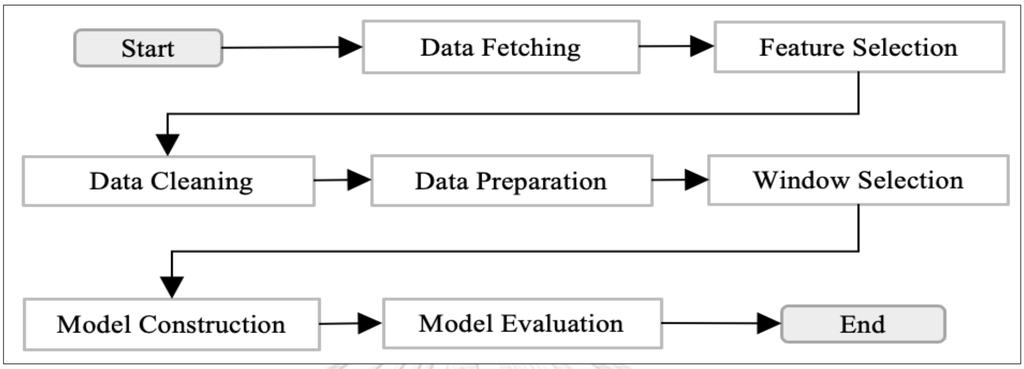

Figure 9 Flow of model construction

3.1 Data Fetching

The data were collected from CoinGecko [2], a website that provides access to cryptocurrency data. The historical data of end-of-day ETH close prices containing 2,734 records which cover the period from August 7, 2015 to January 31, 2023. Figure 10 illustrates the excerpt of end-of-day close prices of Ether.

Date	Market Cap	Volume	Open	Close
2023-01-31	$188,787,565,951	$10,977,781,768	$1,568.65	$1,586.54
2023-01-30	$198,339,307,428	$10,037,370,816	$1,646.52	$1,568.65
2023-01-29	$189,301,295,262	$6,523,949,725	$1,573.06	$1,646.52
2023-01-28	$192,677,648,328	$9,593,062,923	$1,598.47	$1,573.06
2023-01-27	$193,276,131,178	$10,076,436,860	$1,602.85	$1,598.47
2023-01-26	$194,402,223,993	$11,946,053,798	$1,614.68	$1,602.85
2023-01-25	$187,270,576,773	$9,078,639,810	$1,557.06	$1,614.68

Figure 10 Excerpt of end-of-day close prices of ETH

3.2 Feature Selection

The total of 22 technical indicators belonging to momentum factor, volatility factor, and sentiment factor were selected as features that would affect ETH short-term returns. All the values of these indicators are computed and filled into the dataset.

1. *Momentum factor* provides the indicators to measure the speed and strength of price trends such as the rate of change of ETH price over a given period. The chosen indicators are listed as following:
   - Rate of Change over 1 day (ROC1)
   - Rate of Change over 3 days (ROC3)
   - Rate of Change over 5 days (ROC5)
   - Rate of Change over 20 days (ROC20)
   - Rate of Change over 60 days (ROC60)
   - Rate of Change over 250 days (ROC250)
   - Relative Strength Index (RSI)
   - 52 Weeks High Returns (H52W)
   - Above 20 days moving average (isAboveMA20)
   - Above 50 days moving average (isAboveMA50)
   - Above 200 days moving average (isAboveMA200)
   - Uptrend movement of ETH price (isETHUpTrend)

Figure 11 shows the increasing trend of ETH price indicated by Rate of Change (ROC) during July – August 2021. In July, the price rose from $2,169.40 to $2,462.40, or an increase of 13.51% within a month. The price of ETH continued rising and reached $3,440.56 or +31.28% at the end of August. A past rally in ETH is often followed by a rally in the future as well. On contrary, Figure 12 shows the decreasing trend of ETH price indicated by ROC during May – June 2022. In May, the price dropped from $2,817.49 to $1995.94, or –29.16% within one month. The price of ETH resumed falling in June and reached $1,098.91 or −44.94% at the end of June. Historical declines in ETH are often followed by future declines as well.

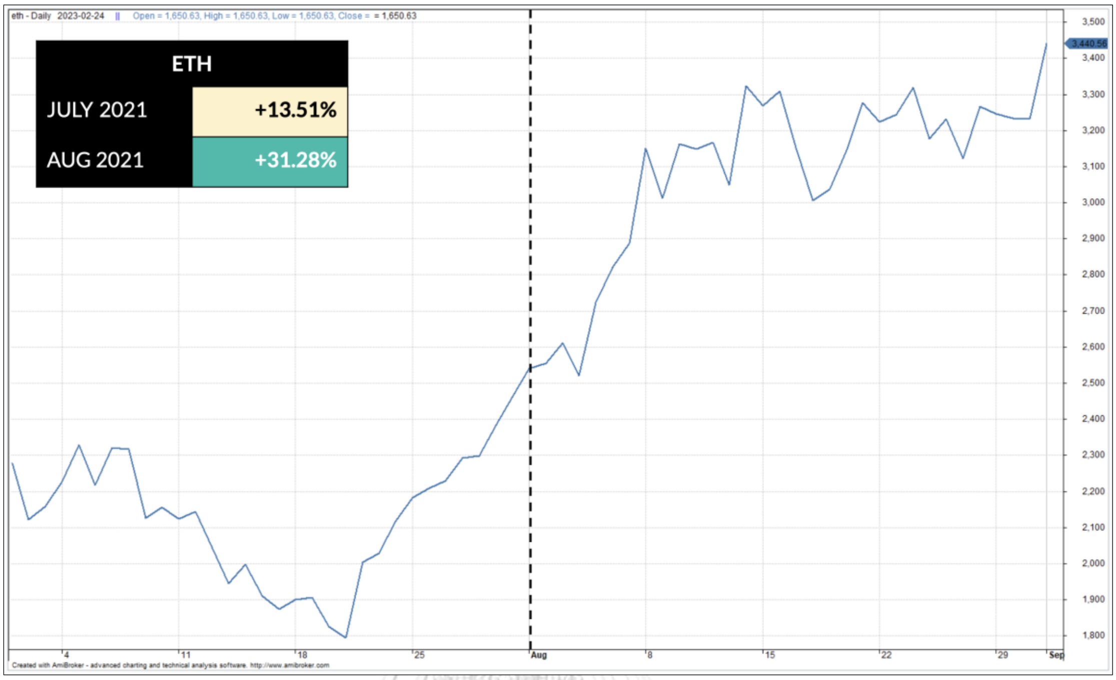

Figure 11 Trend of increasing prices of ETH during July 2021 - August 2021

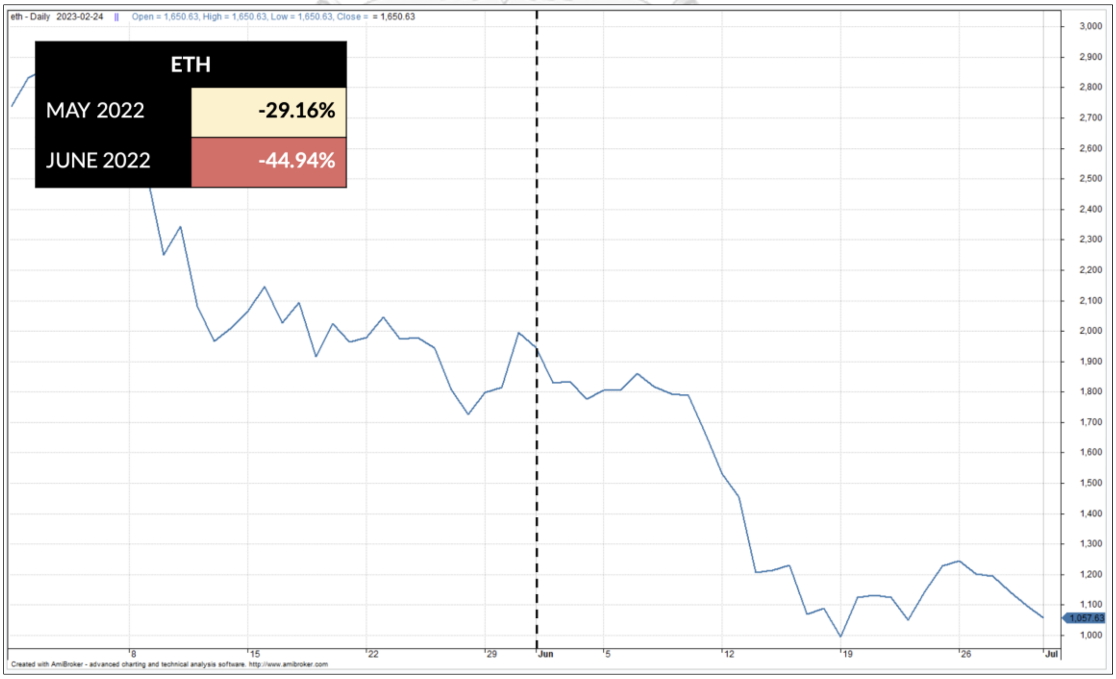

Figure 12 Trend of decreasing prices of ETH during May 2022 - June 2022

Figure 13 Prices of ETH from mid-July 2018 to mid-January 2019

Figure 13 visualizes the price of ETH from mid-July 2018 to mid-January 2019. The price of ETH has dropped from \$449.43 to \$84.47 as of 2018-12-17, or a −81.14% drop within this period. It has been around 6 months. Looking at the RSI (bottom pane), we see that while the price of ETH is falling, the RSI is also declining. The first and second time, the RSI dropped below 30, as marked in the picture, after which the price of ETH started to stall or enter a short sideway.

However, after the RSI fell below 30 last rounds, at the end of 2018, the price of ETH stopped and rise from the lowest point of \$84.47 to \$148.74, or an increase of +76.04% within a period of about one month.

Therefore, it can be noted that whenever the RSI breaks below 70 or enters the oversold zone, the price tends to enter a short-term Sideway or may be followed by a sharp rally from Investor Buyback. This is considered a return to a short-term uptrend as well.

2. *Volatility factor* provides insight into price volatility. In this work, Average True Range (ATR) is chosen to reflect the price movement range. A high ATR value indicates a highly volatile asset price or fast-moving markets where the asset’s price changes rapidly and unpredictably over periods of time.

3. *Sentiment factor* is introduced based on the assumption that the movement of the same type of assets is often related. Sentiment indicators are used to assess the overall market situation based on the movement of the cryptocurrency to determine whether it is in an uptrend or a downtrend price. Following is the list of selected indicators:
   - Correlation over 5 days (Corr5)
   - Correlation over 20 days (Corr20)
   - Correlation over 60 days (Corr60)
   - Correlation over 250 days (Corr250)
   - Rate of Change of BTC price over the past 5 days (BTCROC5)
   - Rate of Change of BTC price over the past 20 days (BTCROC20)
   - Rate of Change of BTC price over the past 60 days (BTCROC60)
   - Rate of Change of BTC price over the past 250 days (BTCROC250)
   - Uptrend movement of BTC price (isBTCUpTrend)

3.3 Data Cleaning

Before calculating any indicators, missing values are handled. In this case, the features containing a missing value of Nan or Null are cleaned by replacing with zero. Figure 14 shows example data cleaning of technical indicator.

Example 1: When computing the indicator of the rate of change over 3 days (ROC3) measures the percentage change in price compared to the previous 3 days to calculate between the current day's closing price and the closing price from 3 days ago.

Example 2: The indicator of correlation over 5 days (Corr5) that measures the relationship between the price movements of two assets over a 5-day period.

Example 3: The indicator calculates the percentage change in the price of Bitcoin over a 5-day period (BTCROC5). It is obtained by subtracting the closing price from 5 days ago from the current day's closing price. All of this makes it possible to calculate the value of the data of the previous period which was not previously available.

Date	Close price	ROC3	Corr5	BTCROC20
2015-08-07	2.83	0.00	0.00	0.00
2015-08-08	1.33	0.00	0.00	0.00
2015-08-10	0.69	0.00	0.00	0.00
2015-08-11	1.07	-62.31	0.00	0.00
2015-08-12	1.26	-5.57	0.00	0.00
2015-08-13	1.83	165.48	-0.20	0.00
2015-08-14	1.83	71.07	-0.22	0.00
2015-08-15	1.67	32.97	-0.79	0.00
2015-08-16	1.48	-19.11	-0.14	0.00
2015-08-17	1.20	-34.07	0.92	0.00
2015-08-18	1.28	-23.31	0.89	0.00
2015-08-19	1.25	-15.16	0.48	0.00
2015-08-20	1.48	23.29	-0.03	0.00
2015-08-21	1.41	9.75	-0.49	0.00
2015-08-22	1.38	10.22	-0.19	0.00
2015-08-23	1.36	-8.69	0.93	0.00
2015-08-24	1.25	-11.36	0.95	0.00
2015-08-25	1.16	-15.93	0.72	0.00
2015-08-26	1.12	-17.38	0.33	0.00
2015-08-27	1.13	-9.06	-0.01	0.00
2015-08-28	1.19	2.42	-0.58	-16.82
2015-08-29	1.17	4.89	0.64	-12.06
2015-08-30	1.32	16.64	0.44	-13.47
2015-08-31	1.35	13.83	0.44	-14.74

Figure 14 Example of data cleaning of missing values

3.4 Data Preparation

The XGBoost models were constructed to predict the class of ETH short-term returns based on the value of Return next X-day which is calculated using the equation as shown in (3). In this work, we focus on short-term returns ranging from 1 to 10 days investment. Table 1 shows example of computing percentage of return next 1-day and next 2-day. For ease of understanding, the values of Return next X-day are discretized into three categories: Downtrend (D), Sideway (S), and Uptrend (U). The distribution of Return next X-day values into classes is summarized in Table 2.

$$Return\ next\ X\ day = \left(\left(\frac{ETH\ close\ price\ next\ X\ day}{ETH\ close\ prie\ at\ day}\right)\right) * 100 \qquad (3)$$

Table 1 Example of percentage of return next 1-day and next 2-day

Date	Close price	Return next 1-day (%)	Return next 2-day (%)
2015-08-07	2.83	-53.00	-75.72
2015-08-08	1.33	-48.33	-19.79
2015-08-09	0.69	55.24	82.76
2015-08-10	1.07	17.73	71.02
2015-08-11	1.26	45.26	45.31
2015-08-12	1.83	0.03	-8.46
2015-08-13	1.83	-8.49	-19.13
2015-08-14	1.67	-11.63	-27.95
2015-08-15	1.48	-18.47	-13.22
2015-08-16	1.20	6.44	4.06
2015-08-17	1.28	-2.24	15.83
2015-08-18	1.25	18.48	12.26
2015-08-19	1.48	-5.25	-6.97
2015-08-20	1.41	-1.82	-3.64
2015-08-21	1.38	-1.85	-9.72
2015-08-22	1.36	-8.01	-14.34
2015-08-23	1.25	-6.88	-10.18
2015-08-24	1.16	-3.55	-2.34
2015-08-25	1.12	1.26	6.19

1) Downtrend (D): The boundary selected from the lowest 25% of return next X-day from all data in each range of returns, calculated by multiplying the total number of returns in this range by 0.25 contributing the bottom 25% of returns. Figure 15 visualizes the boundary of ETH Downtrend group return next 1-day.

2) Sideway (S): The boundary values denote the middle 50% of return next X-day by considering the whole dataset, computed by multiplying the total number of returns from upper limit downtrend range by 0.50 to get the number of returns contributing the middle 50%. Figure 16 visualizes the boundary of ETH Sideway group return next 1-day that range of returns from -2.5% to 2.98% inclusively.

3) Uptrend (U): The boundary value denotes the highest 25% of return next X-day selected from the highest 25% of returns from all data in each range of returns, computed by multiplying the number of returns from the upper limit sideway range by 0.25 to obtain the number of returns contributing the top 25%. Figure 17 visualizes the boundary of ETH Uptrend group return next 1-day ranging of more than 2.98%. Table 3 describes some significant percentiles of ETH return next 1-day.

Table 2 Distribution of 3-class of cumulative return next X-Day

Return next X-day	Downtrend	Sideway	Uptrend
1D	<-2.50	-2.50 to 2.98	>2.98
2D	<-3.20	-3.20 to 4.37	>4.37
3D	<-3.98	-3.98 to 5.75	>5.75
4D	<-4.73	-4.73 to 6.83	>6.83
5D	<-5.36	-5.36 to 8.05	>8.05
6D	<-5.86	-5.86 to 9.21	>9.21
7D	<-6.43	-6.43 to 10.27	>10.27
8D	<-7.13	-7.13 to 11.01	>11.01
9D	<-7.52	-7.52 to 11.98	>11.98
10D	<-8.24	-8.24 to 13.28	>13.28

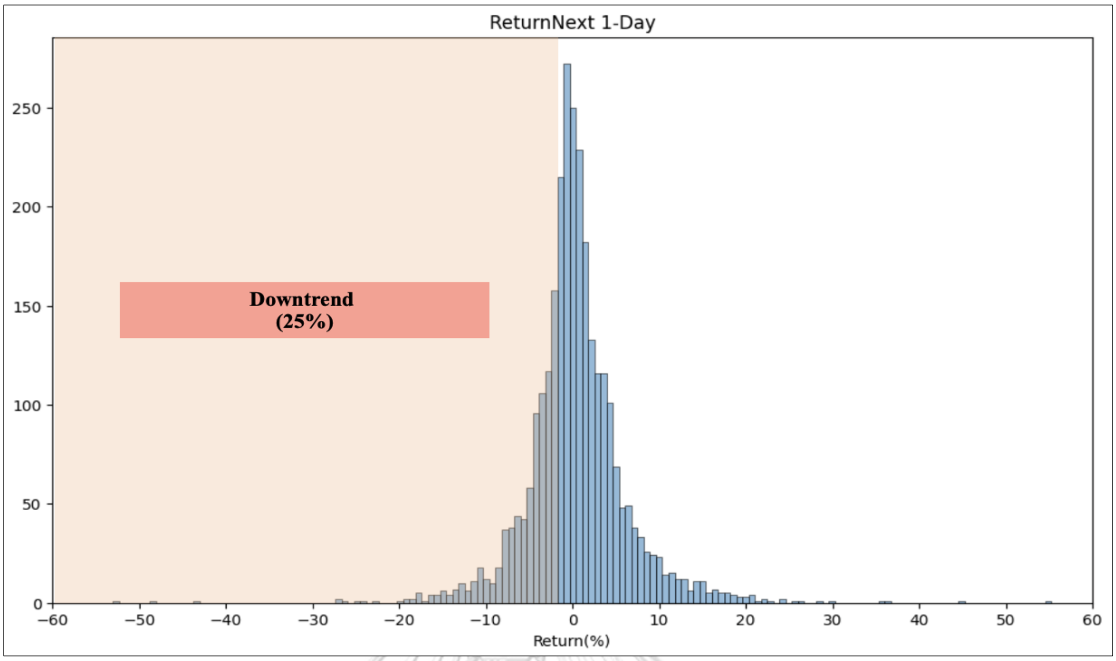

Figure 15 Histogram of boundary ETH Downtrend group return next 1-day

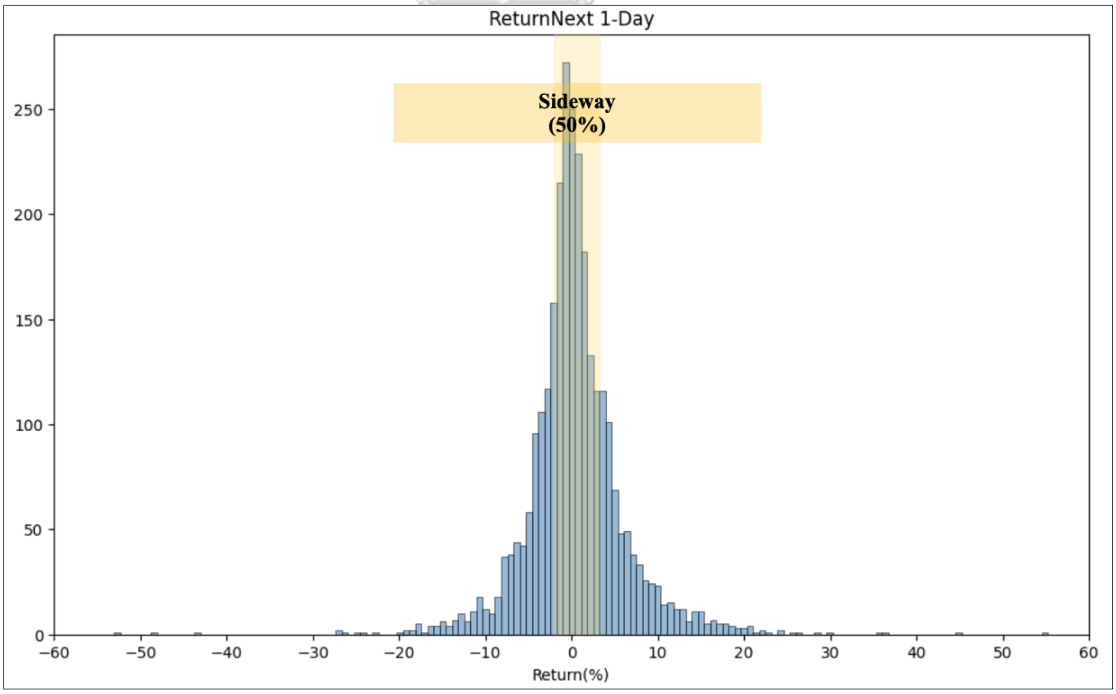

Figure 16 Histogram of boundary ETH Sideway group return next 1-day

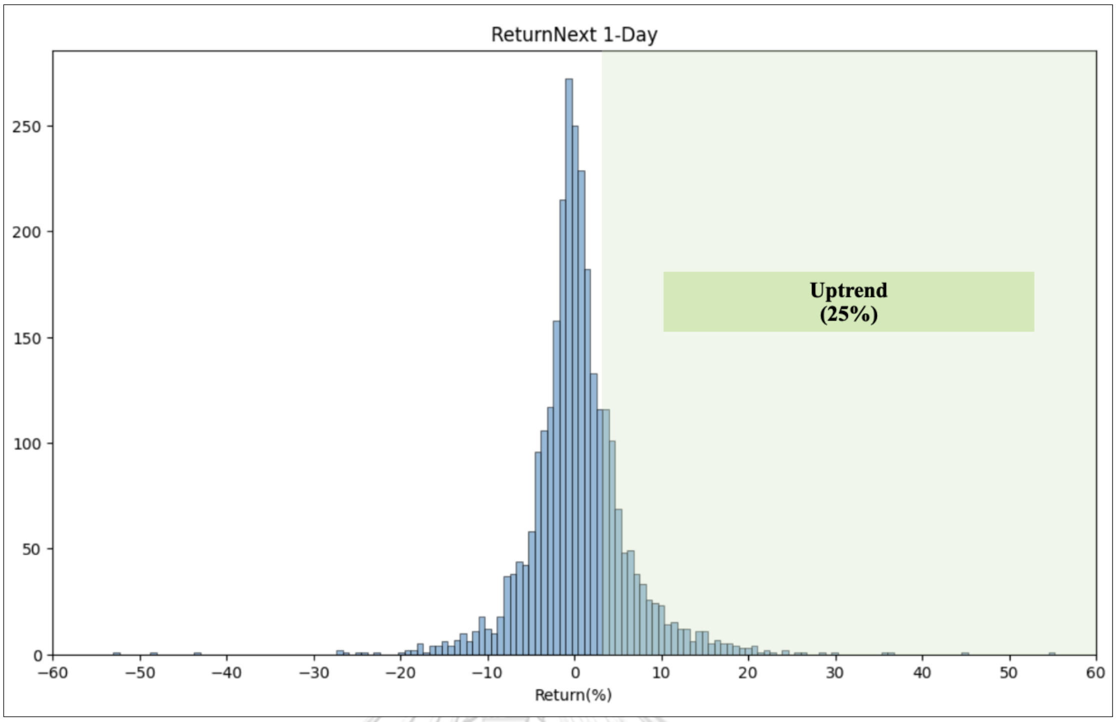

Figure 17 Histogram of boundary ETH Uptrend group return next 1-day

Table 3 Percentiles of ETH return next 1-day

Percentile	Return next 1-day (%)
5%	-8.04
10%	-5.56
25%	-2.50
50%	0.07
75%	2.98
90%	6.76
95%	10.14

Observing that the daily return rate does not signify the market trend. The thresholds of each class listed in Table 2 representing the cumulative Return next X-day. Figure 18-27 illustrate the distribution of ETH returns during the period of 1-10 days. The boundaries for each class based on the specified time periods increase because the computation of returns is cumulative over time. Observing that the longer period, the wider range of possible returns. For periods like 1 day or 2 days, the range of returns for each class is smaller. This is because short-term price movements tend to be less volatile. Therefore, the boundaries for each class are narrower. However, as the periods extend to 3 days, 4 days, and so on, the cumulative effect of price movements increases. Over longer periods, trends and patterns in price movements have more time to develop, resulting in a broader range of possible returns. In summary, the boundaries for each class based on the specified time periods increase because longer time periods capture a greater number of price movements. This leads to larger variations in returns and a broader range of possible outcomes.

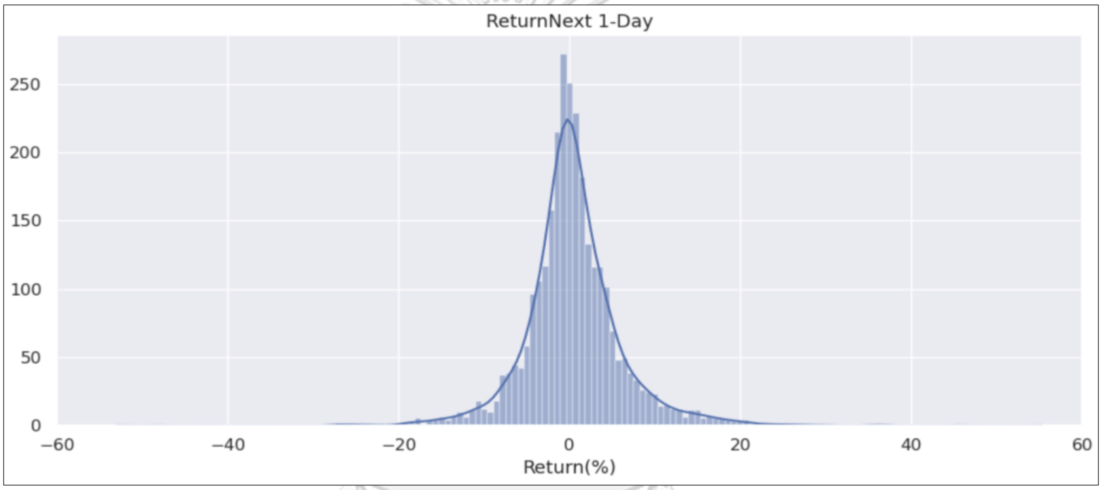

Figure 18 Visualization of distribution of ETH return next 1-day

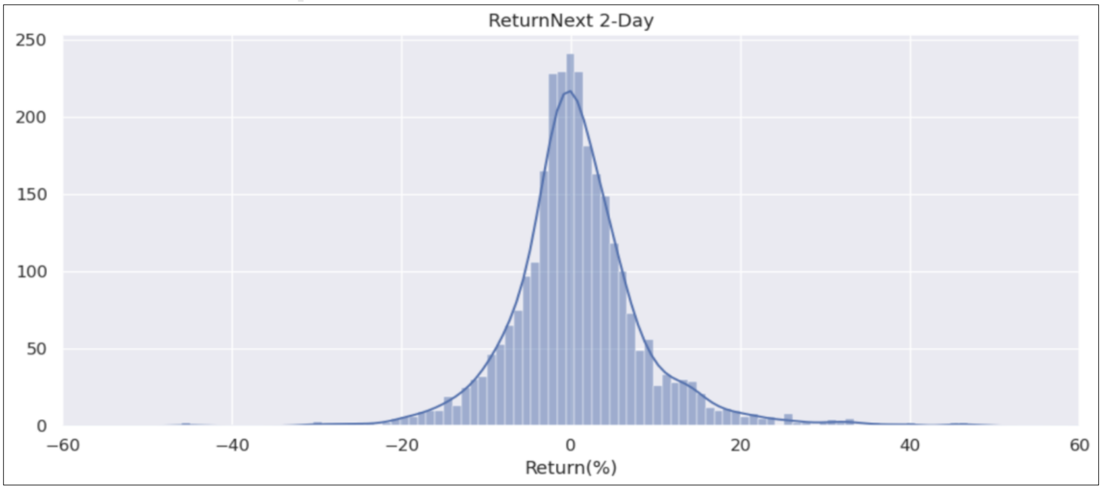

Figure 19 Visualization of distribution of ETH return next 2-day

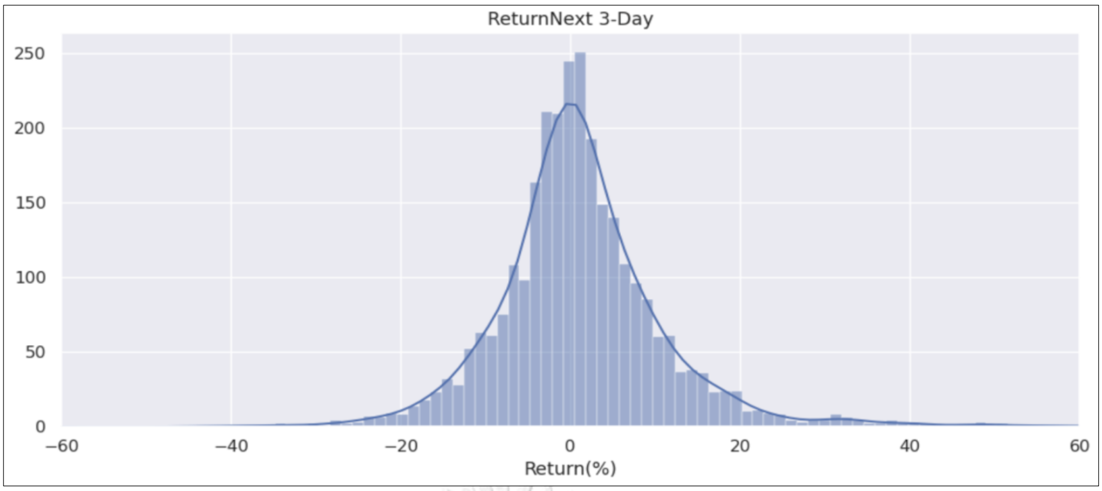

Figure 20 Visualization of distribution of ETH return next 3-day

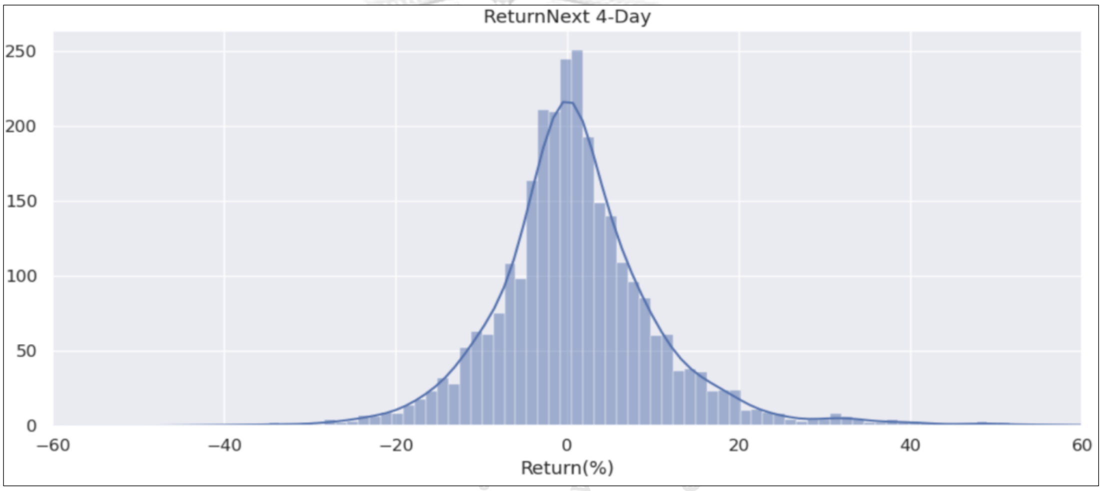

Figure 21 Visualization of distribution of ETH return next 4-day

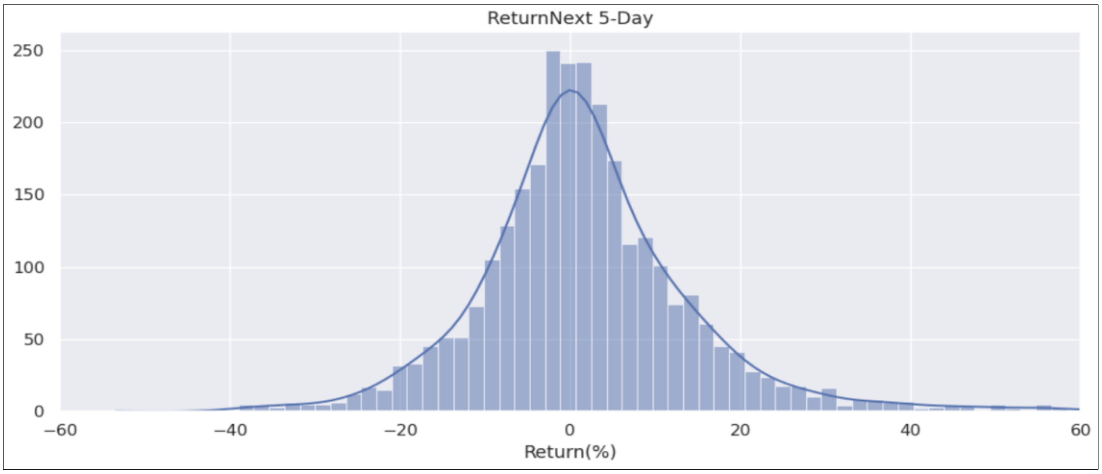

Figure 22 Visualization of distribution of ETH return next 5-day

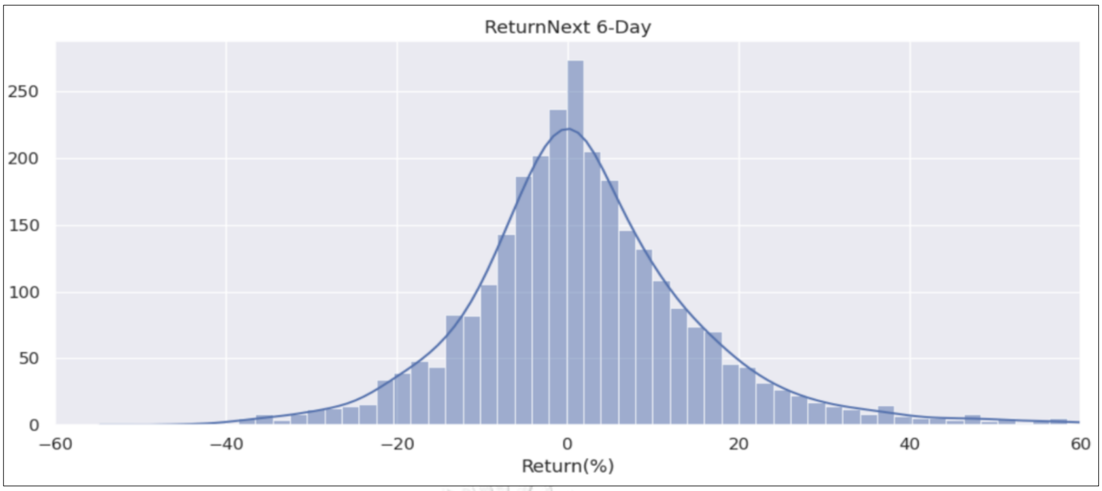

Figure 23 Visualization of distribution of ETH return next 6-day

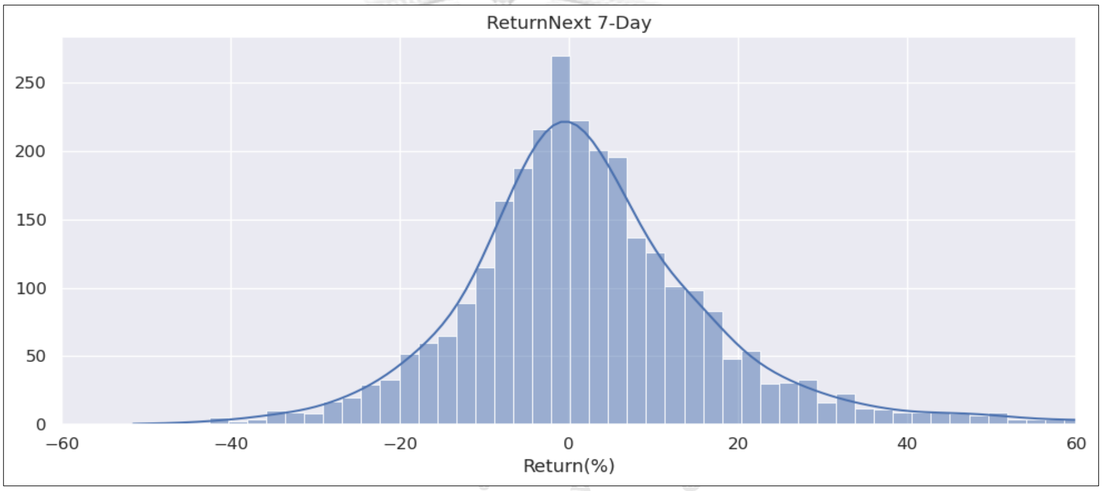

Figure 24 Visualization of distribution of ETH return next 7-day

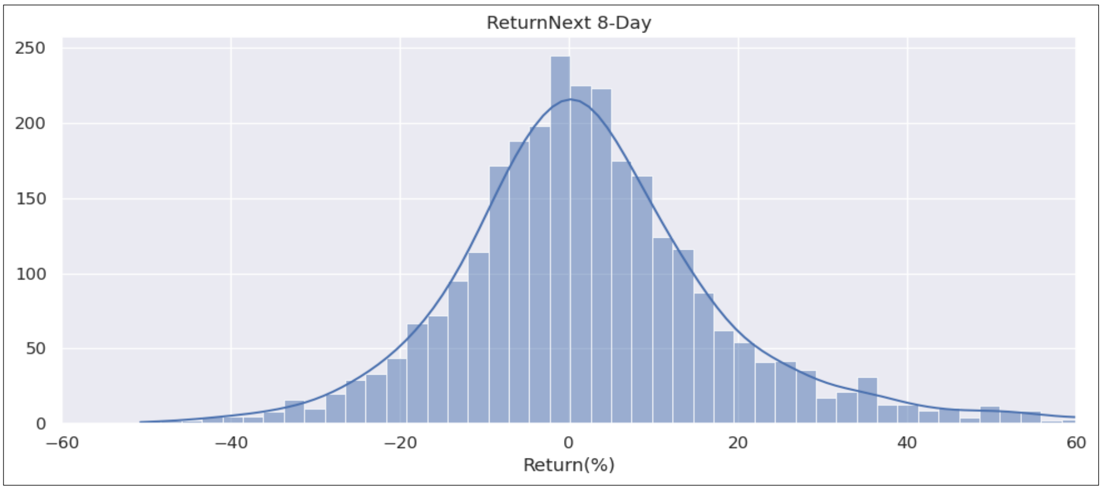

Figure 25 Visualization of distribution of ETH return next 8-day

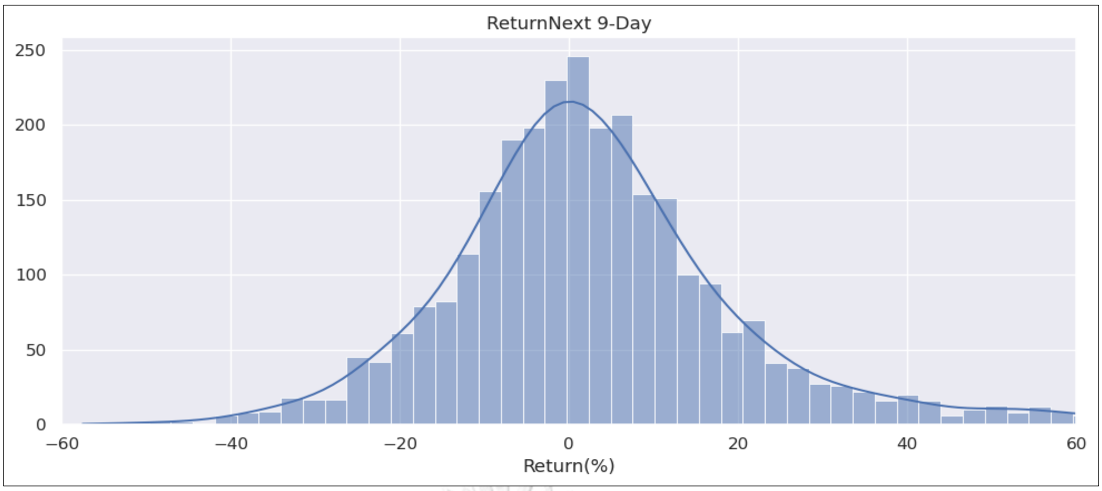

Figure 26 Visualization of distribution of ETH return next 9-day

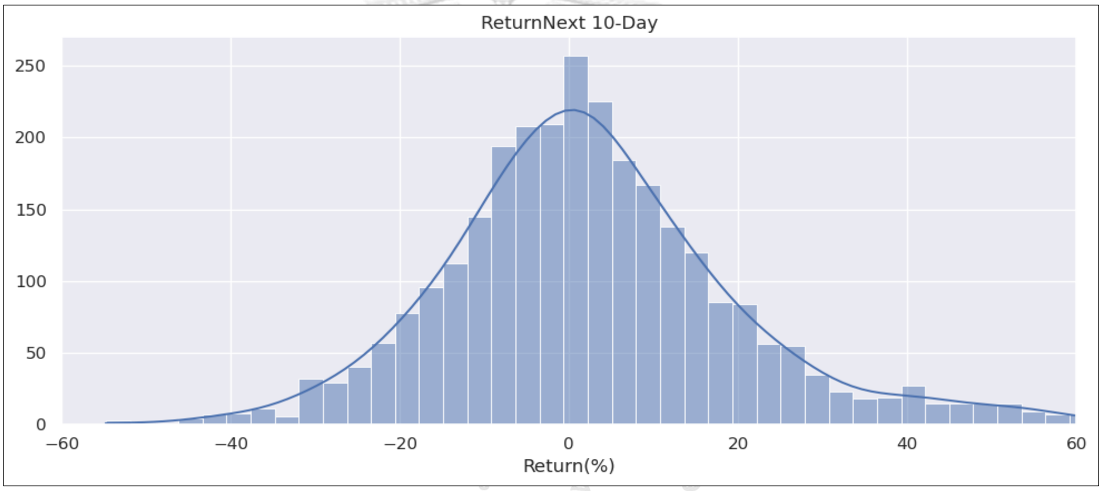

Figure 27 Visualization of distribution of ETH return next 10-day

Table 4 describes scenario 1 of how to compute Return next 10-day, all starting from Return day0 (T0) = 100.

1) Daily Return Day1 (T1) compared to T0 = 2.98%
2) Daily Return Day2 (T2) compared to T1 = 1.00%
3) Daily Return Day3 (T3) compared to T2 = -2.30%
4) Daily Return Day4 (T4) compared to T3 = 1.20%
5) Daily Return Day5 (T5) compared to T4 = 3.40%
6) Daily Return Day6 (T6) compared to T5 = 1.50%
7) Daily Return Day7 (T7) compared to T6 = 0.05%
8) Daily Return Day8 (T8) compared to T7 = 3.00%

9) Daily Return Day9 (T9) compared to T8 = -0.50%

10) Daily Return Day10 (T10) compared to T9 = 4.00%

Considering falling/ rising Daily Returns of T1 to T10 that were associated with fluctuate Return next 1-day, simply informing there were 3 days of Uptrend return and 7 days of Sideway return. While considering accumulative returns, the Return next 10-day ended at 115.09 or +15.09% become Uptrend (compared to 13.28 in Table 2), finally.

Table 4 Scenario1 of Calculation of Return next 10-Day using Cumulative Return

Day	Daily Return	Return 1D Class	Cumulative return 10D	Return next 10D
T0	0.00%		100.00	
T1	2.98%	S	102.98	
T2	1.00%	S	104.01	
T3	-2.30%	S	101.62	
T4	1.20%	S	102.84	
T5	3.40%	U	106.33	
T6	1.50%	S	107.93	
T7	0.05%	S	107.98	
T8	3.00%	U	111.22	
T9	-0.50%	S	110.67	
T10	4.00%	U	115.09	15.09%

Table 5 describes scenario 2 of how to compute Return next 10-day, all starting from Return day0 (T0) = 100. The daily return of day1 compared to T0 (T1) =+1.4% while T2 to T10 is identical +1.25% per day. Starting from Return=100 (T0), the cumulative Return next 10-day (T10) ended at 113.39 or +13.39%. Observing that the return each day was slightly positive, 1.25% denoting Sideway returns. However, when considering the cumulative Return next 10-day of 13.39% > 13.28%, it revealed the Uptrend return market.

Table 5 Scenario2 of Calculation of Return next 10-Day using Cumulative Return

Day	Daily Return	Return 1D Class	Cumulative return 10D	Return next 10D
T0	0.00%		100.00	
T1	1.40%	S	101.40	
T2	1.25%	S	102.67	
T3	1.25%	S	103.95	
T4	1.25%	S	105.25	
T5	1.25%	S	106.57	
T6	1.25%	S	107.90	
T7	1.25%	S	109.25	
T8	1.25%	S	110.61	
T9	1.25%	S	111.99	
T10	1.25%	S	113.39	13.39%

Table 6 describes scenario 3 of how to compute Return next 10-day, all starting from Return day0 (T0) = 100. The daily return of day1 compared to T0 (T1) =+2.98% while T2 to T10 was steady daily return = 0%. Observing that the Return next 10-day still correctly reported cumulative return 10D of 102.98.

3.5 Window Selection

The collected Ethereum close price data were added with the computed 22 indicators, Return next X-day, and class. Figure 28 shows parts of data created for Return next 1-day. Ten datasets were created for individual Return next X-day. Each dataset was then separated into train/ test sets using the ratio 80:20. As depicted in Figure 29, the training dataset contains 80% of the total data, approximately 2,187 records gathered from August 7, 2015 to August 21, 2021. The test dataset contains 20% of the total data, approximately 547 records gathered from August 22, 2021 to January 31, 2023.

Table 6 Scenario3 of Calculation of Return next 10-Day using Cumulative Return

Day	Daily Return	Return 1D Class	Cumulative return 10D	Return next 10D
T0	0.00%		100.00	
T1	2.98%	S	102.98	
T2	0.00%	S	102.98	
T3	0.00%	S	102.98	
T4	0.00%	S	102.98	
T5	0.00%	S	102.98	
T6	0.00%	S	102.98	
T7	0.00%	S	102.98	
T8	0.00%	S	102.98	
T9	0.00%	S	102.98	
T10	0.00%	S	102.98	2.98%

Date (2016)	Close Price	Independent variables																	Return next 1D (%)	Dependent variable
		ROC1	ROC3	ROC5	ROC20	ROC60	ROC250	RSI	H52W	....	Corr20	Corr60	Corr250	BTC ROC5	BTC ROC20	BTC ROC60	BTC ROC250	isBTC UpTrend		Return Class
04-23	8.38	6.65	-1.26	-6.51	-27.23	49.27	595.85	43.15	0.56	....	-0.58	-0.63	0.49	5.44	8.04	7.50	75.76	0.00	-4.56	D
04-24	8.00	-4.56	-1.43	-8.07	-28.53	30.34	523.96	40.60	0.54	....	-0.58	-0.63	0.49	5.40	9.65	8.68	81.73	0.00	-2.28	D
04-25	7.81	-2.28	-0.54	-7.91	-25.02	31.61	523.65	39.40	0.52	....	-0.61	-0.65	0.49	4.20	8.95	8.82	103.78	0.00	-4.97	D
04-26	7.42	-4.97	-11.37	-8.46	-30.39	25.90	400.25	36.90	0.50	....	-0.67	-0.67	0.49	3.72	10.96	9.56	99.39	0.00	4.04	U
04-27	7.72	4.04	-3.39	-1.66	-22.98	21.51	449.29	40.06	0.52	....	-0.67	-0.69	0.50	-0.60	5.63	2.93	91.45	0.00	-6.14	D
04-28	7.25	-6.14	-7.20	-13.45	-25.71	12.57	425.10	36.91	0.49	....	-0.64	-0.71	0.50	-0.54	7.39	3.83	96.11	0.00	3.11	U
04-29	7.48	3.11	0.69	-6.50	-24.48	18.44	451.65	39.36	0.50	....	-0.61	-0.72	0.50	-0.73	9.71	4.50	100.91	0.00	18.96	U
04-30	8.89	18.96	15.12	13.83	0.89	17.63	613.39	51.94	0.60	....	-0.52	-0.72	0.50	-2.73	6.36	3.83	113.19	0.00	-0.68	S
05-01	8.83	-0.68	21.82	18.96	1.89	5.17	660.88	51.45	0.59	....	-0.42	-0.73	0.50	-3.26	7.10	6.62	104.49	0.00	14.44	U
05-02	10.11	14.44	35.21	30.85	34.47	6.34	802.77	60.03	0.68	....	-0.43	-0.72	0.51	-0.07	4.30	5.79	97.32	1.00	-8.14	D
05-03	9.29	-8.14	4.41	28.07	15.84	-7.50	719.04	53.47	0.62	....	-0.45	-0.73	0.51	0.20	6.14	9.99	101.96	1.00	1.39	S
05-04	9.41	1.39	6.59	25.93	11.63	-13.91	691.84	54.31	0.63	....	-0.46	-0.73	0.51	-2.07	5.22	12.56	93.01	1.00	4.39	U
05-05	9.83	4.39	-2.78	10.50	19.75	-10.37	736.82	57.00	0.66	....	-0.47	-0.72	0.51	0.23	4.41	10.87	96.49	1.00	-5.60	D
05-06	9.28	-5.60	-0.09	5.03	8.43	-1.28	601.56	52.56	0.62	....	-0.40	-0.72	0.51	1.99	7.12	11.57	102.41	1.00	1.12	S
05-07	9.38	1.12	-0.35	-7.19	0.62	-4.28	593.17	53.30	0.63	....	-0.28	-0.72	0.52	3.32	7.29	11.50	99.61	1.00	1.58	S
05-08	9.53	1.58	-3.03	2.63	6.35	-20.40	605.46	54.39	0.64	....	-0.19	-0.71	0.52	1.99	7.18	11.39	102.18	1.00	-1.85	S
05-09	9.35	-1.85	0.82	-0.65	7.55	-16.54	631.40	52.82	0.63	....	-0.13	-0.69	0.52	3.27	5.82	10.92	101.95	1.00	-0.03	S

Figure 28 Excerpt of dataset created for return next 1-day

Figure 29 Close price of ETH from August 7,2015 to January 31, 2023

3.6 Model Construction

An individual XGBoost model was constructed using the dataset created for each Return next X-day. The model was trained in Google Colab environment connected to Python3 Google Compute Engine backend with a specific RAM 1.72 GB/12.68 GB and Disk 23.28 GB/107.72 GB. Softmax was used as an objective function for transforming the prediction to multi-class: Downtrend, Sideway, Uptrend. Various hyperparameters were adjusted during the tuning process, such as learning rate was set to 0.1, number of estimators = 5,000, and maximum tree depth = 7.

3.7 Model Evaluation

The model performance was evaluated using the Multi-class AUC-ROC (Receiver Operating Characteristic - Area Under Curve). ROC is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. The higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1. In other words, the higher the AUC, the better the model is at distinguishing between instances with the associated class.

As shown in Figure 30, the ROC curve is plotted with TPR against the FPR where TPR is on the Y-axis and FPR is on the X-axis [24].

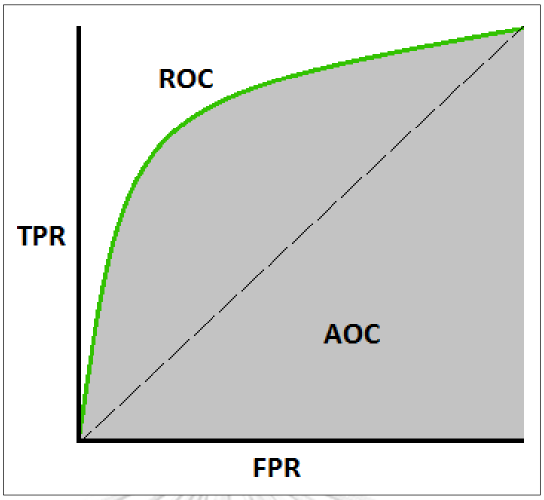

Figure 30 AUC-ROC Curve [24]

CHAPTER 4

Results and Discussion

A library, Stock Indicators for Python [25], is a tool generally used for analyzing financial data and calculating various indicators related to the stock market. The tool provides a variety of functions and methods that can be used for computing indicators such as moving averages, ROC, and relative strength index (RSI). Using test datasets, the prediction results of the ten models are compared and visualized as shown Figure 31-33, reporting the accuracy of predicting the classes of Downtrend, Sideway, and Uptrend. Figure 31 represents the fraction of test dataset that were correctly and incorrectly predicted as belonging to Downtrend class.

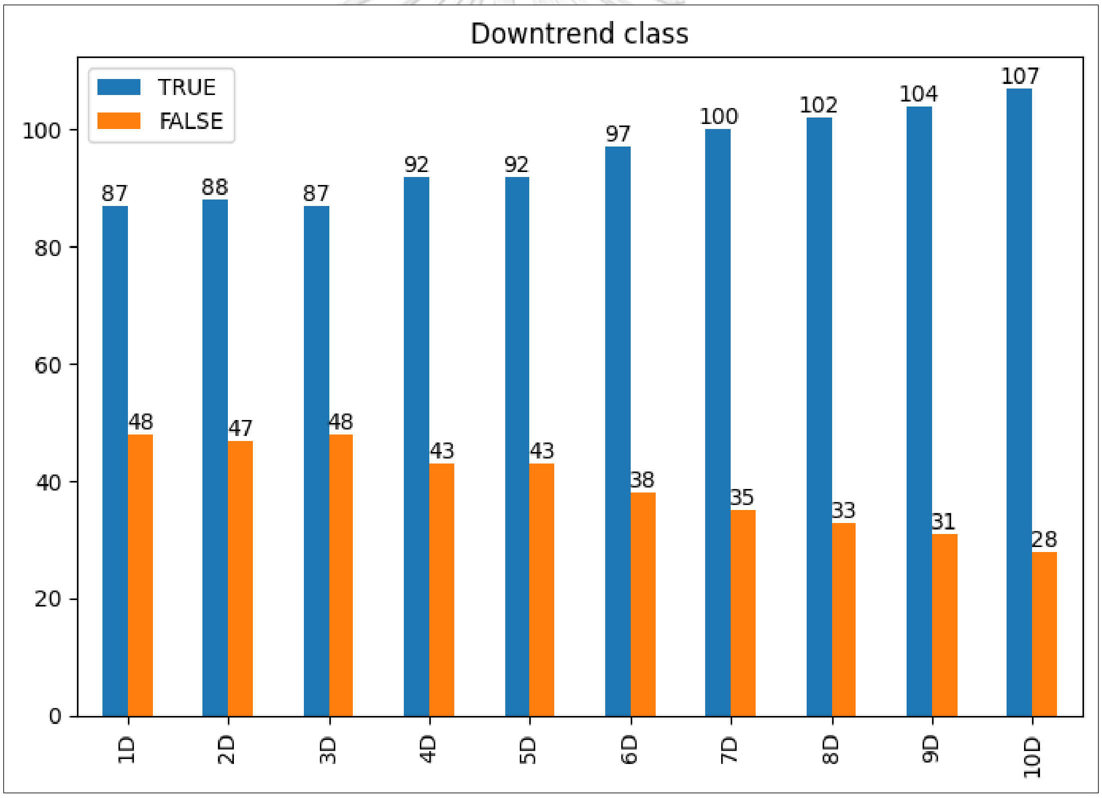

Figure 31 Comparison of Downtrend class prediction on test set

For the model performance for predicting of Downtrend class shows variation in accuracy across different prediction time from 87 to 107 records. The highest correctly predicted is achieved for Returnnext10Day.It means that the model's effectiveness improves for longer time periods in ETH price, indicating potential for predicting price trends over extended duration. And the result indicates the fraction of test dataset that were correctly and incorrectly predicted as belonging to Sideway class as shown in Figure 32.

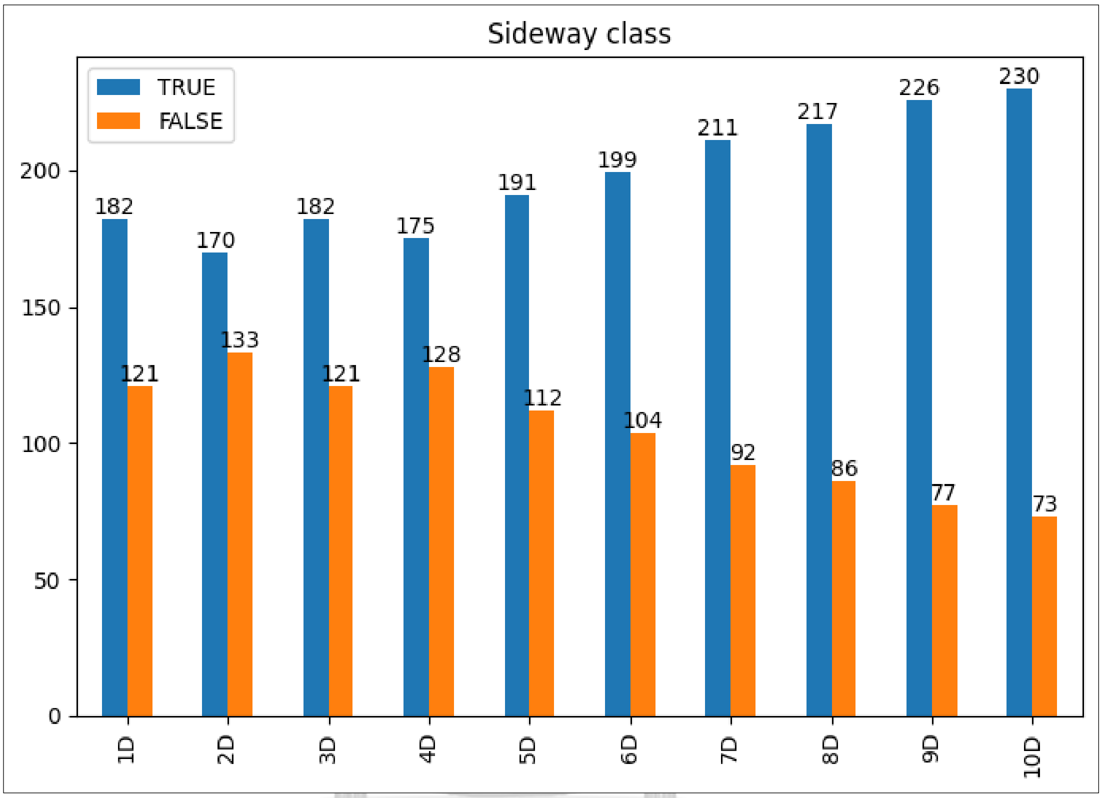

Figure 32 Comparison of Sideway class prediction on test set

For predicting Sideway class, the model performance varies across different prediction time. It achieves correctly predicted from 182 to 230 records. The correctly starts at 182 for the next day prediction and slightly volatility between 170 and 175 for the ReturnNext2D and ReturnNext4D. However, the model improves from the 5th day onwards, with correctly from 191 to 230. This suggests that the model becomes more effective in capturing longer-term price trends associated with the Sideway class.

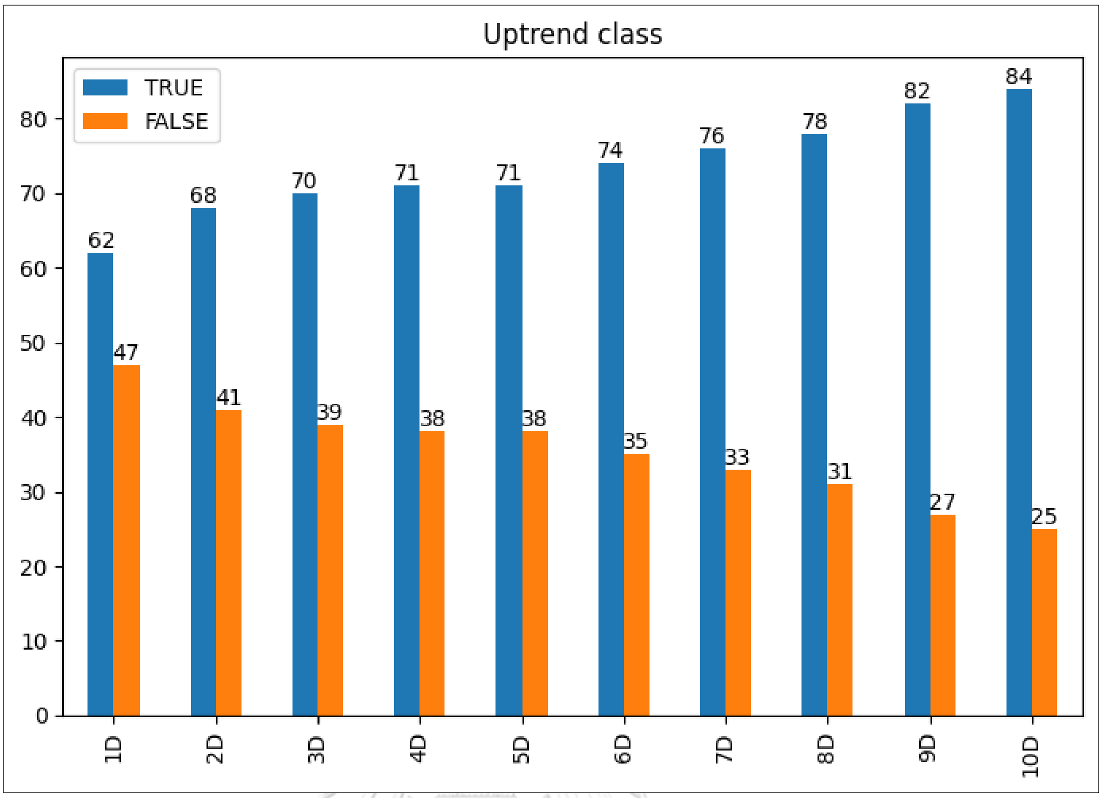

Figure 33 Comparison of Uptrend class prediction on test set

The model performance of predicting the Uptrend class (Figure 33) shows increasing correct prediction, ranging from 62 to 84.

The prediction period extends beyond five days that Returnnext6Day, Returnnext7Day, Returnnext8Day, Returnnext9Day, and Returnnext10Day: The model achieved correctly predicted of 74, 76, 78, 82, and 84 records. This indicates that the model improves in the uptrend class being correctly predicted. In summary, as extend the prediction period beyond ReturnNext5D, the model performance improves. It becomes more accurate in predicting longer-term trends for downtrend, sideway, and uptrend classes.

The findings from Figure 31, 32, 33 revealed that the greater number of next days, the more accuracy the model achieved. When extending the prediction period beyond Return next 5-day, the model performance improves. It becomes more accurate in predicting longer-term trends for Downtrend, Sideway, and Uptrend classes.

A benefit of using XGBoost model is that the algorithm provides a score that indicates how useful or valuable each feature was during the construction of the boosted decision trees. Figure 34 reports the ranking of feature importance for predicting ETH return next 1-day to next 10-day.

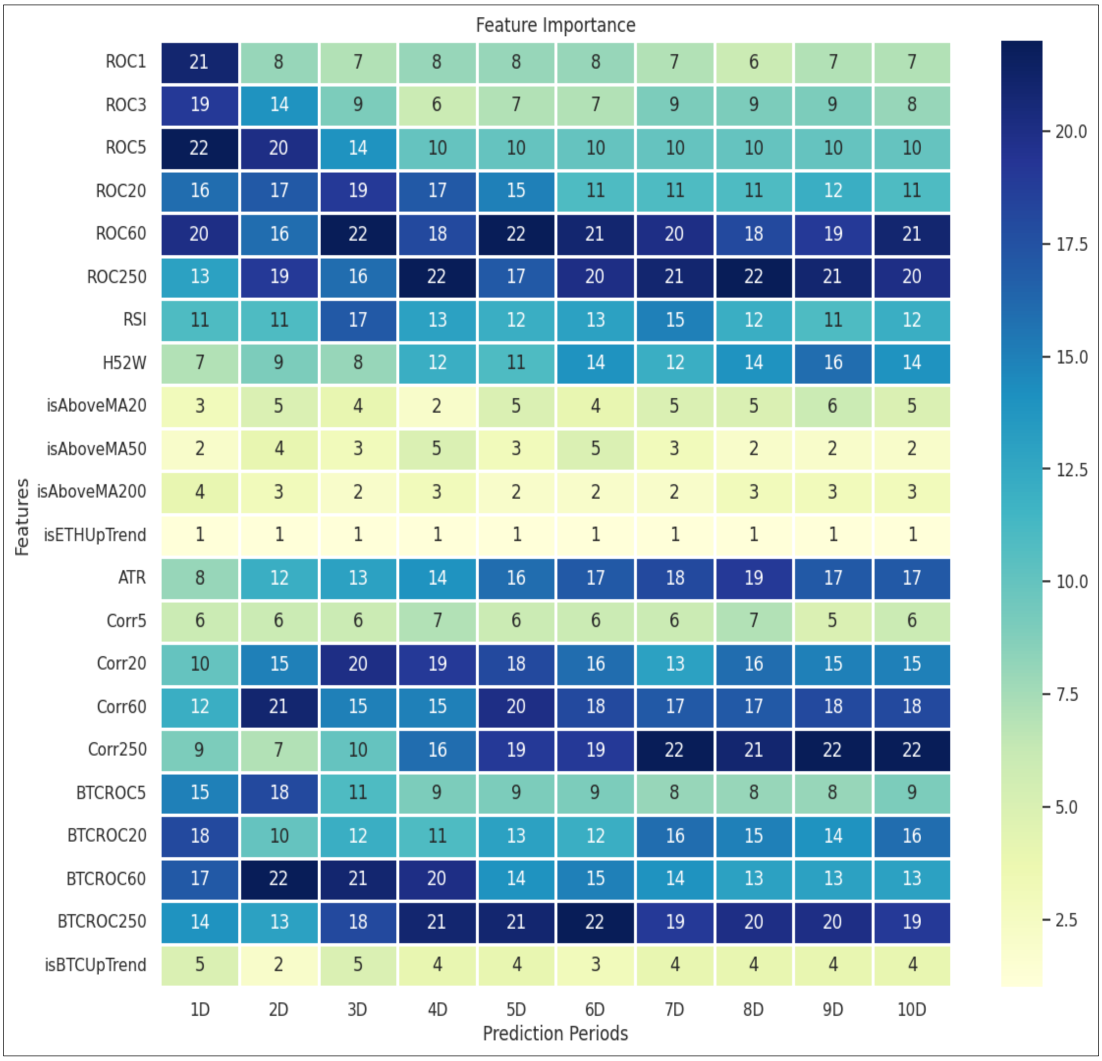

Figure 34 Overall ranks of features essential for classifying ETH returns next 1-10day

Regarding Figure 35, it is noted that the Momentum factor, which is the price trend factor, is important for predicting ETH return in the next 1-day period judged by the high rankings of the group of rates of ROC indicators. These indicators are considered as absolute momentum and used for measuring the rate of asset prices changing over specific time. This study examines various time periods, including 1 day, 3 days, 5 days, 20 days, 60 days, and 250 days. Figure 25 depicts the excerpt of ROC indicators to examine the significance of the momentum factor, specifically the rate of change (ROC), for predicting the returns of ETH in certain periods. It is found that as the return period increases, there is a greater need for long-term momentum factors useful for prediction.

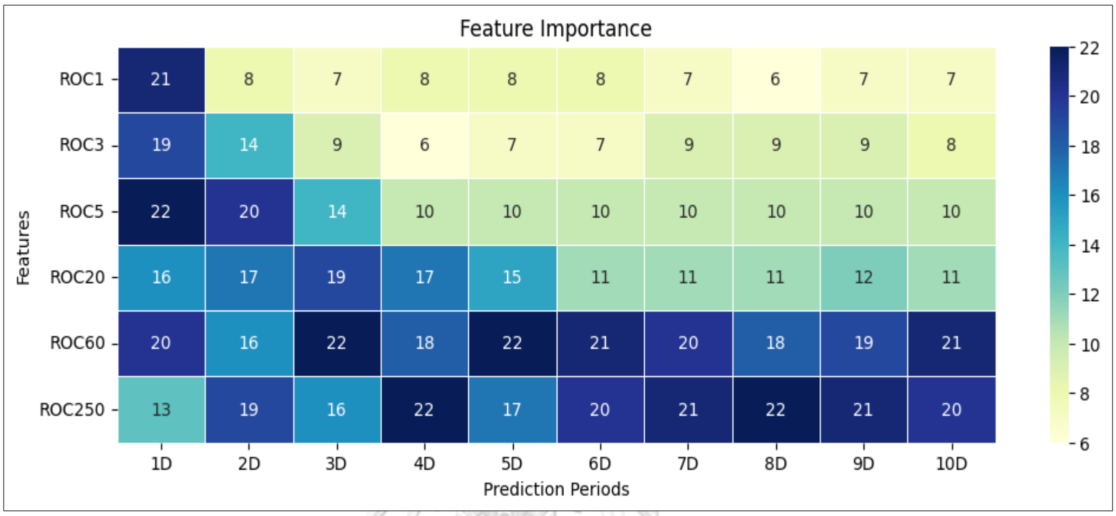

Figure 35 Momentum features essential for classifying ETH returns next 1-10day

Furthermore, the price trend of Bitcoin, which belongs to the same asset class as Ethereum, was found to have a significant influence on Ethereum's prediction. Specifically, the BTCROC values over 5 days, 20 days, 60 days, and 250 days played a crucial role in forecasting Ethereum's price trend. These findings support the hypothesis that market sentiments and the relationship between financial assets have an impact on price movements as shown Figure 36.

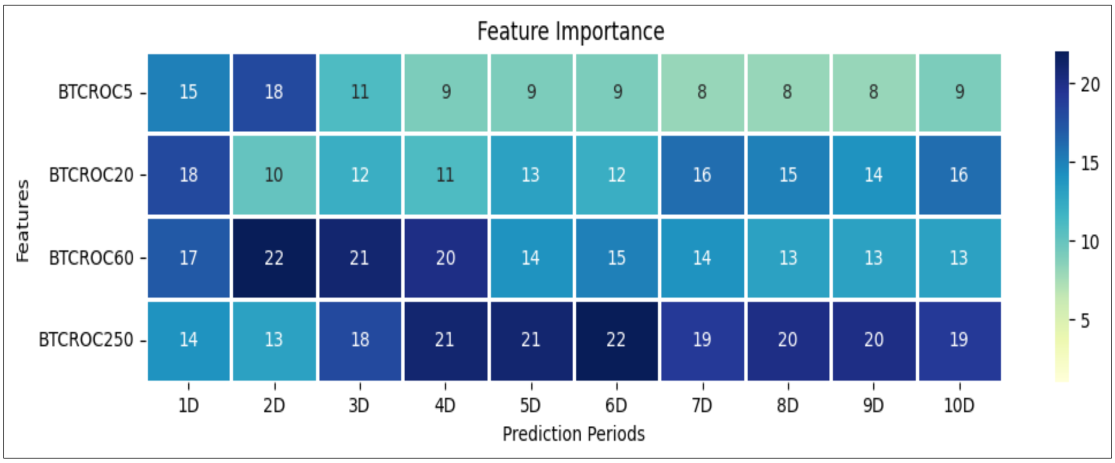

Figure 36 Sentiment features important to classify ETH returns next 1-10day

Figure 37 illustrates part of boolean variables simply indicating true or false consisting of isAboveMA20, isAboveMA50, isAboveMA200, isETHUPTrend, and isBTCUptrend. This group of indicators was found very little importance or almost no impact on predicting the ETH returns next X-day.

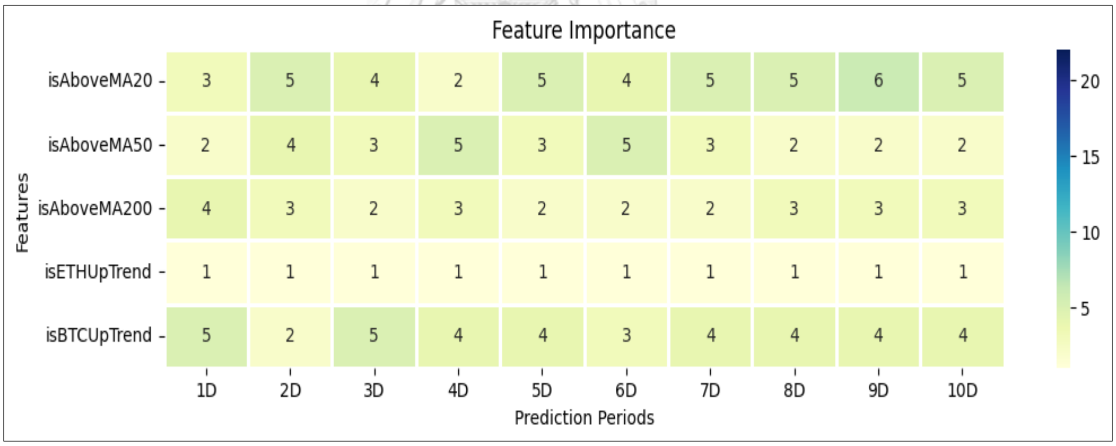

Figure 37 Features nonessential for classifying ETH returns next 1-10day

The model performance evaluated on test set is reported with the AUC-ROC metric. Figure 38–47 illustrate the AUC-ROC curves associated with each model predicting ETH return next X-day ranging from 1-day to 10-day. For each model, the ROC of each class (Downtrend, Sideway, and Uptrend), in addition to the micro average ROC and macro average ROC are reported. The performance of predicting return next X-day is also summarized in Table 7, for ease of understanding.

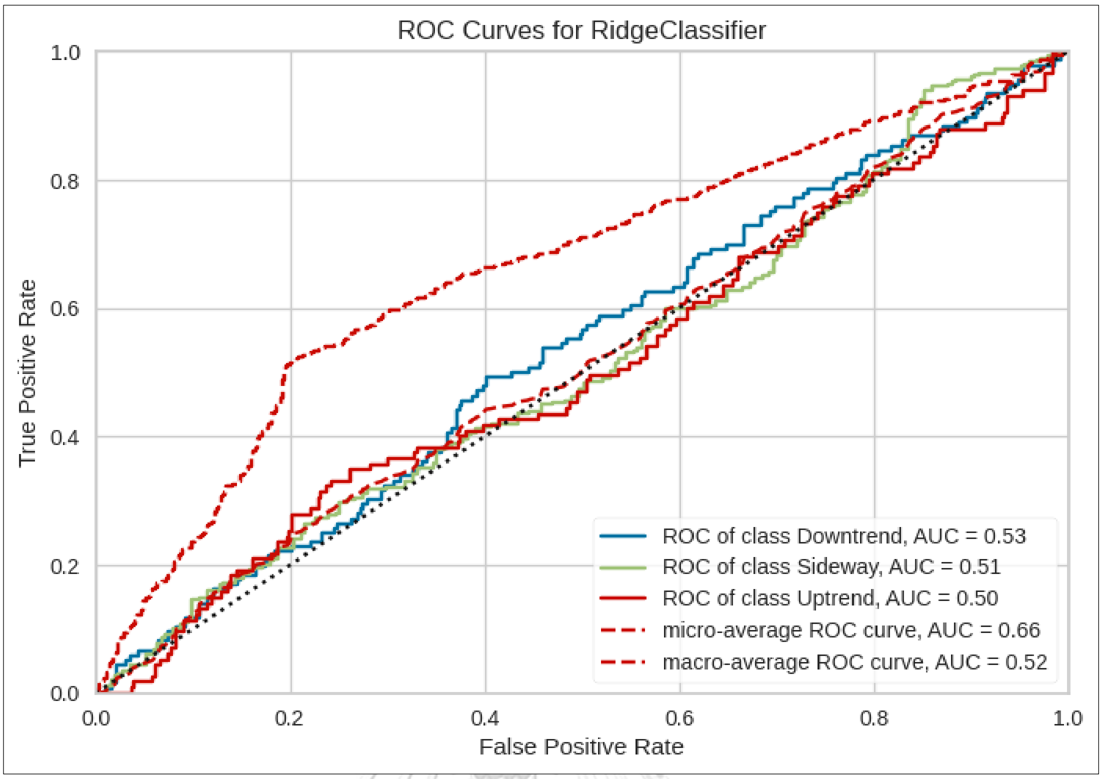

Figure 38 AUC-ROC of ETH return next 1-day

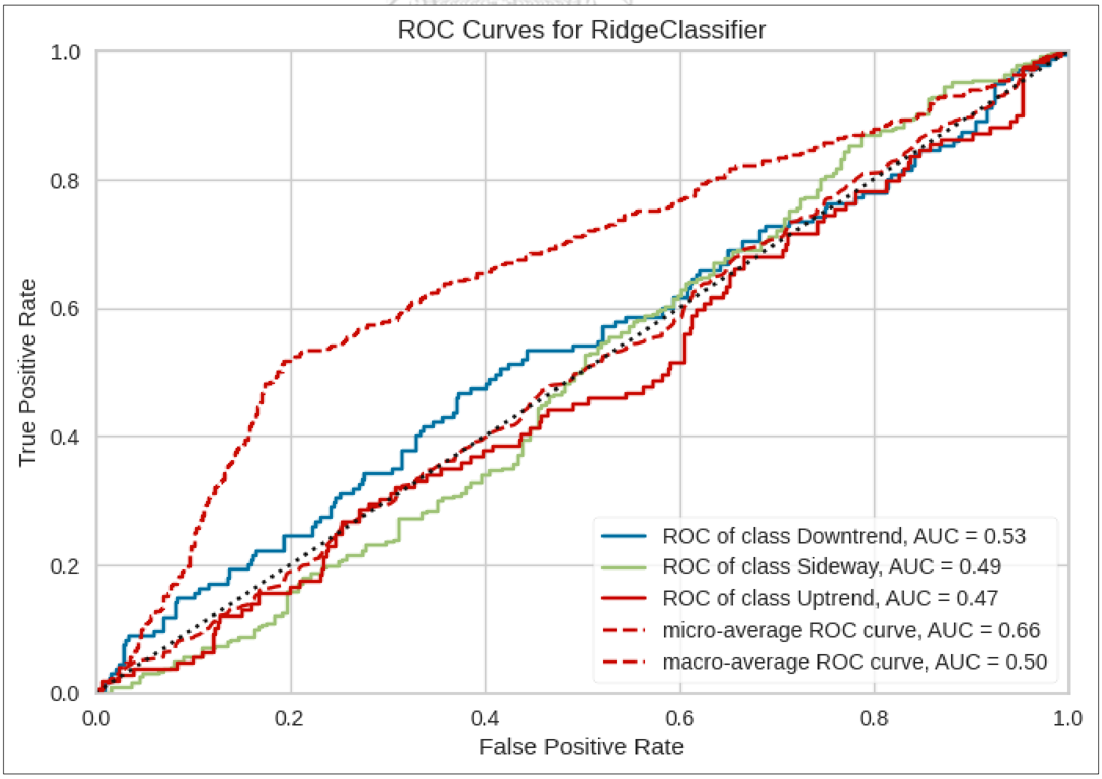

Figure 39 AUC-ROC of ETH return next 2-day

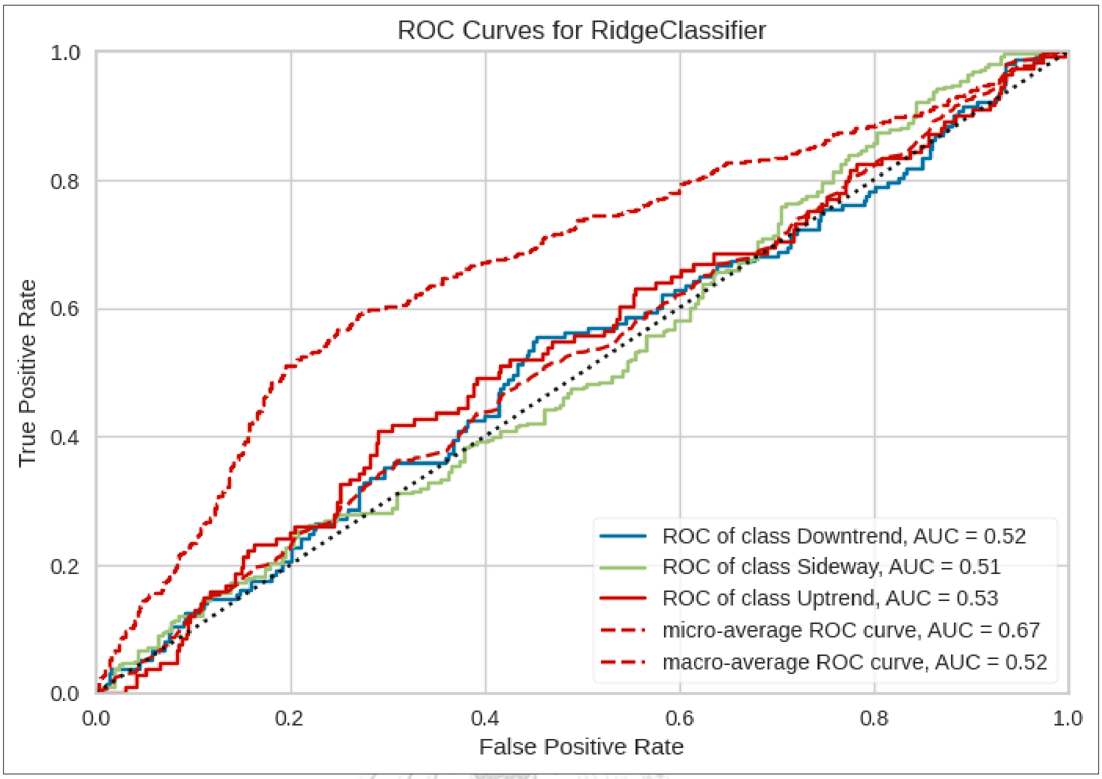

Figure 40 AUC-ROC of ETH return next 3-day

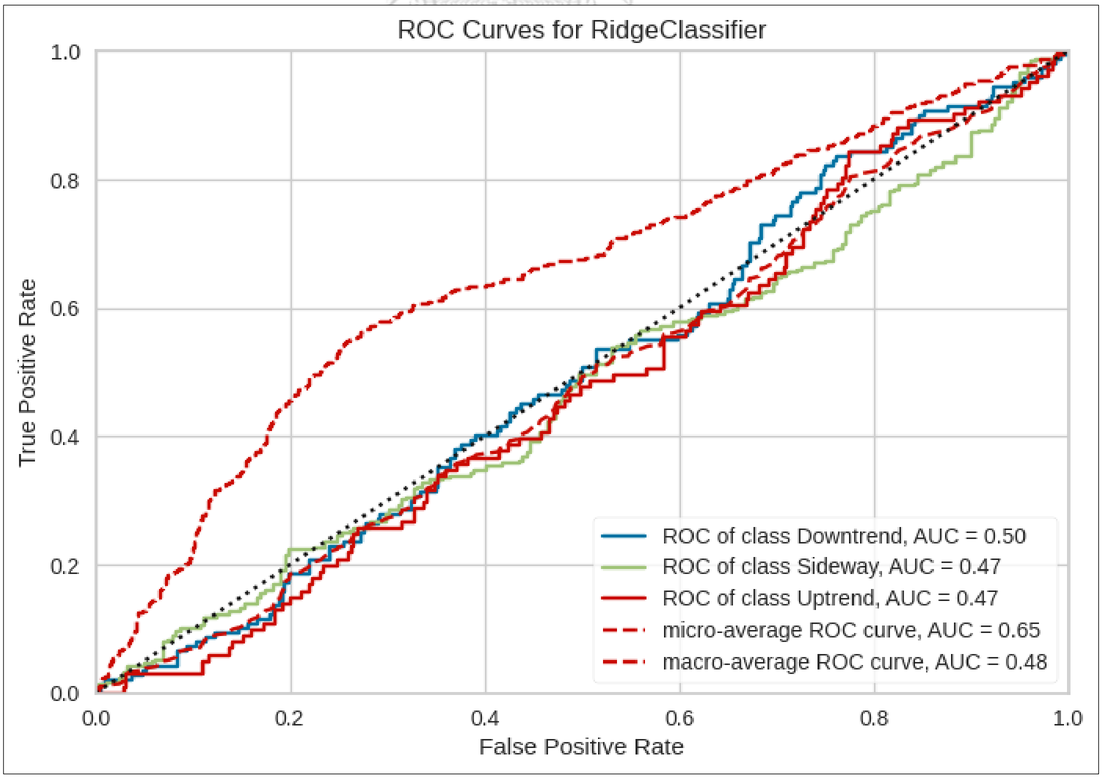

Figure 41 AUC-ROC of ETH return next 4-day

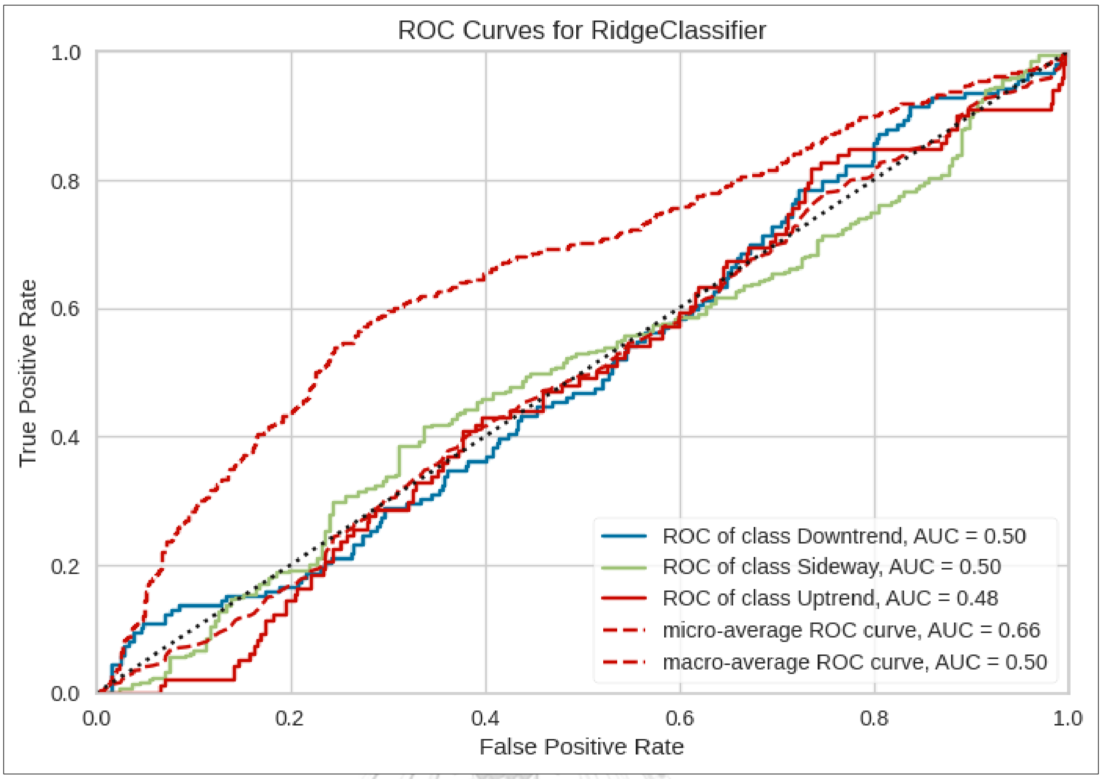

Figure 42 AUC-ROC of ETH return next 5-day

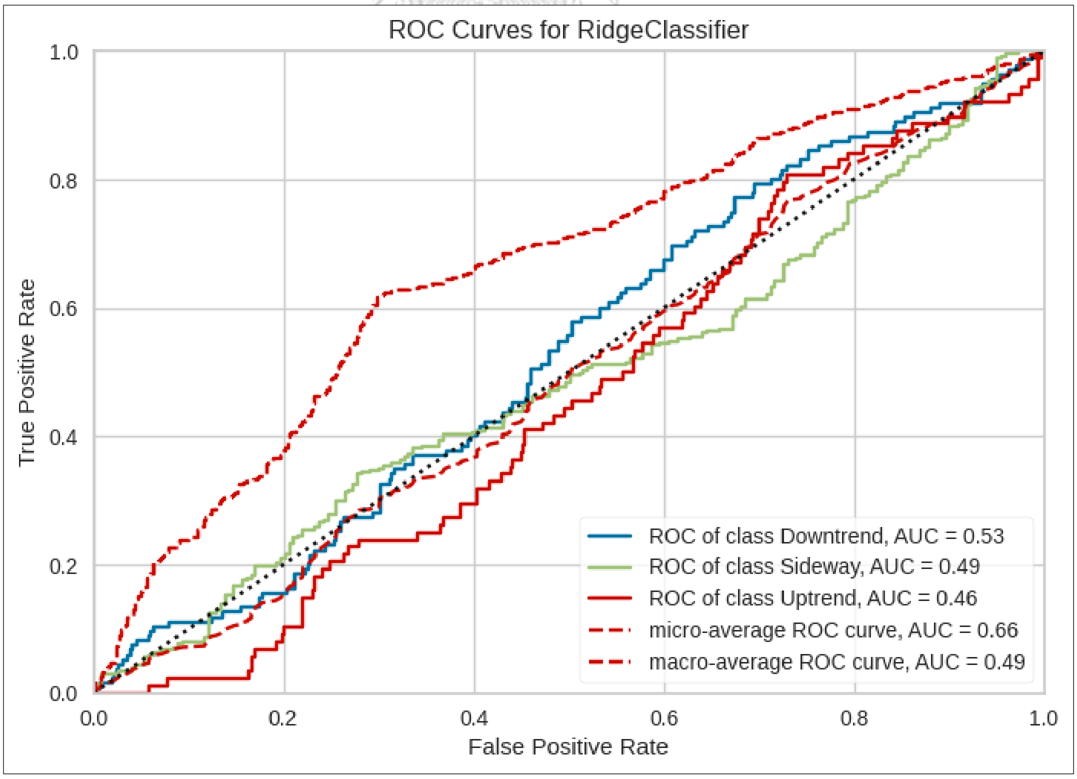

Figure 43 AUC-ROC of ETH return next 6-day

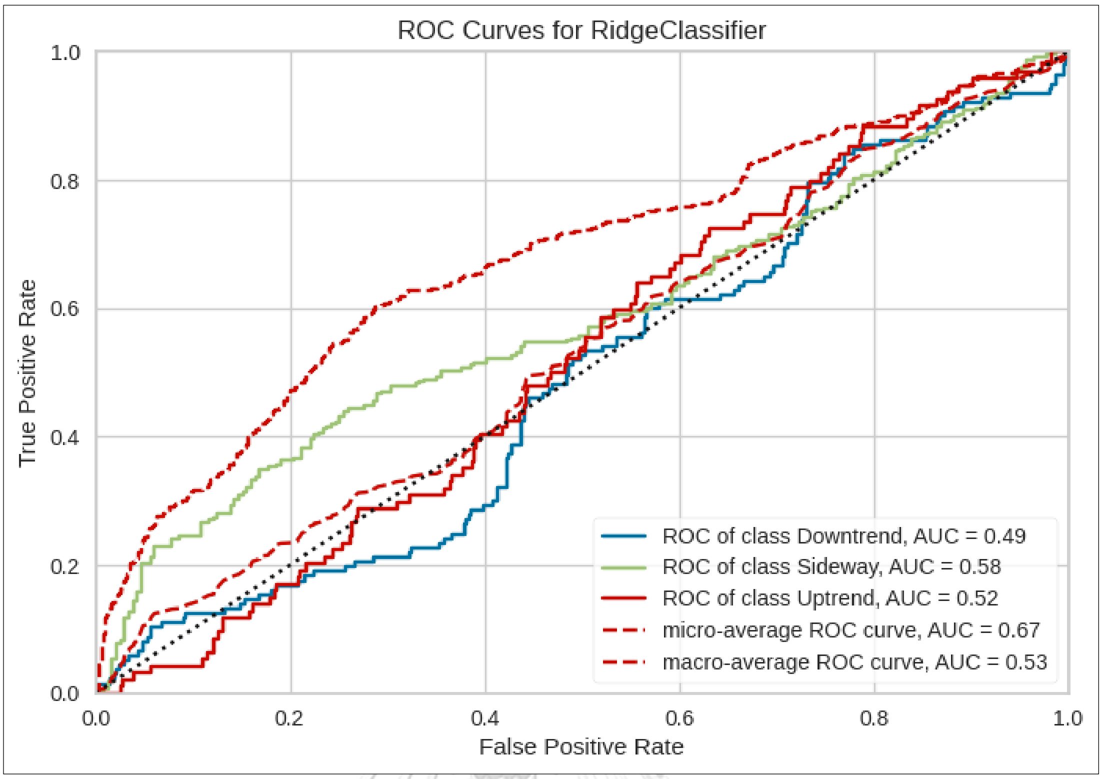

Figure 44 AUC-ROC of ETH return next 7-day

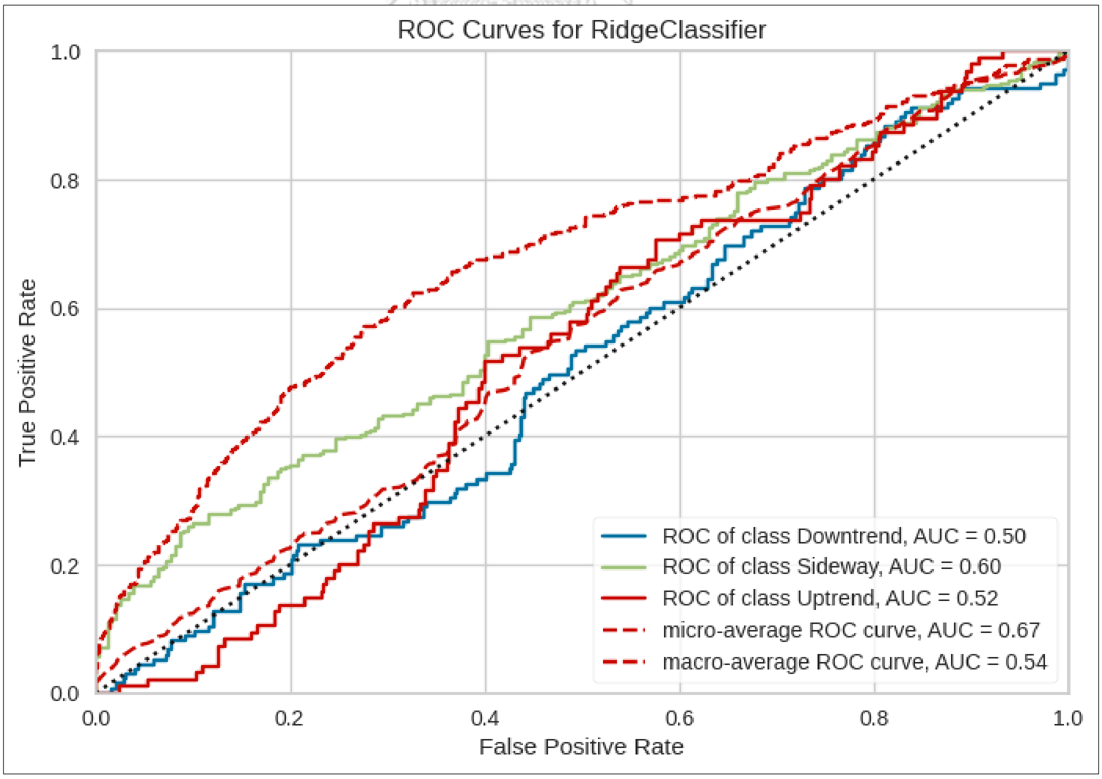

Figure 45 AUC-ROC of ETH return next 8-day

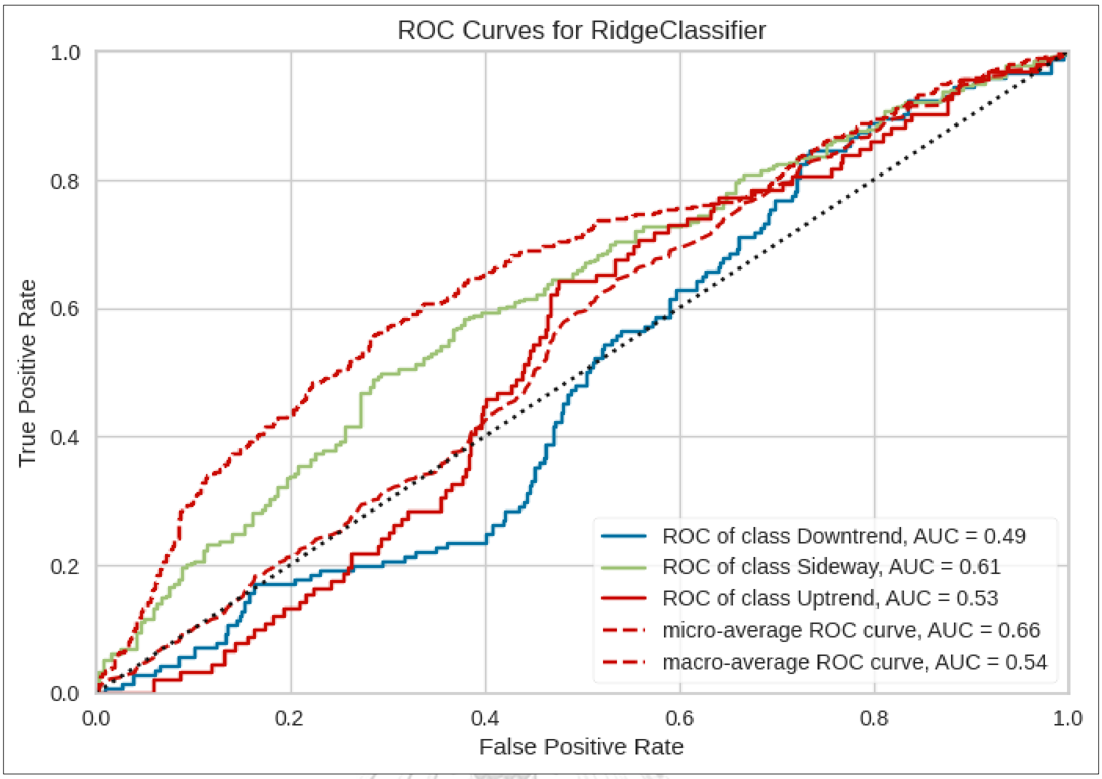

Figure 46 AUC-ROC of ETH return next 9-day

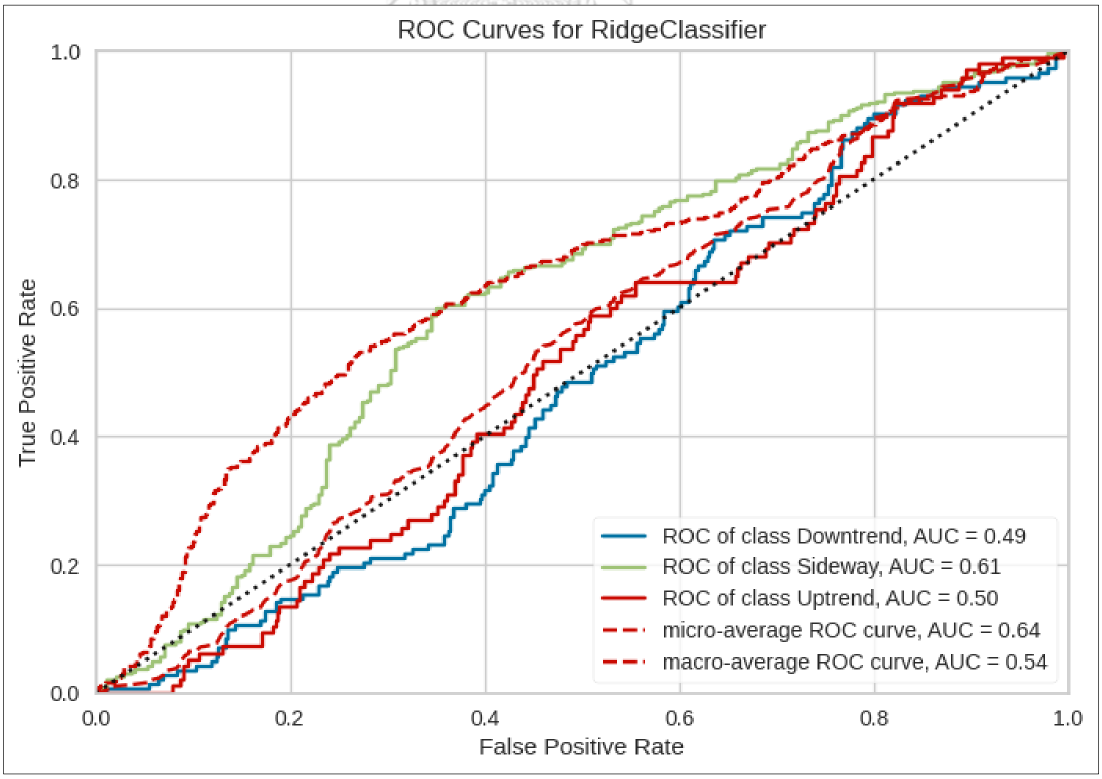

Figure 47 AUC-ROC of ETH return next 10-day

The model performance differs depending on the time period from 1 to 10 days and class being predicted. When predicting the Downtrend class, the AUC ranges from 0.49 to 0.53, indicating moderate to good performance in identifying downward trends. For the Sideway class, the AUC ranges from 0.47 to 0.61, indicating relatively good performance in identifying sideway trends. And the Uptrend class, the AUC ranges from 0.46 to 0.53, suggesting predictive performance in recognizing uptrend movements. The micro-average AUC, which reflects the overall model performance across all classes, ranges from 0.64 to 0.67. and the macro-average AUC, representing the average performance per class, ranges from 0.48 to 0.54.

In summary, these results show that the model performs differently when predicting return directions over different time periods. It performs better in identifying Sideway class compared to Downtrend and Uptrend class.

Table 7 Summary of Prediction Performance of ETH return next 1-10day

	Model Performance				
Return next X-day	Downtrend AUC	Sideway AUC	Uptrend AUC	Micro-average	Macro-average
1D	0.53	0.51	0.50	0.66	0.52
2D	0.53	0.49	0.47	0.66	0.50
3D	0.52	0.51	0.53	0.67	0.52
4D	0.50	0.47	0.47	0.65	0.48
5D	0.50	0.50	0.48	0.66	0.50
6D	0.53	0.49	0.46	0.66	0.49
7D	0.49	0.58	0.52	0.67	0.53
8D	0.50	0.60	0.52	0.67	0.54
9D	0.49	0.61	0.53	0.66	0.54
10D	0.49	0.61	0.50	0.64	0.54

CHAPTER 5
Conclusion

This research presents predictive models trained with XGBoost algorithm. The objective is to classify various short-term returns of Ethereum based on technical factors such as momentum factor, volatility factor, and sentiment factor. The short-term returns denote the investment period between 1 and 10 days. The twenty-two indicators belonging to technical factor were selected as features used for XGBoost learning to predict the class of return next X-day ranging from 1 to 10 days. The model performance was evaluated using the multi-class AUC-ROC metric, which measures the accuracy of the model in predicting three classes: Downtrend, Sideway, and Uptrend. The micro-average ROC curve achieved an AUC of each day between 0.65 and 0.67, indicating that the model performed reasonably well in predicting the overall trend of ETH price. The study discovered that analyzing momentum factors could help investors and traders make informed investment decisions in cryptocurrency. However, it is worth noting that predicting the direction of price movement in cryptocurrency markets could be very challenging due to the high volatility and unpredictability of the market. Despite the model's success in predicting trends, the authors emphasize the need for further research within the field of cryptocurrency trading and investment. Specifically, understanding dynamics or other indicators that drive price movements within this rapidly market.

Appendix

Technical Indicator Measurement

1. Rate of Change over 1 day (ROC1) is the percentage change in price over 1 day.

$$ROC1 = \left(\frac{Close\_Price[p]}{Clos\_\ Price\ [p-n]} \ -1\right) * 100 \qquad (1)$$

where:

- Close_Price [p] = close price of the current day
- Close_Price [n] = close price of the previous day

2. Rate of Change over 3 days (ROC3) is the percentage change in price over 3 days.

$$ROC3 = \left(\frac{Close\_Price[p]}{Clos\_\ Price\ [p-n]} \ -1\right) * 100 \qquad (2)$$

where:

- Close_Price [p] = close price of the current day
- Close_Price [n] = close price of 3 days ago

3. Rate of Change over 5 days (ROC5) is the percentage change in price over 5 days.

$$ROC5 = \left(\frac{Close\_Price[p]}{Clos\_\ Price\ [p-n]} \ -1\right) * 100 \qquad (3)$$

where:

- Close_Price [p] = close price of the current day
- Close_Price [n] = close price of 5 days ago

4. Rate of Change over 20 days (ROC20) is the percentage change in price over 20 days.

$$ROC20 = \left(\frac{Close\_Price[p]}{Clos\_\ Price\ [p-n]} \ -1\right) * 100 \qquad (4)$$

where:

- Close_Price [p] = close price of the current day
- Close_Price [n] = close price of 20 days ago

5. Rate of Change over 60 days (ROC60) is the percentage change in price over 60 days.

$$ROC60 = \left(\frac{Close\_Price[p]}{Clos\_\ Price\ [p-n]}\ -1\right) * 100 \qquad (5)$$

where:

- Close_Price [p] = close price of the current day
- Close_Price [n] = close price of 60 days ago

6. Rate of Change over 250 days (ROC250) is the percentage change in price over 250 days.

$$ROC250 = \left(\frac{Close\_Price[p]}{Clos\_\ Price\ [p-n]}\ -1\right) * 100 \qquad (6)$$

where:

- Close_Price [p] = close price of the current day
- Close_Price [n] = close price of 250 days ago

7. Relative Strength Index (RSI) measures change of price movement to identify overbought and oversold condition.

$$RSI\ = 100 - \left(\frac{100}{1+(1+RS)}\right) \qquad (7)$$

where:

- RS = Average Gain / Average Loss

8. 52 Week High Returns (H52W) is the percentage different between current price and highest price in the past of 52 weeks.

$$H52W\ = \left(\frac{Close\_Price\ [p]}{Close\_Price[hp]}\right) * 100 \qquad (8)$$

where:

- Close_Price [p] = close price of the current day
- Close_Price [hp] = close price of the highest price in the past 52 weeks

9. Is above 20 days moving average (isAboveMA20) measure the average of the current price above 20-day.

$$isAboveMA20 = If\ Close\_Price\ [p] > MA20)\ then\ 1\ else \tag{9}$$

where:

- Close_Price [p] = close price of the current day
- MA20 = (Sum of close price over 20 days)/20

10. Is above 50 days moving average (isAboveMA50) measure the average of the current price above 50-day.

$$isAboveMA50 = If\ Close\_Price\ [p] > MA50)\ then\ 1\ else \tag{10}$$

where:

- Close price [p] = close price of the current day
- MA50 = (Sum of close price over 50 days)/50

11. Is above 200 days moving average (isAboveMA200) measure the average of the current price above 200-day.

$$isAboveMA200 = If\ Close\_Price\ [p] > MA200)\ then\ 1\ else \tag{11}$$

where:

- Close price [p] = close price of the current day
- MA200 = (Sum of close price over 200 days)/200

12. Correlation over 5 days (Corr5)

$$r_{Xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\,(Y_i - \bar{Y})^2} \tag{12}$$

where:

- x = daily close price 5 day of ETH
- y = daily close price 5 day of BTC
- x bar = Average close price return 5 day of ETH
- y bar = Average close price return 5 day of ETH

13. Correlation over 20 days (Corr20)

$$r_{Xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\,(Y_i - \bar{Y})^2} \tag{13}$$

where:

- x = daily close price 20 day of ETH
- y = daily close price 20 day of BTC
- x bar = Average close price return 20 day of ETH
- y bar = Average close price return 20 day of ETH

14. Correlation over 60 days (Corr60)

$$r_{Xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\,(Y_i - \bar{Y})^2} \tag{14}$$

where:

- x = daily close price 60 day of ETH
- y = daily close price 60 day of BTC
- x bar = Average close price return 60 day of ETH
- y bar = Average close price return 60 day of ETH

15. Correlation over 250 days (Corr250)

$$r_{Xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\,(Y_i - \bar{Y})^2} \tag{15}$$

where:

- x = daily close price 250 day of ETH
- y = daily close price 250 day of BTC

- x bar = Average close price return 250 day of ETH
- y bar = Average close price return 250 day of ETH

16. Rate of Change of BTC price over 5 days (BTCROC5)

$$BTCROC5 = \left(\frac{BTC\_Close\_Price[p]}{BTC\_Close\_Price\ [p-n]}\ -1\right) * 100 \qquad (16)$$

where:

- BTC_Close_Price [p] = BTC close price of the current day
- BTC_Close_Price [n] = BTC close price of 5 days ago

17. Rate of Change of BTC price over 20 days (BTCROC20)

$$BTCROC20 = \left(\frac{BTC\_Close\_Price[p]}{BTC\_Close\_Price\ [p-n]}\ -1\right) * 100 \qquad (17)$$

where:

- BTC_Close_Price [p] = BTC close price of the current day
- BTC_Close_Price [n] = BTC close price of 20 days ago

18. Rate of Change of BTC price over 60 days (BTCROC20)

$$BTCROC60 = \left(\frac{BTC\_Close\_Price[p]}{BTC\_Close\_Price\ [p-n]}\ -1\right) * 100 \qquad (18)$$

where:

- BTC_Close_Price [p] = BTC close price of the current day
- BTC_Close_Price [n] = BTC close price of 60 days ago

19. Rate of Change of BTC price over 250 days (BTCROC250)

$$BTCROC250 = \left(\frac{BTC\_Close\_Price[p]}{BTC\_Close\_Price\ [p-n]}\ -1\right) * 100 \qquad (19)$$

where:

- BTC_Close_Price [p] = BTC close price of the current day
- BTC_Close_Price [n] = BTC close price of 250 days ago

REFERENCES

[1] Reiff, N. (2022). "Bitcoin vs Ethereum: What's the Difference." Retrieved 31 Jan, 2023, from https://www.investopedia.com/articles/investing/031416/bitcoin-vs-ethereum-driven-different-purposes.asp.

[2] CoinGecko (2023). "Ethereum Price Char." Retrieved 28 Feb, 2023, from https://www.coingecko.com/en/coins/ethereum/historical_data.

[3] Staff, C. (2022). "Who Created Ethereum and Who Controls It Now.". Retrieved 17 Feb, 2023, from https://www.gemini.com/cryptopedia/co-founder-of-ethereum-founder-vitalik-buterin-ether-eth.

[4] Rodec, D. (2023). "What Is Ethereum? How Does It Work?". Retrieved 1 March, 2023, from https://www.forbes.com/advisor/investing/cryptocurrency/what-is-ethereum-ether/.

[5] ZIPMEX (2022). "What is Ethereum." Retrieved 17 Feb, 2023, from https://zipmex.com/th/learn/facts-about-ethereum/.

[6] Kelly, K. (2023). "What is Ethereum? Explained With Features and Applications." Retrieved 1 March, 2023, from https://www.simplilearn.com/tutorials/blockchain-tutorial/what-is-ethereum.

[7] Chang, S. (2019). "Ethereum Smart Contracts Vulnerable to Hacks: $

4 Million in Ether at Risk." Retrieved 1 March, 2023, from https://www.investopedia.com/news/ethereum-smart-contracts-vulnerable-hacks-4-million-ether-risk/.

[8] CoinDesk (2023). "Ether Price." Retrieved 17 Feb, 2023, from https://www.coindesk.com/price/ethereum/.

[9] Murphy, J. J. (1999). Technical analysis of the financial markets: A comprehensive guide to trading methods and applications.

[10] West, Z. (2020). "Calculating Moving Averages in Python." Retrieved 1 March, 2023, from https://www.alpharithms.com/calculating-moving-averages-in-python-585117/.

[11] Academy, I. (2021). "Moving aveages." The basic of technical analysis. Retrieved 1 March, 2023, from https://www.ig.com/en/ig-academy/the-basics-of-technical-analysis/moving-averages.

[12] Fernando, J. (2023). "Relative Strength Index (RSI) Indicator Explained With Formula." Retrieved 17 Feb, 2023, from https://www.investopedia.com/terms/r/rsi.asp.

[13] Michael, B. (2022). "What Is The Relative Strength Index (RSI) In Stocks?". Retrieved 1 March, 2023, from https://ibkrcampus.com/traders-insight/securities/stocks/what-is-the-relative-strength-index-rsi-in-stocks/.

[14] CRYPTOHOPPER (2023). "Technical Indicators: Rate Of Change (ROC)." Retrieved 28 Feb, 2023, from https://www.cryptohopper.com/resources/technical-indicators/302-rate-of-change-roc.

[15] Yamanaka, S. (2023). "Working Money: Average True Range by Sharon Yamanaka." Stocks & Commodities **20:3**: 76-79.

[16] Above the green line investing with rules (2021). "Average True Range (ATR)." Retrieved 31 March 2023, 2023, from https://abovethegreenline.com/technical-indicators/average-true-range/.

[17] StockCharts (2022). "Average True Rage (ATR)." Retrieved 31 March, 2023, from https://school.stockcharts.com/doku.php?id=technical_indicators:average_true_range_atr.

[18] Carr, M. (2022). "Measure Volitility With Average True Range." Retrieved 31 March, 2023, from Measure Volitility With Average True Range.

[19] Guestrin, T. C. a. C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International conference on knowledge discovery and data mining.

[20] Verma, N. (2022). "XGBoost Algorithm Explained in Less Than 5 Minutes." Retrieved 1 April, 2023, from XGBoost Algorithm Explained in Less Than 5 Minutes.

[21] Jing-Zhi Huang, W. H. a. J. N. (2019). "Predicting bitcoin returns using high-dimensional technical indicators." The Journal of Finance and Data Science **5(3)**: 140-155.

[22] Jie Wu, X. G., Mingqi Fang and JunHao Zhang (2022). Short term return prediction of cryptocurrency based on XGBoost algorithm. 2022 International Conference on Big Data, Information and Computer Network (BDICN).

[23] Li, Q. (2020). Predicting Trends of Bitcoin Prices Based on Machine Learning Methods. ICSEB '20: Proceedings of the 2020 4th International Conference on Software and e-Business, ACM Digital Library.

[24] S. Narkhede. "Understanding AUC - ROC Curve," 5 April, 2023; https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5.

[25] Pypi.org (2023). "stock indicators for Python." Retrieved 28 Feb, 2023, from https://pypi.org/project/stock-indicators/.

VITA

NAME	wipawee nayam
DATE OF BIRTH	9 Sep 1992
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTIONS ATTENDED	Bachelor of computer science, Mae Fah Luang University
HOME ADDRESS	888/68 Casaville Village Ring Road - Ram Inthra, Or Ngoen Subdistrict, Sai Mai District, Bangkok 10200
PUBLICATION	W. Nayam and Y.Limpiyakorn, “XGBoost for Classifying Ethereum Short-term Return Based on Technical Factor”, in Proceedings of the 9th International Conference on Computer Technology Applications (ICCTA 2023), Vienna, Austria W. Nayam and Y.Limpiyakorn, "Prediction of Ethereum Short-term Returns Using XGboost Model", International Journal of Emerging Technology and Advanced Engineering, 13 (8), August 2023.