

T.C.

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL

THE DEPARTMENT OF FINANCIAL TECHNOLOGY

**ENHANCING BITCOIN PRICE FORECASTING ACCURACY THROUGH
RIDGE REGRESSION: A COMPREHENSIVE ANALYSIS INCORPORATING
ON-CHAIN METRICS, ECONOMIC METRICS, MARKET ANALYSIS
METRICS, MARKET SENTIMENT INDICES, AND TECHNICAL
ANALYSIS INDICATORS**

MASTER'S THESIS

HASTI GHIASABADI FARAHANI

ISTANBUL 2024

T.C.

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL

THE DEPARTMENT OF FINANCIAL TECHNOLOGY

**ENHANCING BITCOIN PRICE FORECASTING ACCURACY THROUGH
RIDGE REGRESSION: A COMPREHENSIVE ANALYSIS INCORPORATING
ON-CHAIN METRICS, ECONOMIC METRICS, MARKET ANALYSIS
METRICS, MARKET SENTIMENT INDICES, AND TECHNICAL
ANALYSIS INDICATORS**

MASTER'S THESIS

THESIS ADVISOR

ASSIST. PROF. DR. MERT GÜL

ISTANBUL 2024



T.C.

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL

MASTER THESIS APPROVAL FORM

Program Name:	FINANCIAL TECHNOLOGY (ENGLISH, THESIS)
Student's Name and Surname:	HASTI GHIASABADI FARHANI
Name Of The Thesis:	ENHANCING BITCOIN PRICE FORECASTING ACCURACY THROUGH RIDG REGRESSION: A COMPREHENSIVE ANALYSIS INCORPORATING ON-CHAIN METRICS, ECONOMIC METRICS, MARKET ANALYSIS METRICS, MARKET SENTIMENT INDICES, AND TECHNICAL ANALYSIS INDICATORS
Thesis Defense Date:	5 th June, 2024

This thesis has been approved by the Graduate School which has fulfilled the necessary conditions as Master thesis.

Assoc. Dr. Yücel Batu SALMAN

Institute Director

This thesis was read by us, quality and content as a Master's thesis has been seen and accepted as sufficient.

	Title/Name	Institution	Signature
Thesis Advisor's	Assist. Prof. Dr. Mert GÜL	İstanbul Beykent University	
Member's	Assist. Prof. Dr. Serkan YEŞİLYURT	Bahçeşehir University	
Member's	Assist. Prof. Dr. Bekir AŞIK	İstanbul Beykent University	



I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Hasti Ghiasabadi Farahani

Signature:

ABSTRACT

ENHANCING BITCOIN PRICE FORECASTING ACCURACY THROUGH RIDGE REGRESSION: A COMPREHENSIVE ANALYSIS INCORPORATING ON-CHAIN METRICS, ECONOMIC METRICS, MARKET ANALYSIS METRICS, MARKET SENTIMENT INDICES, AND TECHNICAL ANALYSIS INDICATORS

Hasti Ghiasabadi Farahani

Master's Program in Financial Technology

Supervisor: Assist. Prof. Dr. Mert GÜL

May 2024, 74 pages

This thesis explores the effectiveness of Ridge Regression for forecasting Bitcoin prices, utilizing a machine learning algorithm adept at handling multicollinearity among predictors. It employs a comprehensive dataset drawn from diverse sources, including on-chain metrics, market sentiment indices like the VIX (Volatility Index), GEPU (Global Economic Policy Uncertainty Index), and the Crypto Fear and Greed Index, as well as technical indicators, to enhance the accuracy of Bitcoin price predictions. Ridge Regression is applied to assess the impact of each variable and fine-tune the regularization parameter (α), a critical step to mitigate overfitting and improve model predictability on new datasets. The research is carried out in the Python programming language within the JupyterLab environment, facilitating dynamic code development and data analysis. Preliminary findings affirm that this multifaceted analytical approach markedly boosts the accuracy of Bitcoin price forecasts, enriching financial econometrics literature by demonstrating the benefits of Ridge Regression in overcoming multicollinearity in financial time series. Additionally, the study highlights the model's utility for market participants, especially those involved in short-term trading and cryptocurrency investments, providing a robust tool for navigating cryptocurrency market complexities. This work also sets the

groundwork for future studies on dynamic modeling techniques and the integration of emerging metrics to further enhance the precision of financial market forecasts.

Keywords: Bitcoin Price, Econometric Forecasting, Regression Analysis, Cryptocurrency Market, and On-chain Analysis

JEL codes : C36, C53, C58, G11, G13



ÖZ

RIDGE REGRESYONU İLE BİTCOİN FİYAT TAHMİN DOĞRULUĞUNU ARTIRMA : ZİNCİR ÜZERİ METRİKLER, EKONOMİK METRİKLER, PİYASA ANALİZİ METRİKLERİ, PİYASA DUYARLILIK ENDEKSLERİ VE TEKNİK ANALİZ GÖSTERGELERİNİ İÇEREN KAPSAMLI BİR ANALİZ

Hasti Ghiasabadi Farahani

Finansal Teknoloji Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi Mert GÜL

Mayıs 2024, 74 sayfa

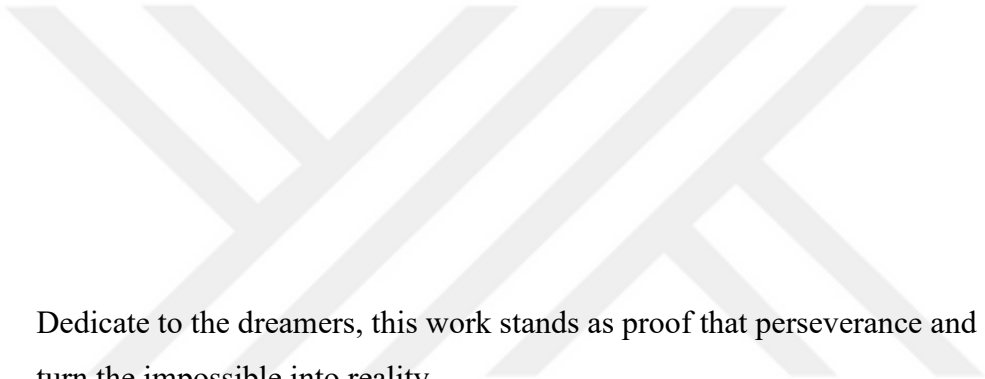
Bu tez, tahmin ediciler arasındaki çoklu doğrusallığı ele almakta usta bir makine öğrenme algoritması kullanarak Ridge Regresyonun Bitcoin fiyatlarını tahmin etmedeki etkinliğini araştırmaktadır. VIX (Volatilite Endeksi), GEPU (Küresel Ekonomik Politika Belirsizliği Endeksi) ve Kripto Korku ve Açgözlülük Endeksi gibi zincir üstü metrikler, piyasa duyarlılık endeksleri ve teknik göstergeleri içeren çeşitli kaynaklardan elde edilen kapsamlı bir veri seti kullanmaktadır. Ridge Regresyon, her bir değişkenin etkisini değerlendirmek ve düzenleme parametresini (alpha) ince ayar yapmak için uygulanmaktadır; bu, aşırı uyumu azaltmak ve yeni veri setlerinde model öngörülebilirliğini artırmak için kritik bir adımdır. Araştırma, dinamik kod geliştirme ve veri analizi kolaylaştıran JupyterLab ortamında Python programlama dili ile yürütülmüştür. İlk bulgular, bu çok yönlü analitik yaklaşımın Bitcoin fiyat tahminlerinin doğruluğunu önemli ölçüde artırdığını ve Ridge Regresyonun finansal zaman serilerindeki çoklu doğrusallığın üstesinden gelmedeki faydalarını göstererek finansal ekonometrik literatürü zenginleştirdiğini doğrulamaktadır. Ayrıca, çalışma modelin piyasa katılımcıları, özellikle kısa vadeli ticaret ve kripto para yatırımlarıyla ilgilenenler için faydasını vurgulayarak, kripto para piyasasının karmaşıklıklarında

gezinmek için sağlam bir araç sağlamaktadır. Bu çalışma, dinamik modelleme teknikleri ve finansal piyasa tahminlerinin doğruluğunu daha da artırmak için yeni metriklerin entegrasyonu üzerine gelecekteki çalışmalar için de temel oluşturmaktadır.

Anahtar Kelimeler : Bitcoin Fiyatı, Ekonometrik Tahmin, Regresyon Analizi, Kripto Para Piyasası, ve Zincir Üstü Analiz

JEL Kodları: C36, C53, C58, G11, G13





Dedicate to the dreamers, this work stands as proof that perseverance and faith can turn the impossible into reality.

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my supervisor, Assist. Prof. Dr. Mert GÜL, for his exceptional guidance, invaluable advice, and unwavering encouragement throughout my research. His dedication, expertise, and insightful criticism have been instrumental to the success of this project. Dr. GÜL's mentorship has significantly enriched my academic journey and personal development, and I am profoundly grateful for his support.

I am deeply grateful to my parents and my uncle for their unwavering support, love, and encouragement. Their belief in me, and sacrifices have been my greatest source of strength and inspiration. I owe my accomplishments to their steadfast presence and am forever indebted to them for their immeasurable contributions to my life.

The data for this study were exclusively supplied by Glassnode, which has not been utilized elsewhere. I extend my sincere gratitude to Glassnode for their collaboration and magnanimity in sponsoring the data for this research.

TABLE OF CONTENTS

ETHICAL CONDUCT	iii
ABSTRACT	iv
ÖZ	vi
DEDICATION	viii
ACKNOWLEDGEMENTS	ix
TABLE OF CONTENTS	x
LIST OF TABLES.....	xiii
LIST OF FIGURES.....	xiv
LIST OF ABBREVIATIONS.....	xv
Chapter 1: Introduction	1
Chapter 2: Literature Review	7
2.1. Forecasting Cryptocurrency Prices.....	7
2.2. Price Forecasting in Traditional Market	9
2.3. Bitcoin	10
2.3.1. Factors Affecting Changes in the Price of Bitcoin	13
2.3.2. Bitcoin Dominance	15
2.4. On-chain Analysis.....	16
2.5. Market Sentiment Analysis	18
2.5.1. Volatility index (VIX).....	19
2.5.2. Crypto Fear & Greed Index	21
2.6. Technical Indicators	22
2.6.1. Relative Strength Index	23
2.6.2. Simple Moving Average.....	23
2.6.3. Exponential Moving Average	24
2.7. Related Work	25
Chapter 3: Methodology.....	34
3.1. Introduction	34
3.2. Research Design.....	35
3.3. Research Philosophy.....	35
3.4. Research Type	36
3.5. Research Strategy	37
3.6. Time Horizon	38
3.7. Sampling Strategy	38
3.8. Data and Data Collection	38

3.8.1.	Dependent variable	39
3.8.2.	Independent variables	40
3.9.	Data Analysis	44
3.9.1.	Data Pre-processing	44
3.9.2.	Model	45
3.9.3.	Conclusion of Optimized Model after Comprehensive Comparative Analysis	48
Chapter 4:	Findings	50
4.1.	Correlation Analysis of Selected Independent Variables	50
4.1.1.	Correlation between Bitcoin price and Balance of Exchange	50
4.1.2.	Correlation Between Bitcoin Price and Mining Revenue	51
4.1.3.	Correlation between Bitcoin Price and HODL	52
4.1.4.	Correlation between Bitcoin Price and Number of Active Addresses	53
4.1.5.	Correlation between Bitcoin price and Net Realized Profit/Loss (NRPL)	54
4.1.6.	Correlation between Bitcoin price and Fee Ratio.....	55
4.1.7.	Correlation between Bitcoin Price and Spent Output Profit Ratio (SOPR)	56
4.1.8.	Correlation between Bitcoin Price and Coin Days Destroyed (CDD).....	57
4.1.9.	Correlation between Bitcoin and the VIX.....	58
4.2.	Correlation Map	59
4.3.	Training Models	62
4.4.	Understanding Alpha in Ridge Regression	63
4.4.1.	Role of alpha	63
4.4.2.	Impact of alpha.....	63
4.4.3.	Choosing alpha	63
4.4.4.	Optimization techniques.....	64
4.5.	Predictor's Coefficients in Ridge Regression	65
4.5.1.	Magnitude and Direction	66
4.5.2.	Relative importance	66
4.5.3.	Impact of regularization.....	66
4.5.4.	Practical interpretation.....	67
4.6.	Model Formulation, Predictive Accuracy, and Residual Analysis in Ridge Regression-Based Bitcoin Price Forecasting.....	68
Chapter 5:	Discussion and Conclusion	71
5.1.	Implications	71
5.2.	Limitation.....	72
5.3.	Recommendation for Future study	74

REFERENCES	75
------------------	----



LIST OF TABLES

TABLES

Table 1 Performance Metrics 77



LIST OF FIGURES

FIGURES

Figure 1 On-chain analysis indicators.....	31
Figure 2 VIF Scores of Different Variables	59
Figure 3 Ridge Trace Plot with Cleaned Data.....	62
Figure 4 Models' Comparison with seen data.....	63
Figure 5 Correlation between Bitcoin price and Balance of Exchange	66
Figure 6 Correlation between Bitcoin Price and Mining Revenue	67
Figure 7 Correlation between Bitcoin Price and HODL.....	68
Figure 8 Correlation between Bitcoin Price and Number of Active Addresses.....	69
Figure 9 Correlation between Bitcoin Price and Net Realized Profit/Loss (NRPL)..	70
Figure 10 Correlation between Bitcoin Price and Fee Ratio.....	71
Figure 11 Correlation between Bitcoin Price and Spent Output Profit Ratio (SOPR)	72
Figure 12 Correlation between Bitcoin Price and Coin Days Destroyed (CDD)	73
Figure 13 Correlation between Bitcoin and the VIX	74
Figure 14 Correlation Map.....	75
Figure 15 Correlation Map (Selected metrics).....	76
Figure 16 Models' Comparison with Unseen Data	77
Figure 17 Optimal Alpha in Ridge Regression	78
Figure 18 Alpha Values Vs. RMSE.....	80
Figure 19 Coefficients for Predicting 1 Day	82
Figure 20 Codes Related to Comparison Real Value of BTC and Forecasted Value with Error via Ridge Regression	84
Figure 21 RMSE Values Obtained for Consecutive Days	88
Figure 22 RMSE Graph of Consecutive Days	88

LIST OF ABBREVIATIONS

AODE	Averaged One-Dependence Estimators
ARIMA	Autoregressive Integrated Moving Average
BTC	Bitcoin
CFG	Crypto Fear and Greed Index
CNN	Convolutional Neural Network
EC	Economic Curve
GEPU	Global Economic Policy Uncertainty Index
	Generalized Auto Regressive Conditional
GARCH	Heteroskedasticity
IGO	Initial Game Offering
IS	Industry System
ISOMAP	Isometric Mapping
PCA	Principal Component Analysis
PTO	Paid Time Off
SCR	Selective Catalytic Reduction
VIX	Volatility Index

Chapter 1

Introduction

The financial markets have experienced a significant transformation in recent years, marked by the advent of cryptocurrencies as a ground-breaking asset class. Among these digital assets, Bitcoin stands out as a central figure, evolving from an experimental concept into a global financial phenomenon. Its journey since inception in 2009 has been characterized by unprecedented growth and dramatic price fluctuations, capturing the attention of investors, analysts, and researchers. This surge in interest aims to unravel the complexities of predicting Bitcoin's future price movements, a task that has proven to be both challenging and highly sought after. Bitcoin often hailed as digital gold, exhibits remarkable volatility, a characteristic that has ignited both fascination and concern within the financial community. This volatility, seen as both a risk and an opportunity, places Bitcoin under intense scrutiny from various stakeholders, including investors, policymakers, and academics. The ability to forecast Bitcoin's price accurately, amidst the fluctuations influenced by a wide range of factors from market sentiment to geopolitical events, is crucial for strategic decision-making and investment planning.

This thesis seeks to navigate the intricate dynamics of the cryptocurrency market by employing advanced statistical methods and exploring novel data sources. Specifically, it utilizes the Ridge Regression model to predict Bitcoin prices, incorporating a carefully selected set of independent variables believed to significantly impact its market behaviour. These variables include on-chain data, Volatility Index (VIX), the Global Economic Policy Uncertainty Index (GEPU) and the Crypto Fear and Greed Index. The selection of Ridge Regression is deliberate, aiming to address the challenges posed by multicollinearity among predictors, thereby enhancing the model's predictive accuracy and reliability. Through this analytical approach, this study endeavors to shed light on the driving forces behind Bitcoin's price movements and assess the effectiveness of Ridge Regression as a forecasting tool within the volatile and speculative realm of cryptocurrencies.

Satoshi Nakamoto, under the pseudonym Satoshi Nakamoto, introduced Bitcoin in 2008 with the paper "Bitcoin: A peer-to-peer electronic cash system."

Bitcoin, the leading example of blockchain technology, effectively addresses the challenges of double-spending and the Byzantine Generals Problem (Nakamoto, 2008). This innovation enables secure online transactions without relying on a centralized authority. This breakthrough led to the emergence of additional cryptocurrencies, contributing to the creation of an entirely new category of assets. Bitcoin has established a secure and cost-effective mechanism for conducting transactions across a decentralized, peer-to-peer network.

Following Bitcoin's introduction, alternative cryptocurrencies, commonly referred to as altcoins, began to emerge from 2011 onwards, with many still actively traded today. These altcoins are predominantly derivatives of Bitcoin, created from the same foundational code, hence the term "alternative coin." Among these, Ethereum stands out as the second most prominent blockchain platform, trailing only Bitcoin, boasting a market capitalization of 267 billion dollars. Ethereum extends the capabilities of decentralized financial transactions to include the creation and execution of rule-based smart contracts (automated code) on a decentralized framework akin to Bitcoin's. The year 2013 marked a notable period of rapid altcoin development, with new cryptocurrencies being introduced nearly every week, leading to the establishment of a broad and highly speculative trading environment (Elendner et al., 2018). The Department of Industry, Science, Energy and Resources in Australia has outlined a National Blockchain Roadmap, highlighting blockchain technology as a major technological innovation with substantial impact across both public and private sectors. Projections within the roadmap suggest that by 2025, the annual business value generated by blockchain technology could exceed US\$175 billion, with expectations of growth to surpass US\$3 trillion by 2030.

While Bitcoin was the inaugural digital currency to achieve significant adoption, the idea of a digital currency predates its existence, tracing back to experiments conducted by Cypherpunks in the 1990s. These early pioneers explored the potential for a digital currency that would operate independently of traditional financial institutions and government oversight. Among the initial endeavours that tackled cryptographic challenges to lay the groundwork for cryptocurrencies were b-money, bit gold, and reusable proof of works. However, the primary reason for the failure of these early projects was their dependence on a centralized entity to function,

a critical vulnerability that Bitcoin successfully addressed by eliminating the need for such intermediaries.

Bitcoin has not only found a smaller niche of its original intention as an alternative currency, yet it is also enjoyed to be a valuable asset (Baur et al., 2018). Many scientists, in their own works, delve into the Bitcoin's positions within the stock markets (anilla or a hedge) and on other stock markers indicators, like Standard & Poor's 500 Index (S&P500 Index) or the Volatility Index (VIX). As for the overall efficiency of the market, experiments performed by Jakub (2015) mean that Bitcoin conforms to the efficient market hypothesis, hence, an asset classification is quite apt, while Lahmiri et al. (2018) dispute the statement and offer one of the strongest evidences against it. The risk of the Bitcoin market basically is that people cannot count on these cryptocurrency markets as reliable hedge means. According to report of Bouri et al. (2017a). To that effect, according to Dyhrberg (2016), BTC can be hedged, and it trades closely with and sometimes displays lower spreads than equities found on major stock exchanges (Dyhrberg et al., 2018).

Finally, the studies conducted by the writers like Al-Yahyaee et al. (2019), Bouri et al. (2017b), Urquhart and Zhang (2019), Shahzad et al. (2019), and Chan et al. (2019) gave the same weight as this on that Bitcoin is a hedge. Nevertheless, recently, the asymmetric relations between market shocks and Bitcoin were noted, the same as the precious metals do, hence by Klein et al. (2018). In addition, it has been observed that the prevalent behaviour of Bitcoin has seemed to correlate with the S&P 500 Index (Georgoula et al., 2015; Conrad et al., 2018; Kjærland et al., 2018; Erdas and Caglar, 2018; Chan et al., 2019).

The Volatility Index (VIX), often heralded as the market's "fear gauge," offers a real-time assessment of market sentiment and anticipations of future volatility. Esteemed for its role in encapsulating expected market volatility, the VIX mirrors the sentiment and risk tolerance within traditional financial realms. An ascending VIX typically signifies a growth in market apprehension or uncertainty, rendering it an instrumental barometer of investor sentiment. Furthermore, the VIX may function as a contrarian indicator; an exceptionally elevated VIX could suggest that fear has peaked, possibly denoting an oversold market condition ripe for a reversal. On the flip side, a notably subdued VIX might indicate a sense of complacency, heralding

potential market adjustments. For risk management endeavours, the VIX and its related instruments stand as pivotal tools. Investors may employ these mechanisms to safeguard against market downturns or to fortify their portfolios amid heightened volatility. Additionally, the VIX can contribute to portfolio diversification strategies. In periods of market turmoil, the VIX's tendency to surge offers a counteracting effect to other portfolio assets, enhancing portfolio resilience. The VIX's prominence was markedly underscored during the 2008 global financial crisis, when it soared to unprecedented levels, accentuating its utility as a marker of market distress. This episode has entrenched its significance in financial market analyses, further highlighting its efficacy in hedging against tail risk — the risk of extreme market upheavals. Serving as a prognostic tool for expected volatility, the VIX is indispensable for traders and investors formulating strategies based on market volatility forecasts. The integral role of the VIX in financial markets lies in its capacity to furnish critical insights into market sentiment and volatility expectations, thus underpinning its value in investment strategies and financial scrutiny. In the realm of cryptocurrency, particularly Bitcoin, deciphering the dynamics between the VIX and Bitcoin's return patterns can unveil profound insights into how general market sentiments influence the volatility and valuation of digital currencies.

The Global Economic Policy Uncertainty (GEPU) Index is a tool that quantifies the uncertainty related to economic policies by analyzing a variety of textual sources, including news articles and reports from different countries. This index evaluates the degree of uncertainty linked to economic policy decisions, regulatory shifts, and similar factors. Central to the GEPU Index are three core components: the extent of news media coverage on economic policy uncertainty, the uncertainty introduced by tax code provisions affecting businesses, and the range of economic forecasts provided by professional economists. Distinct from other indices that might concentrate on specific nations or regions, the GEPU Index offers a worldwide view, compiling data across multiple countries to gauge the global level of economic policy uncertainty. This index is utilized by researchers, policymakers, and investors to gauge the effects of economic policy uncertainty on financial markets, economic development, and investment strategies. Elevated uncertainty levels can impact business investments, consumer behavior, and the broader economic climate. Furthermore, the GEPU Index is often examined in conjunction with financial market

indicators, noting that heightened uncertainty can lead to increased market volatility as investors respond to unpredictable policy environments. Businesses and investors might incorporate the GEPU Index into their strategic planning, potentially postponing investments or expansions during times of heightened uncertainty until the policy outlook stabilizes. Governments and economic policymakers also consider the GEPU Index when devising economic strategies, recognizing that high uncertainty levels could necessitate actions aimed at diminishing ambiguity and fostering economic steadiness. In essence, the GEPU Index serves as a comprehensive economic indicator, quantifying the impact of economic policy uncertainty through the analysis of textual materials and data from numerous countries. It sheds light on the manner in which policy uncertainty can influence financial markets, investment decisions, and economic growth at an international scale, offering valuable insights into the dynamics between economic policy uncertainty and the stability of financial markets, including the potential effects on Bitcoin's price.

This study integrates not only traditional financial metrics but also sentiment-driven indicators like the Crypto Fear and Greed Index providing by AlternativeMe to capture the unique market sentiment inherent to the cryptocurrency sphere. The Crypto Fear and Greed Index, a pivotal sentiment gauge, evaluates the collective mood and perceptions of cryptocurrency market participants, with a particular focus on Bitcoin. It aims to discern the prevailing market sentiment, oscillating between fear, which denotes caution, and greed, indicative of overconfidence. This index amalgamates various data points and sources, such as price volatility, trade volumes, social media chatter, surveys, and sentiment analysis tools, to distil a singular sentiment score. Presented on a spectrum from 0 to 100, the scale interprets 0 as maximum fear and 100 as maximum greed, with intermediate values depicting the market's fluctuating emotional states. Investors and traders might leverage the Crypto Fear and Greed Index as a strategic market-timing mechanism. An exceptionally high greed level could imply an overbought market poised for a downturn, whereas a pronounced fear level might signal an opportune moment for entry. Mirroring its utility in traditional financial contexts, this index can act as a contrarian indicator, where extreme fear suggests potential market undervaluation and extreme greed indicates possible overvaluation. Given the notorious volatility of cryptocurrency markets, sentiment can significantly sway price movements. The Crypto Fear and Greed Index aids in

decoding the complex dynamics of market sentiment's effect on cryptocurrency valuations. However, it's crucial to acknowledge the inherent limitations of this index, rooted in its reliance on subjective and sentiment-driven data, potentially susceptible to the ephemeral nature of social media trends and market speculation. Predominantly centred on Bitcoin and the wider cryptocurrency landscape, the index does not extend its insights to other asset categories or the traditional financial markets. Ultimately, it furnishes an additional layer of analysis for comprehending market sentiment, facilitating informed decision-making amidst the volatile domain of digital currencies.



Chapter 2

Literature Review

2.1.Forecasting Cryptocurrency Prices

A variety of techniques are a result of calculations of Bitcoin and other cryptocurrency prices. Being an example, Katsiampa, (2017) used only closing prices which are daily for assessing Generalized AutoRegressive Conditional Heteroskedasticity (GARCH)-type models and recognized that AR-GARCH is the most effective model after the best fitting metrics were used for choosing. In particular, Mallqui, & Fernandes, (2019) developed ensemble machine learning algorithms to forecast the exchange rates of Bitcoin. The ANN-GARCH combination, which combines Artificial Neural Network and the Auto-Regressive Conditional Heteroskedasticity was proven to be the most revealing model in a total of 12 formulas, and it was capable of forecasting Bitcoin volatility from the price series (Kristjanpoller, & Minutolo, 2018). McNally et al. (2018) used the Bayesian-optimized RNN and LSTM scheme, a representation of recurrent neural network and long short-term memory, trained on the market price data to predict three scenarios of price changes (up, down and flat) compared to traditional Autoregressive Integrated Moving Average (ARIMA) model, obtaining better results in time series prediction.

Alonso-Monsalve et al. (2020) made thorough review of neural network architectures to facilitate deep learning utilization in Bitcoin prediction of which Convolutional Neural Network (CNN)s were seen to be technically capable. Nevertheless, exclusively using the past price-data, as a place to start, is often unsatisfactory when it comes to forecasting the cost of Bitcoin. The latest studies therefore incorporate many sources of the data, as well as sentiment-based data, in an attempt to improve the forecasting precision. Kim et al. (2016) found out the people's comments on cryptocurrency online communities and labelled them accordingly. Also, Averaged One-Dependence Estimators (AODE) was employed by the researchers for price prediction. Stenqvist, & Lönnö, (2017) classified sentiment about Bitcoin on Twitter into Positive, Neutral and Negative using these categories to forecast price up or down move replacing them with neutral and notes high overall prediction accuracy. Likewise, Li et al. (2019) took the help of sentiment information from Twitter rather than technical analysis to predict the Bitcoin price movements. As the experts noted,

semantic analysis for forecasting incurs lower scope for performance improvement due to the fact that it is often difficult to classify sarcasm's peculiarities as a linguistic detail.

Mohapatra et al. (2019) approached the issue of Bitcoin price prediction by means of a real-time cryptocurrency price prediction platform, which operated with Twitter sentiments for Bitcoin price forecasting (Cheuque Cerda, & Reutter, 2019). Jain et al. (2018) posited that the price of the coin varies along these indicators, but economy sentiment also takes part in the nature of the cryptocurrency. While (as Rebane et al. (2018) showed) classifier models like Seq2Seq produce better results than autoregressive ones like ARIMA when combined with additional inputs from, e.g., Google Trends, Altcoin data or Bitcoin prices, it is important to note that there is still (as Hansen, & Lunde, (2005) point out) substantial potential left for AI models in finance. Sure, even though the Seq2Seq model works excellently just before the price increase, it could be severely affected by quick Bitcoin falling prices. Jang, & Lee, (2017) applied BNN method for prediction of Bitcoin prices with a problem of colinearity for macro variables and blockchain-derived features. By demonstrating the larger accuracy in relation to LR (Logistic Regression) and SVR (Support Vector Regression) models, they laid groundwork to the behaviour of BNN in both time-fluctuation of price and volatility predictions.

Amjad, & Shah, (2017) took bid and ask prices of Bitcoin and developed simple real-time machine learning models to predict three price states. They showed that proposed model can be better than economic curve (EC) and autoregressive integrated moving average (ARIMA) models. Similar prediction method was proposed by Guo et al. (2018) where probability-based temporal mixture models were introduced predicting 2-week volatility of Bitcoin prices and order book data, where bid orders (maximum price to buy) and ask orders (minimum price to sell) from exchange databases were taken into considerations. The accuracy of the models was higher compared to the base models presented by statistics and machine learning. This study applies new dimensions i.e. are drawn from exchange blockchain transactions to yield Bitcoin price predictions and analyzes their significance in the case in point. To the extent of our knowledge, it (Brain-Net integration) is a lead one among those who try at crossing such data to predict Bitcoin price.

2.2.Price Forecasting in Traditional Market

It is no coincidence that forecasting Bitcoin prices has similarities with predicting the future behaviour of other financial markets, namely, foreign exchange, stocks, commodities, and so on. Different methods were marked to be used in resolving the same, including the use of autoregression models to study predictability in stock returns (Hansen, & Lunde, 2005). Indeed, despite the fact the two processes are found non-linear and non-stationary the actual auto regressions models can often show us models which can appropriately forecast the prices of the stocks. Addressing these deficiencies has been achieved by through different machine learning strategy (Hagenau et al., 2013; Nam, & Seong, 2019; Hassan et al., 2013; Nayak et al., 2015). Hassan et al. (2013) constructed an Hidden Markov Model (HMM)-based adaptive fuzzy inference system to tackle price trends analysis, whereas Nayak et al. (2015) came out with an hybrid model that gone through the combination of SVM's two approaches and K-nearest neighbours to balance complexity and errors and, in this way, became a useful tool in finding out the price trend of stock market indices.

Deep learning, a powerful approach has already found a niche within the finance sector. This methodology has transformed this sector and others one after another. A set of studies have found that Artificial Neural Network (ANN) models have outperformed ARIMA models by number of times. Jiao, & Jakubowicz, (2017) consequently applied the conventional supervised machine learning methods which included neural network to predict stock prices based on the daily data of 463 stocks belonging to the S&P 500. Faccini presented additional tool, the representative investor's implied relative risk aversion, insider their study (Faccini et al., 2019). Further, the study by Neely et al. (2014) showed that both technical indicators (which consider past price and volume patterns) and macroeconomic variables, for example, dividend-price ratio, could improve the RMSE value. For extracting multi-frequency trading patterns State Frequency Memory (SFM), Zhang et al., (2017) having similar structure to Discrete Fourier Transform (DFT) have been used.

Sul et al. (2017), tried social sentiments and news for forecasting a stock price but they concatenated their vectors into one vector while doing so; thereby could be a loss of the relationships that could be inherent features. On one hand, Li et al. (2015) utilized stock data to compare the tensor-based approach with Selective Catalytic

Reduction (SCR) + Principal Component Analysis (PCA), and Isometric Mapping (ISOMAP) +SVR. On the other hand, Katsiampa, (2017) has applied SCR+PCA and ISOMAP+SVR on the specified dataset. They gave the evidence that the tensor-based approach seemed more powerful than the others. A distinction feature of Bitcoin in comparison to traditional financial markets is that it remains open 24/7, which means that Bitcoin trading never stops. This is different from the traditional stocks and forex. Besides, offline data and events can second the cryptocurrency price variations automatically, whereas the stock is excluded from these trading-hour confines. A variance of 23.97 can be seen in the consideration of stock dataset in the aforementioned work Zhang et al., (2017) against a mean of 50 stock prices. Compared to Bitcoin, the usual standard deviation for most traditional assets stands at 59.55 while that for Bitcoin is much greater with the standard deviation being 3923.15. This stresses upon a new model of forecast methodology which can tackle these price fluctuations that are way-too-much extreme.

2.3.Bitcoin

Bitcoin, the King of Cryptocurrencies, had its first mention by Satoshi Nakamoto (2009), also referred to as a nom de plume. In that year, there is a paper published on cryptography written by Nakamoto which appeared on a forum further developing the idea of a currency that was later known as Bitcoin. In early 2009, the Bitcoin source code and the compiled computer code were openly released by the creator on the "www.sourceforge.net." By June 2009, Nakamoto created the peer-to-peer Bitcoin network, which permitted the participants to monitor all transactions and mine Bitcoins. In the beginning period, number of miners in the network was very low, that graphics processing unit (GPU) then was able to collect a lot of Bitcoin.

Chinea-Rios et al. (2023) research checked the cryptocurrency's data and discovered that Nakamoto owned between 980.000 and 1.000.000 Bitcoins. Interestingly, not a single Satoshi had been spent, so the underlying purpose remained unknown. Nonetheless, the fact that Nakamoto spent these Bitcoins would lead to his identity leak because blockchain transactions are transparent among all network participants, and therefore tracing these Bitcoins could get Nakamoto's real-world identity (Chinea-Rios et al., 2023). Nakamoto 's decision to create Bitcoin as an autonomous system is in line with the prevailing public scepticism towards third-party

accountability resulting from 1990s events. According to Nakamoto, in the next 10 years virtual currencies will replace the conventional money.

Bitcoins serve as digital money working on the principle of protocol and cryptographic methods that make transactions secure and new units (Tschorsch, & Scheuermann, 2016). It stands out as the first technology that allows to do the transfers and transactions without relying on third-party processing and confirmation. The digital nature of Bitcoin is one of its most outstanding features as it allows for direct peer-to-peer transactions not possible with traditional currencies.

Bitcoin's popularity obviously comes from the fact that a third-party intermediary like in a whole banking system, is absent which in return makes one whole system free from central control (Peng, 2013). In such system, which runs based on the belief that more than majority nodes as such, general users, are honest, all transactions are traced. Those engaged in such transactions can track each transaction independently without the intervention of a middleman, and this is why any transaction involves several existing nodes for processing. These miner nodes are assigned the responsibility of two broad functions, namely, the recording of transactions and to store them in a technology called the blockchain. As a result of mining, miners are awarded for confirming and recording transactions in the blockchain. For this, utilization of in-computing is fundamental, as miners compete, and the verified process finally selects the winner for the purpose of updating the blockchain. Remember, all users own the system, each getting a personalized ledger which is to maintain in real-time.

Lately, Akbar et al. (2021) proposed a new proof of work process intending to increment decentralization and lower possibility of 51% attacks without increasing vulnerability for sybil attack that involves by sidelining identity and reputation system. From algorithmic complexity aspect, Renato and Dos identified that the complex systems theory has already classified the Bitcoin system as a chaotic turbulent and cascade types, thus raising the doubts of the possibility of chaotic phase of the system.

The benefits of operating in a blockchain based network include data transparency, getting rid of the reliance on a third party, speeding up international payments, privacy as a user, those transactions being irreversible, users not paying any transaction fees, the low cost of transactions, and minimizing theft. Cryptocurrency makes its way on more than 40 over-the-counter exchanges globally and its market value is currently

estimated at US\$ 16 billion. It is because of its entry of a large number of users of ordinary type and being related to nothing that has seen Bitcoin to rise as one of best investment choices. This led to the fact that Price anticipating tools would be crucially beneficial for investors. Given such magnitude of the issue, many experts around this field of crypto have recently research about Bitcoin price prediction.

An artificial neural network (ANN) model's performance in forecasting the price of Bitcoin was reviewed by (Almeida et al., 2015). They used the previous price and volume for one day as predictors. Nevertheless, the strength of their approach was defined with the lack of sufficient data in order to produce accurate forecast. McNally et al. (2018) developed machine learning approaches for Bitcoin price forecasting. He applied patterns including, RNN, ARIMA, and LSTM. The performance results of the RNN, ARIMA, and LSTM models were reported as 5.45 %, 53.47%, and 6.87%, respectively (Faghih Mohammadi Jalali, & Heidari, 2020). Othman et al. (2020) utilized network characteristics of the Blockchain to construct an ANN for predicting Bitcoin future prices, and they got about 55% of accuracy on average. Zhang et al. (2022) integrate a classification technique introduced by Chen et al. (2012) in the nonparametric mode to forecast price movements, which leads to a 89% accuracy of the strategy based on a Bayesian regression.

The prediction of Bitcoin prices was employed by Madan et al. (2020) using Support Vector Machine (SVM) algorithms, binomial logistic regression classifiers, and random forests with a 55% accuracy rate in the Bitcoin blockchain (Madan et al.,2020). Atsalakis et al. (2019) evaluated the determinants of Bitcoin rates along with Twitter SVM emotional analysis and discovering a positive correlation between Bitcoin prices and some factors such as the number of Wikipedia hits and hash rates. The other study carried out by Brauneis et al. (2019) was focused on correlating Bitcoin trading volumes with aggregate of the sentiments that was extracted from the twitter posts. They observed that there was a significant relationship between the BTC prices which were going to increase and the number of the daily tweets. Moreover, the study results showed that online search for the term "Bitcoin" affected the Bitcoin price (Georgoula et al., 2015).

In the proposed forecasting method, the objective is to use both the empirical mode decomposition (EMD) and support vector regression (SVR) algorithms in order to add

the intrinsic components of the signal and perform future forecasting. This method implements intrinsic component decomposition as a noise removal block at the training data training stage.

2.3.1. Factors Affecting Changes in the Price of Bitcoin. The fast spread of the cryptocurrency market spurred many researches to study Bitcoin in specific, trying to find the price influencers of such a volatile and new market. Zhou, (2021) claims that external shocks such as interest rates fluctuation and exchange rates changes have an impact on the way Bitcoin is priced. The other finding was that Gonçalves, & Alves, (2022) Bitcoin doesn't correlate with gold, fiat currencies, and major stock market indices which are traditional financial assets. Speculative, Mokni et al. (2022) mentions the fact that investor sentiment in Bitcoin plays a very important role in its short-term volatility. Agarwal et al. (2018) proposed that real-world factors like transactional utility and changes in money supply underlie long-term price dynamics of Bitcoin, and the investors' need for cryptocurrency confirming this relationship.

Besides that, Agosto et al. (2022) used the Google search term queries associated with Bitcoin to show that globally Bitcoin interest usually comes before the prices go up and the price moves down is mostly preceded by the market panic. Hakim das Neves, (2020) emphasize a significant relationship between public interest and Bitcoin prices, Pearson's correlation coefficient accurately shows how Twitter sentiment is used to predict Bitcoin price moves (Gao et al., 2021). On the same note, Saleem et al., (2024) finds that Twitter sentiments and text volume also play a significant role in Bitcoin prices. Hence, forecasting the Bitcoin price solely from these separate factors comes with substantial constraints.

Similar to any other asset price, the price of Bitcoin is determined by a wide range of factors, from internal issues within the crypto market to those that impact the world economy.

- **Supply and Demand Dynamics:** The basic economic principle of supply and demand plays a significant role in Bitcoin's price fluctuations. Bitcoin's supply is limited by its algorithm, with a maximum cap of 21 million coins. Changes in demand, driven by factors like investor sentiment, institutional adoption, and macroeconomic conditions, can affect its price.

- **Market Sentiment and Speculation:** Investor sentiment and market speculation heavily influence the price of Bitcoin. Positive news, such as regulatory developments favoring cryptocurrencies or institutional adoption, can drive up prices, while negative news, such as security breaches or regulatory crackdowns, can lead to price declines.
- **Adoption and Acceptance:** Increasing adoption of Bitcoin as a store of value, means of exchange, or investment vehicle can positively impact its price. This includes acceptance by mainstream institutions, businesses, and individuals.
- **Regulatory Environment:** Regulatory developments and government policies regarding cryptocurrencies wield substantial influence over Bitcoin's price. Favorable regulations can boost investor confidence and adoption, while unfavorable regulations can lead to uncertainty and sell-offs.
- **Technological Developments:** Advances in Bitcoin's technology, such as scalability improvements, upgrades to the underlying blockchain, or developments in the broader cryptocurrency ecosystem (e.g., decentralized finance), can influence its price by affecting its utility, security, and adoption.
- **Market Liquidity:** Bitcoin's price can be influenced by market liquidity, which refers to the ease with which it can be bought or sold without significantly affecting its price. Higher liquidity typically leads to lower price volatility, while lower liquidity can exacerbate price swings.
- **Macroeconomic Factors:** Bitcoin is sometimes viewed as a hedge against traditional financial assets and inflation. Economic events such as monetary policy decisions, inflation rates, geopolitical tensions, and currency devaluations can influence investor perceptions of Bitcoin's value relative to fiat currencies and other assets.
- **Market Manipulation:** As with any financial market, Bitcoin is susceptible to manipulation by large traders or entities. Activities such as spoofing, pump-and-dump schemes, or coordinated trading efforts can artificially inflate or deflate prices in the short term.
- **Halving Events:** Bitcoin undergoes halving events approximately every four years, reducing the rate at which new coins are created. These events historically have led to supply shortages and subsequent price increases, as seen in the past halving cycles.

- **Technical Analysis and Trading Patterns:** Traders often use technical analysis and trading patterns to make decisions, which can influence short-term price movements. Psychological factors such as support and resistance levels, moving averages, and trading volumes can affect market sentiment and price action.

It's important to understand that the price of Bitcoin is determined by a variety of complex factors, and its value can change dramatically because it is a relatively new and still developing market.

2.3.2. Bitcoin Dominance. The derivation "BTC dominance" is the assimilation of the BBC (Bitcoin's market capitalization as a percentage of the total market capitalization of Bitcoin) (Suryawijaya et al., 2023). Thus, BTP or Bitcoin Dominance Index is obtained by dividing the market cap of just Bitcoin to the market cap of all the cryptos in the world, and finally it is expressed as a percentage. For instance, if Bitcoin's market capitalization is \$1 trillion and the total cryptocurrency market capitalization is \$2 trillion, then Bitcoin dominance would be: An instance of cryptocurrency capitalization where the market cap of Bitcoin is \$1 trillion and cryptocurrency total market cap is \$2 trillion indicate that Bitcoin dominance will be:

$$BTC\ dominance = \left(\frac{Bitcoin\ market\ capitalization}{Total\ cryptocurrency\ market\ capitalization} \right) \times 100 \quad (2.1)$$

$$BTC\ dominance = \left(\frac{1\ trillion\ \$}{2\ trillion\ \$} \right) \times 100 = 50\% \quad (2.2)$$

BTC dominance is an important metric in the cryptocurrency space and is often used by investors and analysts to assess the relative strength of Bitcoin compared to other cryptocurrencies. A rising BTC dominance indicates that Bitcoin is outperforming the broader cryptocurrency market, while a falling dominance suggests that altcoins (alternative cryptocurrencies) are gaining traction relative to Bitcoin. Investors may interpret changes in BTC dominance as signals of shifting market sentiment and adjust their investment strategies accordingly. However, it's essential to consider other factors such as market trends, news events, and technological developments when analysing cryptocurrency markets.

2.4. On-chain Analysis

On-chain analysis is the process of gaining insight into the data that has been written onto the blockchain with respect to the fact that blockchain analysis though is a more general term, the former is a more specific branch of it (Jagannath et al., 2021). It is worth to mention that the whole process is hermetic and irreversible since they are inscribed on blockchain so, this is designed in such a way like that it will meet the needs of the users and open for public.

In this way, the thoughts and investment direction of all system users can be predicted. To a Bitcoin Merchant, it gives another idea of where the price might go depending on the transaction patterns, the type of Bitcoin owner making the transaction, and where the coins might move. This analysis is similar to technical analysis but is more fundamental. The data used to make the relevant analysis can be broadly classified into three different categories (Akgul et al., 2022):

- Data on the transactions is an example of where the sender and receiving addresses, the quantity of crypto transfer and the balance that is related to the certain addresses are identified and analysed.
- Data of blocks with the mention of timestamp, miner fee, and benefits.
- Basically, the rules of a smart contract are written here, and thus the system is given its business rules.

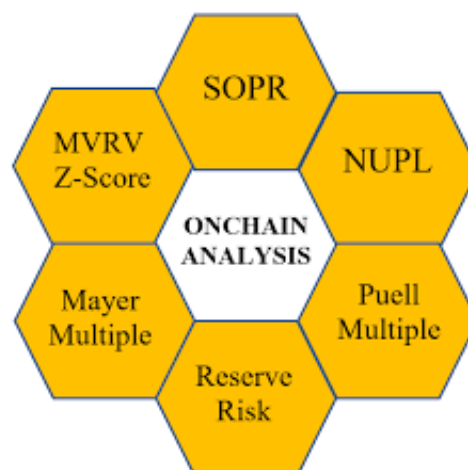


Figure 1: On-chain analysis indicators (Nava-Solis, & Treviño-Saldívar, 2021).

Looking at its history, it dates back to 2011. By comparing the value of the network with the transaction volume recorded on the blockchain, it is possible to determine when a cryptocurrency is overvalued. For this, various ratios and indicators have been developed. In this way, market sentiment analysis is made, and the amount of traded /untraded crypto money or token can be determined because whales (people or organizations who own large amounts of crypto currency) sometimes pose a great danger for the crypto money world. Popularized by Coin Metrics, Chris Burniske and Jack Tatar developed the Network Value-to-Transaction (NVT) ratio in the summer of 2017 to measure the utility value of a cryptocurrency, specifically its transactional utility. It was one of the commonly used measures for cryptocurrencies. According to the NTV ratio if the network value is high, the NTV ratio decreases. Conversely, as the value of the network decreases, the NTV ratio increases. The explanation for this is that, if the network value is extremely low, it may indicate that a more substantial price is warranted given the trading volume (Bergesen, & Palm, 2018).

The so-called UTXO (unspent transaction outputs) is a very important concept. It represents the amount of crypto money left over from each transaction. By using this, it is possible to see how long these funds are kept in the wallet. Based on the UTXO data, HODL waves are drawn. Using UTXO, the amount of unused crypto money in the wallets are determined and for how long they are kept.

Holder with an increasing number of investments means that the circulating supply is lower, that is, they do not decrease the sale and the circulating supply can said to be less. This indicator also reflects the investor psychology, allowing us to observe whether people want to sell coins or not. For example, considering investors who have started to sell the coins they have kept in their wallets for more than 1 year, Holder decreases, which means that the investor is selling his crypto asset because they are making a profit. With this indicator, investments can be made by looking at the buying/selling times. Warm colors, such as red, orange, yellow, represent Bitcoin that has been used within 1 year, while cold colors represent Bitcoin that has not been used for more than 1 year (Akgul et al., 2022).

Another indicator used in on-chain analysis is MVRV Z-SCORE. This indicator shows the period when Bitcoin is overvalued and undervalued. Certain values in the past are considered and the formula is as follows:

$$MVRVZ - SCORE = \frac{MarketCap\ USD - Realized\ Cap\ USD}{Std.Dev(Or\ Market\ Cap\ USD)}$$

(2.3)

The MVRV Z-score is a function obtained by dividing the difference between the total market value and the realized market value by the standard deviation of the market value. This shows how many standard deviations the market value differs from the realized value. Market value indicates a peak if it is significantly higher than realized value, but a bottom vice versa. In other words, it shows whether the realized value with the market value is over or undervalued than the Bitcoin price. The ultimate goal is 7. The probability of a Balloon asset above 7 is high.

The Z-Score is a number used to measure the relationship of a value with the group average of that value. It is measured with a standard deviation, and for example, Z-Score 0 means the value is the same as the group mean, and if the Z-Score is 1, that value is 1 standard deviation from the group mean. It is more important in determining extraordinary movements. If the Z-Score -Red Line enters the pink box, it means that there is an excessive increase beyond the true value. The probability of the price falling increases (Akgul et al., 2022).

2.5. Market Sentiment Analysis

Textual information can be categorized into two main types: reports, views or analysis. Fact spells as an objective statement which usually describes what happened, while opinions may include individuals' feelings in addition to their views and evaluations (Bhardwaj et al., 2015). Sentimental analysis is a thinking process that mainly deals with the evaluation of feelings expressed in the people's opinions (Liu, 2022). NLP, Computer linguistics and text analysis are among the techniques used in this process. Such study is generally employed for mining and examination of the subjective text that comes from the Web, mostly, social media. By creating an index of sensitivity or reactions of the large majority to a certain product, a person or an idea, the presentation of the polarity of information, termed as sentiment analysis, becomes apparent.

Web especially Facebook works as a primary arena where the individuals can submit knowledge and tell what they think. This trend is a useful tool to businesses

and organizations to conduct sentiment analysis, which they can use to monitor their reputation and customers' satisfaction, among others. Sentiment analysis can help businesses deepen understanding of consumer opinions/preferences regarding their products and thus pinpoint weak areas and make improvements if necessary. Moreover, organizations can use the sentiment analysis tool to understand the mood in the public and they may can foretell the election results which probably assist that goal of the candidates to make successful election campaign strategies. Thus, the increase in the use of sentiment analysis across all industries has progressed to the incorporation of companies specializing in offering the sentiment analysis services which are at around 20 to 30 in number only from the United States (Bhardwaj et al., 2015).

Moreover, recent reports show the high rate of importance business people input in the info obtained through platforms like Facebook. For example, the Cambridge Analytica, known for its participation in political campaigns with the like of Donald Trump and Ted Cruz, came dig correlation between the customs of Facebook users and their social networks, bribe them as an academic investigation. This information was deemed "extremely valuable" [Source: For example, just triple the number of women in the STEM department or in art classroom budget.

2.5.1. Volatility index (VIX). Stability is a determinative factor for many brokers and investors in financial systems, amongst other things. Hurricane chart measures the degree of the fluctuation of returns over a certain period according to the annualized standard deviation of daily returns, calculated for a specified number of the previous trading days before, measured as a yearly rate. Implied volatility is the value derived from the observation of current option contracts' prices and this voluntarily represents the market's expectations concerning future movements in volatility (Kiapekos, 2021).

Investors will use historical volatility in order to interpret past market momentum to analyze the current market. However, they also accept volatility indexes like VIX to go real-time on their search for oscillations in market opinion as well as future estimates of volatility. In 1993 the CBOE® (now affected by Chicago Board Options Exchange) started creating of the VIX as a measure of fear. It conducts 30 days volatility. First, composed to understand the volatility of S&P 100 Index by using

options price options by the money, the VIX has been then quickly accepted as the important indicator of US stocks market volatility (Rashid et al., 2019).

By the year 2003, the VIX (volatility index) underwent revisions aimed at boosting accuracy for volatility traders profiting and hedgers. After that, the VIX has been constructed by using options on the S&P 500 (SPX), the main index for the US equities. The VIX comprises the weighted average prices of SPX put and call options which are calculated across diverse strike ranges. This strategy not only created an instrument for mimicking a volatility using a collection of SPX options but also it made the VIX from theory to a standard of practice - playing, hazing and pledge process (Yoldas Ayan, 2023).

VIX, the acronym for volatility index, is a futures contract, thus the markets use the spot VIX, which is a representation of a future volatility time which can be derived from SPX options. The VIX, which is updated every 15 minutes until 5 p.m., by using the option prices of up to 20 stocks in the S&P 500 to evaluate the expected realized volatility 30 days ahead, employs the option strip of SPX to measure and assess the daily changes in the S&P 500. Finally, the problem of SPX options basket is its dynamic demonstration, which in turn is caused by the costs of the system and impracticality in the long term. Cboe has Upgraded the VIX Made Schme VIX Energy in 2014 by entering SPX weekly options Expirations in its formulation. Among all the option expires on a Friday, which incorporate both usual and weekly types, only SPX options get into account of VIX. Non-Friday expirations are not part of VIX compounding formula (Alexander, & Imeraj, 2019).

This move to the weekly options of SPX index accommodates the calculation of the VIX in option series with the same period length for the expected volatility in the next 30 days. To do this, the VIX considers SPX Options with maturity dates anywhere from over 23 days to less than 37 days and extracts an interpolation in between two of the points in the S&P volatility term structure.

We know that there are lots of research that discuss the VIX parameters for illustrative purposes, e.g. the valuing of volatility components of derivatives or predicting directional movements of the stock indices such as S&P500 (as reflected by Dotsis et al (2007)). Yet, we can infer that only a few investigations deal with the direct forecasting of implied vol One of the hybrid models proposed by Ahoniemi (2006)

incorporates both an ARIMA and GARCH to make point forecasts for the VIX index. Konstantinidi et al. (2008) on the other hand introduce the forecasting capabilities of the economic variables model, vector autoregression (VAR) model and finally the autoregressive fractionally integrated moving average (ARFIMA).

The outcomes of both studies demonstrate that the coming models make a better explanation for the S&P500 and VIX volatility indices, and their empirical out-of-sample forecasts are used for trading purposes. Hedging is positioned to be performed on an equity index by Clements and Fuller (2012) via the use of semi-parametric forecasts that are intended to cover increases in VIX. On the other hand, Fernandes et al. (2014) employ the heterogeneous autoregressive (HAR) process proposed by Corsi (2009) in order to model the VIX (Chicago Board Options Exchange Volatility Index) but involving a number of macroeconomic variables from the US.

The rationale behind the HAR process adoption is that the long memory observed in both implied and realized volatility of the options share similar properties (Koopman et al., 2005, Bandi and Perron, 2006; Corsi, 2008). What is more, Fernandes et al. (2014) present an HAR model based on a neural network and assume there are some nonlinear anomalies of unknown shape which in this model are considered. They decide on this because giving a consideration that some macro-finance variables, for instance, the dollar index cannot have a statistically important relation with VIX after including non-linear associations. Nevertheless, however they are designed, their strength of contribution to the index remains within a linear line of thought.

2.5.2. Crypto Fear & Greed Index. Alternative.me was among the first websites that present the fear and greed index (FGI) in February 2018 to investors. The indicator, characterizing the Bitcoin market closely, became popular among investors in a short time. Fiscal Fundamental Indicator (FFI) has a range from 0 to 100, which depicts whether investors are moved by their fear or greed during a particular time. The lowest point of the scale means the flee stage when investors are active, while the highest value is the greed level at the wave top (Łęć et al., 2022).

Meanwhile the FGI looks beyond the purely sentiment measures and for the purpose of gauging the investors' sentiments it captures a more comprehensive picture. Also, the FGI uses public interest as registered in social media, and along with this, it

uses investment behavior metrics such as volatility, market momentum, and volume, as well as user interest from analysis of surveys, dominance, and trends. The market situation created by the FGI has attracted much attention of cryptocurrency market participants and the indicator is the central focus of many studies undertaken in the field of cryptocurrency research. For instance, during the pandemic which led to the onset of COVID, Güler undertook an emended EGARCH model, which showed that the FGI had a positive impact on Bitcoin returns and volatility Furthermore, she points out that alterations in both decent and indecent opinions lead Bitcoins to experience such returns.

Some as well examine the since correlation between crypto and FGI. Mokni et al. (2022) and Gaies et al. (2023) studied the causality effects between Bitcoin and the FGI, while Bourgehelle (2022) unigveked this collective sentiment to Bitcoin's volatility. The other more recent study (Lin et al., 2023) focus on how FGI and market risk interact in four digital currencies. Another element of this research as accentuated by Johnson et al. (2024) is the price and quantity of trading of five cryptocurrencies in relation to the Financial Stability Index. Notwithstanding, the aforementioned article certainly links this sentiment index to both cryptocurrencies yields or volatility, by means of conducting empirical analyses over short sample periods, mostly because of the data limitation.

2.6. Technical Indicators

The Technical indicators, RSI, EMA and SMA are widely used in stock market analysis for generating trading signals and understanding the behavior of the market. These fluctuating indicators are vitally important in detecting reliably the direction and possible price corrections. The incorporation of these KPIs into ML models represents a huge step and points to the move away from pure prices in research. With by making use of them, models can identify subtle relations of the indicators with prices and dust improvements of their performance.

The underlying reason behind the incorporation of these four technical indicators as input features in this thesis is their noble position, wide use, and demonstrated performance – regardless of the market conditions. In addition, they are the easiest to calculate and analyse owing to the novice trader's availability. These indicators are

applicable for not one but all kinds of both short and long-term research based on the duration (Dongrey, 2022).

2.6.1. Relative Strength Index. Technical analysis is the tool that rests on the basis of RSI indicator to measure the strength of the price movements in different assets (Rodríguez-González et al., 2011). Find the RSI, by dividing trading range by average range. It creates the sense of these alterations and allows you to control the pattern if necessary. Write an article about public health and the effects of air pollution. Discuss issues like higher rates of respiratory diseases, cardiovascular ones, and upgraded risk of cancer, and back up with associated facts. Among the possible uses of the RSI by the traders and the analysts is to recognize the stock price run up or run down and, what is more, predict the corrections or reversals as well as the price.

RSI integration as only one of the exported neural-based models may make the model able to forecast more accurately by (Zhao et al., 2019). By including RSI in my feature set, I ramp up the model's capacity of information utilization, thereby improving the model's forecasting accuracy.

The RSI is obtained by a formula that analyzes current increase compared to previous one within the time frame selected. The RSI calculation formula used in this thesis is as follows: This thesis utilizes the RSI calculation formula below:

$$RSI = 100 - \left(\frac{100}{1 + RS} \right) \quad (2.4)$$

Where RS (Relative Strength) is Average Gain by Average Loss.

2.6.2. Simple Moving Average. The Simple Moving Average (SMA) shows the price of a security or property over a time span calculated by adding the prices of a given set of time periods and dividing by the number of units for this time interval. A possibility of providing the SMA imputations might unveil the price shift of the asset prevailed during the specified period of time. Such would actually help to identify both long and short-term dependencies which in turn would further improve the accuracy of the proposed machine learning model.

To figure SMA, we simply include within our calculations a particular number of prices over a limited duration and then take a sum of each value and divide that number by the number of periods (Hansun, 2013). The SMA calculation formula

utilized in this thesis is as follows: In this treatment of the subject, the SMA calculation formula represents the SMA formula as thus:

$$SMA = \frac{\sum_{i=1}^N P_i}{N} \quad (2.5)$$

Where N is the number of periods used in the calculation, P_i denotes the price of the asset at period i.

2.6.3. Exponential Moving Average. The Exponentially Moving Average (EMA) happens to be one of the weighted moving averages, whose value is more related to the recent price, instead of the old ones (Nakano et al., 2019). Nevertheless, EMA is slightly different from SMA that each data point is assigned the same weight. The EMA is determined by adding the weighted average of the current days opening price and the previous days EMA value, which is multiplied by the smoothing factor. So, smoothing factor refers to the previous EMA constant. This weighting factor is calculated on each and every instance.

The two types of investors who consider long term and short term as well as volatile traders can make use of EMA. EMA ascertains the tasks through the ability of detecting a price movement of considerable value and also capturing reversals of the market; therefore, it is easy to use for both short-term and long-term forecasting (Menezes et al., 2023).

The calculation of EMA involves two key components: within the momentum indicator calculation, there is a buy back indicator, calculated by EMA and the prior EMA. The smoothing factor is what gives the most recent observation the greatest weight and usually it is based on either a certain time period or a specific number of observations. Nevertheless, the distinction is that EMA gives the advantage of recent data in the case, and also it has rapid reactivity to the price changes amongst all the indicators whereas SMA does not have these characteristics (Menezes et al., 2023).

The EMA calculation formula used in this thesis is as follows: The formula used in this thesis to calculate the EMA is the following one:

$$EMA = (Price - EMA_{previous}) * Smoothing Factor + EMA_{previous} \quad (2.6)$$

Where $EMA_{previous}$ is the EMA value from the previous calculation "Smoothing Factor" is a constant that determines the weightage given to the most recent data point.

2.7.Related Work

There are many studies and methods for determining the future price of a security in traditional financial markets. However, after a brief introduction to the research topic in the first part, the second part includes a literature review, followed by a basic and technical analysis, the studies on chaos theory, on-chain analysis, and sentiment analysis.

Hacinliyan and Kandiran (2015) investigated the possible fractal behaviors in the Istanbul Stock Exchange indices within the scope of chaos theory. To observe whether there were any chaotic and fractal behaviors, Higuchi and Katz methods were used to analyze the monofractal behavior of the selected indices while the Transformed Width (R/S) and Adjusted Fluctuation (DFA) analyses were used to examine the chaotic behavior. As a result, it was concluded that there is chaotic behavior in the relevant indices.

Alpar and Özge (2016) used the data between 1988-2004 in order to prove the existence of chaos theory in the stock market. Using the Lyapunov exponential model, the result was found to be 18. The analyzed time series have showed stock market regularity, in which the direction of stock price tendency is caused the movements of observed in previous 2 days repeating themselves 4 days later. In 2018 Biswas et al. developed a clear and easy understanding of chaos theory in relation to scientific domains and the methods to use in prevention of chaos in these areas. The study defined the area of research, involved appropriate literature reviews, and made observations that corresponded with chaos theory.

Abraham et al. (2018) had done sharp work, integrating Tweeter and Google trends information, so as to understand the price patterns of both Ethereum and Bitcoin. They used sentiment analysis with VADER (a dictionary that evaluates the valence of words for sentiment reasoning) to process the obtained data. VADER in particular was selected to perform sentiment analysis because not only it is able to assign a text a positive, negative or a neutral sentiment but also to quantify how the language intensity is. On the other hand, the research team did not include sentiment analysis of tweets in their studies as they believe that it might be an inaccurate index during the time of cryptocurrency price drops. Conversely, tweet volume was more of a focus on.

The research showed that a correlation of high order was present regarding both Google Trends and tweet volume with the value of cryptocurrencies. Strikingly, this relation persisted across all the cases, even the ones corresponding periods of price changes, illustrating the power against high variance and nonlinearity. Direct contrasts were made using a linear model, because although the terms in the input exhibited some non-linearity, they were similar to the response variable for most parts. The study concluded that there was an existence of such a connection that was either directly caused or influenced by the fact that their observation happened around the time that the price of all cryptocurrencies was on the rise.

Additionally, the study was carried out using Twitter data to detect the general tone of the cryptocurrency's users, the positive sentiment held even if the prices would go down unexpectedly. the author also spotted relationship between the cryptocurrency prices and google searches. This relation was illustrated by price fluctuations that conformed to the model line.

For the chaos and randomness degree evaluation in Bitcoin's marketplace price and returns dynamics I did the Lyapunov spectrum, Shannon entropy, and generalized Hurst's exponent models. Peculiarly, both returns and prices in the mentioned study displayed non-linear dynamics though the latter could not be predictable based on the observations till date. The analysis focused on three digital currencies: Bitcoin. Digital Cash. Ripple, summing up that all these cryptocurrencies had fractal dynamics, long memory, and self-thoroughness.

Jain et al. (2018) sought to examine whether social variables can help predict cryptocurrency rates. They did that using the Litecoin and Bitcoin prices two hours in the future which were predicted based on the real-time sentiment given in tweets. For their research purpose, they did use a Multiple Linear Regression (MLR) model to model an average price at 2 hours from accumulated positive, neutral and negative tweets between March 1-11, 2018.

Among the number of other works that was carried out is that of Bouri et al. (2019) whose study dealt with the role of herd behavior in the cryptocurrency market. The research supplies the necessary evidence allowing to claim that the herd behavior was

present in this market. The result of a logistic regression analysis revealed that when uncertainty was higher, people are more likely to display a tendency of herd behavior.

To understand the relationship between Bitcoin and Ethereum news and the price prediction, Vo et al. (2019) conducted an analysis based on the assumption that there is a relationship between the mood of the public and the cryptocurrency market, like traditional financial markets. Data were obtained using daily time series data from July 30, 2017, to October 5, 2018. For the sensitivity analysis, the news of the last 7 days was collected. An algorithm was created with a dictionary-based approach, with the thought of distinguishing the effect of the news about cryptocurrencies on the price movement, and with the thought that it can provide investors with a buying and selling advantage. They created a model that can directly predict the price direction by specifying whether to buy, sell or hold, and they concluded that the model they created correctly predicts the price movement in two cryptocurrencies.

Holiachenko et al. (2022) aimed to explain cryptocurrencies using fundamental and technical analysis. A technical analysis method was developed with different currencies in the same stock market to achieve maximum gain. In the conclusion part of the study, where many scenarios and graphics related to arbitrage were created, it was stated that the developed method was resistant to fluctuations in oil and gold prices.

Hudson and Urquhart (2021) conducted a technical analysis of Bitcoin, Ethereum, Bitstamp and Litecoin in their study and although predictability continues in other cryptocurrency markets, it has been concluded that there is no predictability for Bitcoin in the out-of-sample period. In order to test the validity of the efficient market hypothesis in the crypto money market, Kang et al. (2019) investigated the random walk theory. They concluded that among crypto exchanges established before November 2017, large exchanges are more likely to satisfy weak and semi-strong forms of market efficiency.

Pietrych et al. (2021) investigated the existence of chaos theory in cryptocurrencies. The nonlinearity and chaos in cryptocurrencies (Bitcoin, Ethereum, Ripple and Litecoin) were tested and concluded that these time series show strong

evidence supporting the hypothesis that these time series come from an unknown production process that behaves nonlinearly and chaotically.

In their study carried out in 2021, Yue et al. employed the method of bibliometric analysis to find out the effect of cryptocurrency on the economy. The research targeted literature on the issue giving an overview of the crypto-economics interfacing now and then. Bibliometric analysis aims to reveal previously unknown patterns by collecting a large number of relevant information in a specific area or a specific area of the subject. In this context, starting from the keywords in the literature, using the CiteSpace 5.7. the research outline of the economic effects of the study was created. Data obtained from the literature studies were clustered according to events that could affect the value of Bitcoin and its effects on the economy, and made available for analysis. The results obtained for direct economic effects are as follows:

- While the impact of speculative trading on cryptocurrency differs in different studies, Bitcoin price fluctuations will affect macroeconomic policy to some extent.
- The hedging and safe-haven characteristics of cryptocurrencies continue to change over different periods of research. Some research finds that Bitcoin can act as a speculative asset.
- External events such as the Covid pandemic can increase price fluctuations and improve the hedge effectiveness of cryptocurrencies.

Another valuable result of the study is as follows: While computer science literature and interdisciplinary fields are concerned with the economic phenomenon caused by technologies, it has been concluded that previous researches are more concerned with economic results in economic fields. In another study, Gozde (2021) conducted sentiment analysis on tweets connected with "Bitcoin" on Twitter. By Orange Data Mining, they discovered there is an upbeat feeling related to Bitcoin generally, with users conveying 'happy' feeling when they get Bitcoin. Gurrib and Kamalov (2021) tried fluctuations in the crypto money market to be predicted by using sentiment analysis. It was determined that tweets about crypto money in a 3-week period were directly related to the market value of coins, and the model used in this study was determined to be critical. It is emphasized that it is a good model for predicting the value of PTO (Paid Time Off) money.

Jagannath et al. (2021) conducted an on-chain analysis study in order not to predict Ethereum prices. The selection of Ethereum for the present case study is based on its popularity as the second most famous cryptocurrency (after Bitcoin) and its remaining workings of an open network structure. We ran data collection from the Ethereum blockchain's public domain and APIs of various online applications. The study investigates the association of no key Ethereum on-chain metrics with price by use of the Pearson and Spearman correlation coefficients. On top of that, the study comprises a model analysis in which machine learning models will be applied to forecast Ethereum price.

In the author's study, it is shown that the network uses the Long Short-Term Memory (LSTM) neural network model with self-adaptation features such as three hyperparameter optimization techniques for Ethereum price forecasting. In addition, the LSTM model vaunting self-adaptive algorithms outperform the traditional LSTM models in giving speedier and more precise price predictions of Ethereum.

While Gu et al. (2022), in their own study, did another On-Chain analysis and checked transaction amounts that were unusual on cryptocurrency exchanges. To correlate the results, the most substantial factors influencing trading volumes in diverse exchanges were analyzed. Moreover, a model for predicting was designed, which allows to take into account the most influencing factors on the total trading volume. The case study on the detected anomalies, raised the fact that the activities held, policy changes, and legal issues had correlation. Further, the scale included determining a difference between real and estimated transactions through the application of deep learning methods to assess the abnormal transaction amounts.

El Montasser et al. (2022) calculated the closing prices of different cryptocurrencies in his study. The results of the authors who made balloon studies in terms of market size were interpreted by comparing them with the Covid 19 period. As a result of the study, a high correlation was found between the Covid 19 period and the crypto market. The study analyzes crypto bubbles and COVID-19 pandemic influence on the market efficiency during the last year of extraordinary events. The paper discusses the evolution of market anomalies throughout the period. Outcomes evidencing tangible changes in the trading activity of the most traded cryptocurrencies which were directly linked to the declaration of the COVID-19 pandemic. Three key

moments in the history of cryptocurrency were identified: the end of 2017, the beginning of 2018, and the period around July 2020. The study claims that these crypto-bubbles, though they are the essence of the distributed finance, just a tiny part of the market efficiency. Study is supposed to be the core content of research and here researcher must evaluate the usefulness of cryptocurrencies and look forward to price bubbles.

Nie (2022) applied the network method for discovering active (critical) events in correlation dynamics of the cryptocurrency, new sequence around January 6, 2021 was provided to demonstrate the changes of local and dramatic correlation network structures. The application of this methodology helps the recognition of the developing markets' dynamics as well as in research on underlying instability/fragility of the correlated assets in the cryptocurrency market. The initial analysis is based on the vector multiplication of the powering of the leading crypto industry system (IS). The theoretical structure of the markovian jumps indicates market volatility before and after different critical events. It is worth noting that correlations observed between the dynamics of the correlation matrix and market conditions play a key role in clarifying the nature of jumps associated with difficult predictable positions. The association between correlation values and network structure remains to be seen additionally, implying a positive correlation between these two metrics.

Wasiuzzaman et al. (2022) conducted research which aimed to measure the performance of IGO (Initial game offering) cryptocurrencies against the backdrop of 2020 bear cryptocurrency market. The investigation involved the acquisition of the pricing data for the three Islamic gold-backed digital currencies such as OneGram, HelloGold and X8X as also the traditional gold-backed cryptocurrency PaxGold and Bitcoin, the traditional cryptocurrency backed with fiat, during the period December 2019 to November 2020. Study using Autoregressive Moving Average Generalized Exponential GARCH (ARMA-G-EGARCH) models showed that in the bear market, returns for all cryptocurrencies decreased, most noticeably for the Islamic gold-backed cryptocurrency. On the other hand, we see that the volatility went up for all five cryptocurrencies and more significantly so only for the traditional ones.

Rani et al. (2022) observed that Long Short-Term Memory (LSTM) models demonstrated higher accuracy compared to Linear Regression in market trend

predictions when analyzing features such as open, close, high, and low prices. However, they cautioned that these findings might be subject to change upon the consideration of more comprehensive variables. The study highlighted the dynamic nature of the cryptocurrency market, noting its susceptibility to influence from social media and other external factors, thereby challenging the sufficiency of datasets for predictive accuracy. Rani et al. emphasized the potential for improved outcomes through the integration of new data facilitated by technological advancements, suggesting a continuous evolution in research methodologies and findings in this domain.

Ferdiansyah et al. (2019) demonstrated that their proposed model, leveraging time series analysis techniques, successfully predicts Bitcoin prices using data from the Yahoo Finance stock market. The model, which divides data into training and testing segments as detailed in their study, has shown capability in forecasting future price movements. However, they noted a limitation regarding the model's precision, as indicated by Root Mean Square Error (RMSE) values that are less than ideal—potentially under one hundred or approaching a score of 50 RMSE. This limitation underscores the inherent difficulty of predicting stock market behaviors, which are subject to a wide array of uncertain factors including political and economic issues on both local and global scales.

The authors acknowledge that the unpredictability of such factors makes relying solely on LSTM (Long Short-Term Memory) networks for Bitcoin investment decisions insufficient. They propose future research directions aimed at enhancing the model's accuracy. These include modifications to the LSTM architecture, such as adjusting the number of layers, incorporating dropout techniques, altering the epoch count, and testing the model against more volatile datasets. Additionally, they suggest integrating sentiment analysis with LSTM methods to evaluate the impact of market uncertainties on Bitcoin's value more effectively.

Akgul et al. (2022) elucidate that On-Chain analysis necessitates the consideration of diverse metrics for the insightful examination of Bitcoin and other cryptocurrencies. This methodology is often synergized with technical analysis by traders and investors to ascertain optimal entry and exit points for crypto assets in the short term. Distinct from speculative approaches, On-Chain analysis enhances the

precision of targeting for either short or long positions through the systematic interpretation of metrics and analyses, which are consistently shared. It parallels technical analysis in its foundation but differentiates by exclusively analyzing transactions recorded on the blockchain.

Furthermore, On-Chain analysis introduces specific indicators, much like those found in technical analysis, to offer a current snapshot of the market and forecast future trends. On-Chain sentiment analysis, in particular, focuses on the interpretation of cryptocurrency transfers on the blockchain. These innovative analytical techniques are posited by Akgul et al. to potentially provide the most reliable price predictions, reflecting the inherent characteristics of the cryptocurrency market.

Overall, On-Chain analysis presents a compelling instrument for analysts to delve into real-time blockchain network insights, thereby affording them the opportunity to navigate a more transparent and data-enriched cryptocurrency market landscape.

In the analysis of the three articles, a consensus emerges on the importance of incorporating external factors to enhance predictive accuracy in cryptocurrency price models. One article highlights the superiority of Long Short-Term Memory (LSTM) models over Linear Regression in terms of accuracy. In contrast, another article raises doubts about the effectiveness of LSTM models for Bitcoin price prediction, pointing out the complexity and unpredictability of the cryptocurrency market influenced by a myriad of uncertain external factors such as political and economic events at both local and global levels.

Furthermore, the discussion suggests that while LSTM models have demonstrated potential in forecasting price movements, their performance, as indicated by Root Mean Square Error (RMSE) metrics, suggests room for improvement. The inherent limitations of these models, as underscored by the articles, reflect the dynamic and multifaceted nature of cryptocurrency markets, which are affected by factors beyond historical price data.

To synthesize the viewpoints presented in these articles: it is acknowledged that for enhanced accuracy in cryptocurrency price predictions, models must account for external variables that influence market dynamics. Although LSTM models offer a

methodological advancement over traditional Linear Regression, their efficacy as standalone tools for investment decision-making in the Bitcoin market is questioned. Future research directions propose modifications to LSTM architecture and the integration of sentiment analysis to better capture the impact of external uncertainties on cryptocurrency values.



Chapter 3

Methodology

3.1.Introduction

In the pursuit of advancing our understanding of cryptocurrency price dynamics, this thesis endeavours to construct a predictive model for Bitcoin prices, leveraging the statistical prowess of Ridge Regression. Bitcoin, as the progenitor and most widely recognized cryptocurrency, presents a fascinating subject for analysis due to its volatile price movements and significant impact on the global financial landscape. The primary objective of this study is to harness historical price data and various market indicators to forecast future price movements, thereby providing insights that could be invaluable for investors, policymakers, and scholars alike. This introductory section outlines the data collection process and the methodological framework adopted in the study, setting the stage for a detailed exploration of the predictive capabilities of Ridge Regression in the context of Bitcoin price prediction.

The methodology section delves into two crucial components: data acquisition and preprocessing, followed by the application of Ridge Regression for predictive analysis. Initially, a comprehensive dataset is compiled, encompassing daily Bitcoin prices, trading volumes, and other relevant market indicators over a specified period. This dataset is subjected to rigorous preprocessing to ensure data quality and relevance, including handling missing values, normalization, and feature selection to enhance the model's predictive accuracy. Subsequently, the study employs Ridge Regression, a variant of linear regression known for its effectiveness in addressing multicollinearity and overfitting by introducing a penalty term to the loss function. The choice of Ridge Regression is motivated by its ability to handle the high volatility and noise inherent in cryptocurrency markets, providing a more stable and robust predictive model. This section will detail the selection criteria for model parameters, the training and testing process, and the evaluation metrics used to assess the model's performance, laying the groundwork for the empirical analysis that follows.

3.2. Research Design

The research design adopted in this thesis is a quantitative approach aimed at establishing a predictive model for Bitcoin prices using Ridge Regression. This design is meticulously crafted to navigate the complexities of financial datasets, specifically the volatile nature of cryptocurrency markets. The model's development is structured in a phased manner, starting with the collection of a robust dataset that includes not only historical Bitcoin prices but also a range of financial indicators considered influential in price determination. This dataset then undergoes a comprehensive pre-processing phase to ensure data integrity and relevance, followed by the division into training and testing sets to facilitate a rigorous evaluation of the model's predictive accuracy. The core of the research design revolves around the iterative optimization of the Ridge Regression model, meticulously tuning hyperparameters to strike a balance between bias and variance, thereby enhancing the model's generalizability to unseen data. This quantitative framework is not only geared towards achieving high predictive accuracy but also aims to shed light on the relative importance of various predictors, offering insights into the underlying dynamics that drive Bitcoin prices.

3.3. Research Philosophy

The research philosophy underpinning this thesis is rooted in positivism, reflecting a belief in the empirical observation and quantification of phenomena to derive meaningful insights. This philosophy is particularly well-suited to the study of financial markets, including cryptocurrencies like Bitcoin, where data-driven approaches can uncover patterns and relationships that are not immediately apparent. By adhering to a positivist framework, this research assumes that the cryptocurrency market operates within a structured reality governed by underlying laws and patterns that can be discovered through systematic analysis and modelling.

This positivist stance is operationalized through the use of Ridge Regression, a quantitative method that allows for the empirical testing of hypotheses concerning the determinants of Bitcoin prices. The reliance on statistical techniques to analyse historical price data and market indicators embodies the objectivist belief that knowledge is best gained through observable and measurable facts. Accordingly, this research philosophy emphasizes the importance of rigorous data collection, pre-

processing, and validation methods to ensure the reliability and validity of findings. Through this empirical and objective approach, the study aims to contribute to the body of knowledge on cryptocurrency markets by providing evidence-based insights into the factors that influence Bitcoin prices, thereby offering a model that stakeholders can rely on for decision-making purposes.

3.4. Research Type

This thesis predominantly employs a deductive research approach. Deductive reasoning starts with a hypothesis derived from existing theories or knowledge and then tests this hypothesis through empirical observation and analysis. In the context of predicting Bitcoin prices using Ridge Regression, the study begins with theoretical assumptions about factors believed to influence Bitcoin prices, such as market demand, investor sentiment, regulatory announcements, and macroeconomic indicators. These assumptions are informed by economic theory, previous empirical findings, and the theoretical underpinnings of financial markets' behaviour.

The deductive method is evident in the structured research design, which systematically tests these predefined hypotheses against the collected data on Bitcoin prices and relevant market indicators. The application of Ridge Regression serves as a tool to empirically assess the validity of these hypotheses, examining the strength and nature of the relationships between independent variables (predictors) and the dependent variable (Bitcoin price). The study's aim is to either confirm or refute the initial hypotheses based on the statistical evidence derived from the model's output, adhering to the principles of deductive reasoning.

This approach aligns with the positivist research philosophy underlying the study, which prioritizes the use of quantitative methods and empirical data to test and validate theories. By adopting a deductive approach, the research seeks to contribute to the body of knowledge in financial economics by providing empirical evidence on the determinants of cryptocurrency prices, thereby enhancing our understanding of digital asset markets through a theory-driven analysis.

3.5. Research Strategy

The research strategy for this thesis on predicting Bitcoin prices through Ridge Regression encompasses a series of systematic steps designed to ensure a rigorous, transparent, and replicable study. This strategy integrates elements of quantitative research with the deductive approach outlined earlier, focusing on empirical data analysis to test theoretical hypotheses. The following outlines the key components of the research strategy:

- Literature Review to identify existing theories and empirical studies related to financial market predictions, with a focus on cryptocurrencies and, more specifically, Bitcoin. This step helped in formulating hypotheses and selecting relevant predictors for Bitcoin prices.
- Hypothesis Formulation to establish clear, testable hypotheses based on the literature review, focusing on potential determinants of Bitcoin prices such as market demand, investor sentiment, and macroeconomic indicators.
- Data Collection to compile a robust dataset that includes historical Bitcoin prices and potential predictor variables over a defined period.
- Data Pre-processing to prepare the dataset for analysis, ensuring it is clean, normalized, and suitable for the application of Ridge Regression.
- Model Development and Testing to apply Ridge Regression to the pre-processed data, optimizing the model to predict Bitcoin prices accurately. Dividing the dataset into training and testing sets and using the training set to develop the Ridge Regression model, carefully tuning hyperparameters to manage multicollinearity and overfitting.
- Analysis and Interpretation part to analyze the results obtained from the model, focusing on the significance and impact of different predictors on Bitcoin prices. Interpreting the model coefficients, assess the model's overall fit, and evaluate the predictive power of individual variables.
- Compare findings with existing literature to validate or challenge theoretical assumptions.
- Documentation and Presentation for documenting the research process and findings comprehensively, presenting the implications of the study for investors, policymakers, and scholars. Prepared a detailed thesis report that

includes an introduction, methodology, results, discussion, and conclusions. Highlight practical implications and suggest areas for future research.

This research strategy aims to ensure that the study is grounded in solid theoretical foundations, employs rigorous quantitative methods, and yields insights that are both scientifically robust and practically relevant to the field of cryptocurrency market analysis.

3.6. Time Horizon

Implementing a longitudinal design involves observing the same subjects over a period of time, potentially ranging from a few years to several decades, to identify patterns, changes, and causality in behaviors, conditions, or outcomes. Enabling researchers to discern how variables change over time, offering insights into long-term effects, developmental trends, and the dynamics of change. The time horizon in a longitudinal study is critical, as it must be sufficient to capture the relevant changes in the population or phenomena under investigation.

3.7. Sampling Strategy

The data was sourced from Glassnode, one of the most reputable firms in on-chain analysis, along with CoinMarketCap, Yahoo Finance, and Alternative.me, through a non-probability sampling method, ensuring a comprehensive and trustworthy dataset for analysis.

3.8. Data and Data Collection

The data collection method for this study incorporates a targeted, non-probability sampling approach, specifically designed to gather quantitative and analytical data from leading, authoritative sources within the financial and cryptocurrency sectors. By harnessing detailed quantitative data from Glassnode, renowned for its comprehensive on-chain analytics, alongside CoinMarketCap for real-time market capitalization figures, Yahoo Finance for a wide array of financial indicators, and Alternative.me. The methodology ensures an extensive, data-driven foundation for the analysis. This deliberate selection of sources enables the collection of precise, numerical data critical for statistical analysis and predictive modeling, facilitating a robust, analytical examination of cryptocurrency market trends and behaviors. The emphasis on

quantitative data ensures that the study is grounded in measurable, objective evidence, providing a solid basis for drawing reliable, insightful conclusions about the dynamics at play within the cryptocurrency market.

The dataset utilized for this study comprises daily observations of Bitcoin prices along with various on-chain metrics, market sentiment indices, technical indicators, market analysis indices and economic metrics that potentially influence Bitcoin price. The dataset spans from December 31, 2019, to March 31, 2023, on daily basis encompassing a total of 1189 observations. Categories below presents delineates the variables encompassed within the dataset, categorizing each according to its respective domain.

3.8.1. Dependent variable. The dependent variable in this study is the daily closing price of Bitcoin (BTC), while the independent variables include the aforementioned indices. The data were sourced from various reputable crypto analytics platforms as Glassnode, Alternativeme, CoinMarketCap and YahooFinance ensuring a comprehensive view of the factors affecting Bitcoin's price dynamics.

Predicting the direct price of Bitcoin, as opposed to its volatility, particularly for investors, financial analysts, and the wider public engaged in cryptocurrency. Here are some advantages and unique aspects of choosing to forecast Bitcoin's price directly:

- **Clear Investment Guidance:** Direct price forecasts allow for straightforward investment decisions. By having a potential future price, investors can tailor their strategies for buying, holding, or selling, aiming to optimize returns or minimize losses without needing to interpret volatility measures.
- **Ease of Understanding for Broader Audiences:** Price predictions are more accessible and easier to grasp for most people compared to volatility forecasts. This accessibility can democratize Bitcoin investment, making it appealing to a wider range of investors, not just seasoned traders or finance professionals.
- **More Direct Application in Risk Strategies:** While volatility forecasts offer insights into market risk, direct price predictions can be immediately applied in setting up risk management strategies, such as stop-loss and take-profit

orders, offering investors a practical tool for controlling their exposure to market movements.

- **Immediate Relevance to Portfolio Management:** For diversified portfolios including Bitcoin, knowing its future price can directly impact decisions on asset allocation and help in evaluating the portfolio's future value, assisting in strategic planning and management.
- **Indication of Market Sentiment:** Price trajectories can reflect broader market sentiments towards Bitcoin. A prediction of price increase suggests a bullish market sentiment, which can be a valuable indicator for investors and analysts alike, providing context beyond the numbers.
- **Utility in Comparative Analysis:** Price predictions allow for straightforward comparisons with other investment assets, facilitating benchmarking and performance evaluation across different asset classes, enhancing Bitcoin's integration into traditional financial analysis frameworks.
- **Integration into Financial Models:** Direct price forecasting makes it possible to incorporate Bitcoin into established financial theories and models, such as those used in portfolio management, thus broadening the understanding of Bitcoin's role and impact within the larger financial ecosystem.
- **Encouragement for Adoption as Payment:** Predicting Bitcoin's price can also promote its acceptance as a form of payment. Predictive insights into its future value may reduce hesitancy associated with its volatility, encouraging more businesses and individuals to transact using Bitcoin.

3.8.2. Independent variables

- **Economic index:** Global Economic Uncertainty Index (GEPU) measures uncertainty in economic policy, which can impact financial markets and investment decisions. Including GEPU in this study allows for assessing external economic influences on Bitcoin's price.
- **Market analysis indicators:** Bitcoin (BTC) dominance is a metric that quantifies Bitcoin's share of the total market capitalization of all cryptocurrencies. Calculated by dividing Bitcoin's market capitalization by the total market capitalization of all cryptocurrencies and multiplying by 100, it provides a percentage indicating the extent of Bitcoin's market value relative

to the entire cryptocurrency market. An elevated Bitcoin dominance metric typically reflects a consolidation of market capital into Bitcoin, which is often regarded as a more secure asset amidst periods of economic instability or high market fluctuations. On the other hand, a reduction in Bitcoin's market dominance could signify a growing trust in alternative cryptocurrencies, hinting at a broader spread of capital allocation across the cryptocurrency landscape.

- **Market sentiment metrics:** The Volatility Index (VIX) is commonly recognized as a measure that gauges the stock market's expectation of volatility over the near term. Originating from option prices on the S&P 500 Index, it is often referenced as a barometer for investor sentiment and market stress.

In parallel to the VIX in traditional markets, the cryptocurrency sector has devised the Crypto Fear and Greed Index. This index is an aggregation of data points that are intended to represent the general sentiments within the cryptocurrency market. It assimilates various sources of information such as volatility, market momentum, social media chatter, and trends to construct a numerical index that swings between fear (indicating potential investor concern and a possible buying opportunity) and greed (reflecting possible market over-exuberance and a potential selling opportunity).

Both indices serve as pivotal market sentiment indicators, offering insights into investor psychology and potential shifts in market dynamics. They are instrumental for analysts and traders in making informed decisions, reflecting the collective emotions and attitudes that drive market movements.

- **On-chain analysis:** On-chain analysis refers to the examination of publicly available data recorded on a blockchain. This process involves analysing various metrics such as transaction volume, wallet addresses, the movement of cryptocurrencies between addresses, token issuance, and more. The primary goal of on-chain analysis is to gain insights into the financial health, trends, and potential future movements of blockchain networks and the digital assets they support.

Key components of on-chain analysis include:

- **Transaction Analysis:** By studying the volume and size of transactions over time, analysts can identify patterns, spikes in activity, or unusual transactions that may indicate significant market events or shifts in investor sentiment.
- **Wallet Address Tracking:** Examining the behaviour and balances of wallet addresses can reveal how individuals or entities are interacting with a cryptocurrency. It can help in identifying large holders (often referred to as "whales"), their impact on the market, and potential accumulation or distribution phases.
- **Network Health Metrics:** Metrics such as hash rate, node distribution, and active addresses can provide insights into the security and decentralization of a blockchain network. A healthy, secure, and well-distributed network is often seen as more attractive to investors.
- **Token omics:** Analysis of a token's issuance schedule, supply mechanisms (such as inflation rates or deflationary mechanisms), and distribution can offer insights into its long-term value proposition.
- The leading method of fasting and analyzing the cryptocurrency market, including traders, investors and researchers, the on-chain analysis enables the unique views only possible with blockchain technology. This particular kind of analytics complements the regular financial and technical analyses and adds a greater level of understanding about the market.

Technical indicators: like Moving Averages (MA) and Exponential Moving Averages (EMA) tend to be rather important in the analysis process. MA proposes the trend of an asset's price in a smoothed form by taking an average over a period against EMA which gives more weight to the recent prices, making it to act faster to the price changes.

Moving Average Convergence Divergence (MACD) is a comparatively advanced indicator which is used to indicate the interaction between two moving averages of an asset price. It meshes up with the movement, momentum and strength of a stock price plummet.

The Relative Strength Index (RSI) has the role of a momentum oscillator - it is used to measure how fast are the price movements increasing or decreasing. RSI fluctuates in the range of zero to one hundred and is employed to identify oversold or

overbought conditions in investors' basket. Taken all together these indicators give that much needed insight into how markets are moving and may also offer certain trade opportunities.

These technical indicators are pivotal in chart analysis, offering traders insights into potential future market movements and helping them to identify trade entry and exit points. Each indicator provides a different perspective on price and momentum, allowing for a multifaceted analysis of market conditions.

For better understanding, indices are organized into specific categories based on their analytical focus. These include:

Economic Analysis: This category features indices like the Global Economic Uncertainty Index (GEPU) which tracks global economic trends.

Market Analysis: Indices such as Bitcoin Dominance fall under this category, assessing the market share of Bitcoin relative to other cryptocurrencies.

Market Sentiment Analysis: This includes the Volatility Index (VIX) and the Crypto Fear & Greed Index, which gauge the market's mood and sentiment.

Technical Analysis: Tools like the Exponential Moving Average (EMA), Moving Average (MA), Moving Average Convergence Divergence (MACD), and Relative Strength Index (RSI) are used to analyze price movements and market trends.

On-chain Analysis: This comprehensive category includes Balance on Exchanges, BTC Options Volume Put/Call Ratio, Bitcoin Transfer Volume from and to Whale entities, Coin Block Destroyed (CBD), Coin Days Destroyed (CDD), Fee Ratio, HODL Ratio, Miner Revenue, Net Realized Profit Loss, Number of Active Addresses, Spent Output Profit Ratio (SOPR), and Short-Term Holder SOPR (STH-SOPR), which provide deep insights into blockchain activities and the behavior of cryptocurrency holders.

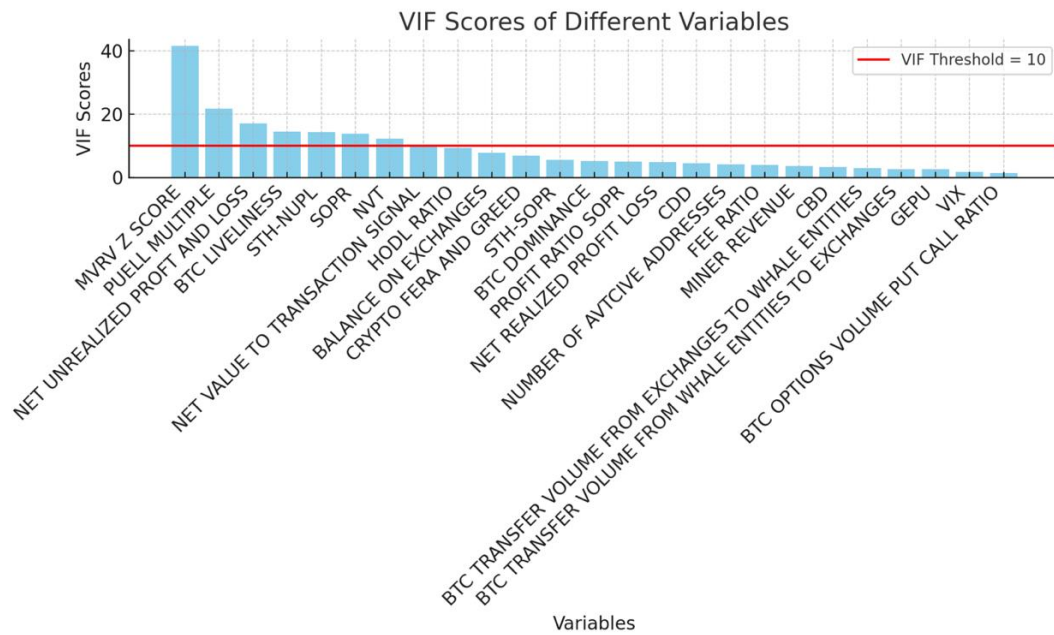


Figure 2. VIF Scores of Different Variables (Authors own calculation).

In this dissertation, the primary focus was on leveraging publicly accessible data. However, for the on-chain analysis segment, data was obtained from Glassnode. Consequently, certain indicators might require an update to ensure accessibility. Glassnode offered access to a comprehensive suite of metrics for the purpose of this study, yet this research selectively employed specific metrics based on Variance Inflation Factor (VIF) scores within the threshold. Additionally, the selection process was informed by recommendations found within other scholarly articles and studies. This approach allowed for a nuanced analysis that was both rigorous and tailored to the objectives of this study.

3.9. Data Analysis

3.9.1. Data Pre-processing. Prior to modelling, the dataset underwent several pre-processing steps to ensure data quality and compatibility with the ridge regression model:

- **Cleaning:** The dataset was cleaned to remove any rows with missing values, ensuring a complete dataset for analysis.
- **Normalization:** Given the wide range of values across different variables, feature scaling was applied using the MinMaxScaler method which is widely

used in Python's scikit-learn library for data pre-processing. This normalization ensures that each feature contributes proportionately to the model's predictions.

- Feature Selection: All available variables were included as predictors for the initial model to capture a broad spectrum of influences on Bitcoin's price.

The code was developed using the Python programming language within the JupyterLab environment, version 3.6.3. To facilitate the analysis, several libraries were utilized, including Pandas for data manipulation and analysis, NumPy for numerical operations, yfinance for fetching historical market data, TA-Lib for technical analysis indicators, Matplotlib and Seaborn for data visualization.

3.9.2. Model. In this research, Ridge Regression is utilized, which is a variant of regularized linear regression that integrates a regularization component into the cost function. Also referred to as L2 regularization, this method is part of a broader suite of regularization techniques designed to reduce overfitting within linear regression frameworks. The utility of Ridge Regression is particularly pronounced in the development of machine learning models endowed with a substantial number of parameters, notably when these parameters are associated with considerable magnitude. The inclusion of the regularization term is instrumental in curtailing the magnitude of the coefficients, thereby simplifying the model's complexity and averting the risk of overfitting. The application of Ridge Regression is especially pertinent in instances where predictor variables exhibit multicollinearity, or when the quantity of predictor variables surpasses the number of data observations.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (3.1)$$

- Y is the dependent variable (the variable to predict).
- $X_1, X_2, \dots, X_n$ are the independent variables (the predictors).
- $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ are the coefficients related to the model. β_0 is the intercept terms while others are the slopes associated with each independent variable.
- ϵ is the error term, which accounts for the difference between the predicted and actual values.

The Ridge Regression model was fitted using the following steps:

1. Alpha Selection: The regularization strength, alpha, plays a crucial role in Ridge Regression. A range of alpha values was explored through cross-

validation (RidgeCV in scikit-learn) to find the optimal value that minimizes the cross-validated mean square error (MSE).

In Ridge Regression, choosing the right regularization strength, denoted by α is key to balancing the model's complexity with its performance on unseen data. This selection process is not straightforward and relies heavily on experimentation. The goal is to find an α that minimizes the model's Mean Squared Error (MSE) on validation data, striking a balance between underfitting and overfitting. Here's how this is typically done:

- **Setting Up α Values:** Begin with a set of potential α values. These could be chosen based on previous experience, recommendations, or by using a systematic approach like logarithmic spacing that covers a broad spectrum of possibilities from very low (almost no regularization) to very high (strong regularization).
- **Employing Cross-Validation:** With a range of α values at hand, you employ a technique known as k-fold cross-validation. This involves partitioning the training data into k subsets. Each subset is used once as a validation set while the model is trained on the remaining $k - 1$ subsets. This process is repeated for each α value, and the MSE is recorded for every iteration.
- **Average MSE Calculation:** After running the cross-validation for all α values, average out the MSEs for each α to get a sense of which one generalizes best on the dataset. This averaging step helps in smoothing out variations that might occur due to the random partitioning of data in the cross-validation process.
- **Optimal α Selection:** The α that yields the lowest average MSE is chosen as the optimal one. This α represents the best compromise between fitting the training data accurately and maintaining a level of simplicity (or generality) that helps in preventing the model from overfitting.

Conceptually, Ridge Regression aims to minimize a cost function that comprises two parts: one that penalizes the model for not fitting the training data well enough (sum of squared residuals) and another that penalizes the model complexity (the sum of squared coefficients, scaled by α). The selection of α is essentially about finding the sweet spot where the model is complex enough to capture underlying patterns in the data but not so complex that it becomes overly tailored to the training set. This

explanation reiterates the core steps involved in selecting α for Ridge Regression in an intuitive manner, emphasizing the balance between model complexity and generalization.

Mathematically, the process of finding the optimal α does not translate to a direct equation but is more about this iterative procedure. However, if we were to summarize the objective function that Ridge Regression minimizes (including the regularization term), it would look something like this:

$$\beta \{ \sum_{i=1}^n (y_i - x_i^T \beta)^2 + \alpha \|\beta\|_2^2 \} \quad (3.2)$$

Here, β represents the model coefficients, y_i are the observed values, x_i are the feature vectors, and α is the regularization parameter to be optimized. The first term is the sum of squared residuals (a measure of how well the model fits the data), and the second term is the regularization penalty (α times the sum of squares of the coefficients), which prevents overfitting by keeping the coefficients small.

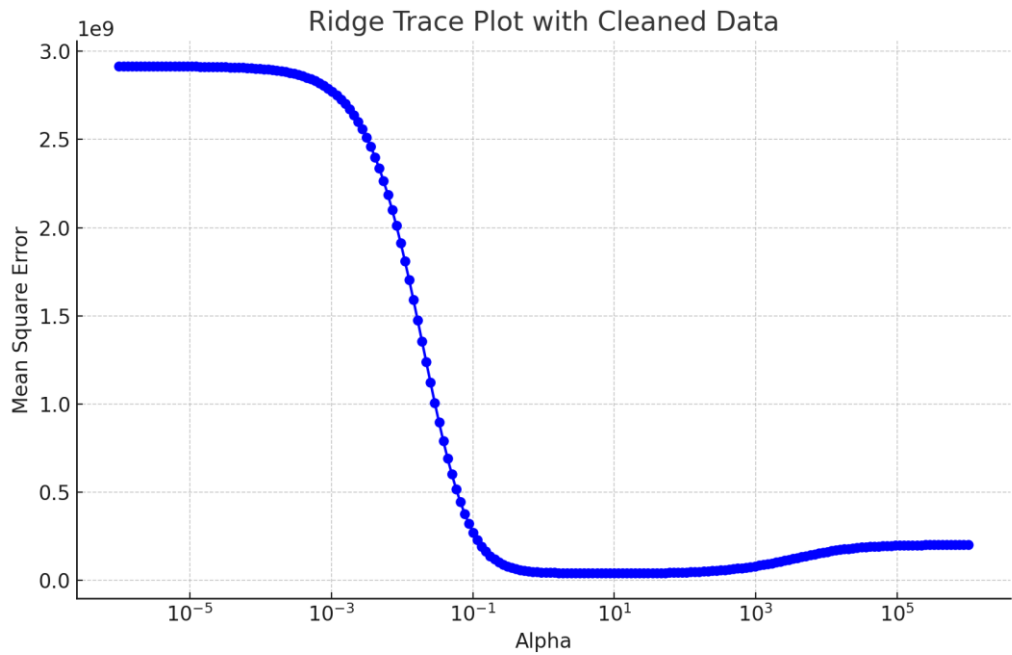


Figure 3. Ridge Trace Plot with Cleaned Data (Authors own calculation).

2. Model Fitting: With the optimal alpha determined, the Ridge Regression model was fitted to the training data. The dataset was split into training (80%) and testing (20%) subsets to evaluate the model's performance on unseen data.

3. **Evaluation:** The model's predictive performance was assessed using the root mean square error (RMSE), R-squared (R^2) and mean absolute error (MAE) metrics on the testing set. These metrics indicate the model's accuracy and the proportion of variance in the Bitcoin price that can be explained by the independent variables.

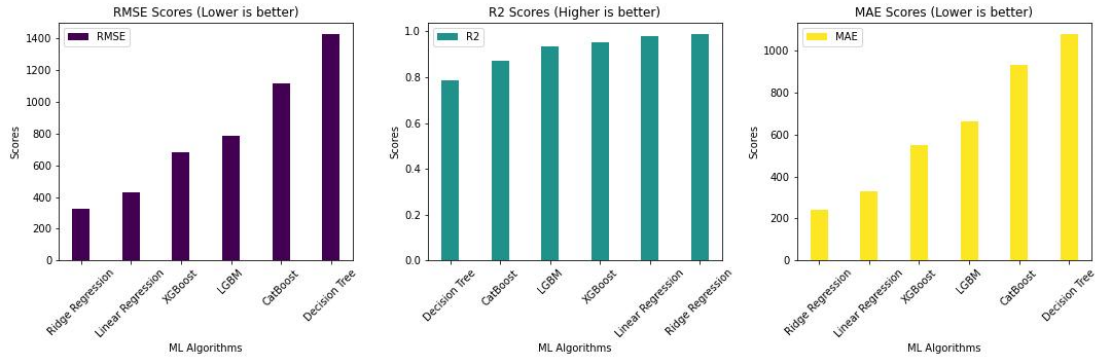


Figure 4. Models' Comparison with seen data (Authors own calculation).

3.9.3. Conclusion of Optimized Model after Comprehensive Comparative Analysis. In the process of evaluating various predictive models, including Ridge Regression, Linear Regression, LightGBM (LGBM), XGBoost, CatBoost, and Decision Tree Regressor a comprehensive comparative analysis was conducted based on key performance metrics such as the Root Mean Square Error (RMSE), R-squared (R^2), and Mean Absolute Error (MAE). This analytical assessment aimed to identify the model that demonstrates superior predictive accuracy and generalization capability in the context of our dataset. The outcome of this comparison revealed that Ridge Regression outperformed the other considered models across these evaluation metrics, showcasing its enhanced effectiveness in capturing the underlying patterns within the data while maintaining a balance between bias and variance. Consequently, Ridge Regression was selected as the primary model for further exploration and analysis within our study, underpinning our decision with empirical evidence of its optimal performance characteristics.

This study presents a comprehensive analysis of factors influencing Bitcoin's price using Ridge Regression. The methodology outlined combines rigorous data pre-processing with a systematic approach to model selection and evaluation, providing

insights into the predictive power of various market and non-market indicators on Bitcoin's price.



Chapter 4

Findings

Through comprehensive analysis and interpretation, this chapter unveils the significant findings and outcomes of the conducted research. A ridge regression was applied to forecast the price of Bitcoin (BTC). The optimal regularization parameter (alpha) was determined through cross-validation, ensuring the minimization of prediction error across various training subsets. The final regression model demonstrated a coefficient of determination R^2 of 0.95, indicating that 95% of the variance in Bitcoin prices could be explained by the model.

4.1. Correlation Analysis of Selected Independent Variables

The correlations between selected independent variables and the dependent variable have been presented as detailed below:

4.1.1. Correlation between Bitcoin price and Balance of Exchange.

Changes in the amount of Bitcoin stored on trading platforms are considered significant indicators of market trends. A reduction in Bitcoin balances on these exchanges usually signals a bullish sentiment, suggesting that investors might be moving their assets to private wallets for long-term holding in anticipation of rising prices. On the other hand, an increase in these balances might suggest that investors are preparing to sell, potentially leading to bearish market conditions. This relationship has been highlighted by trends where large outflows from exchanges typically precede periods of price appreciation. Godbole, O. (2024, March 8). Bitcoin Worth \$1B Leaves Exchanges in Largest Single-Day Outflow in 12 Months.

Over a period of four years, there has been a notable decrease in the supply of Bitcoin on exchanges, correlating with a bullish outlook in the market. This trend has been reinforced by the emergence of Bitcoin-focused financial products like ETFs, which facilitate greater investment from the traditional financial sector. Moreover, events such as Bitcoin halvings, which reduce the rate of new Bitcoin creation, also contribute to bullish sentiment as they lead to expectations of a supply shortage, further influencing market dynamics. Pokima, D. (2024, March 20). Bitcoin (BTC) Supply on Exchanges Slips To 4-Year Low, Impact On Price. These analyses underscore the

importance of monitoring Bitcoin balances on exchanges as a key metric for understanding and predicting investor behaviour and market movements.

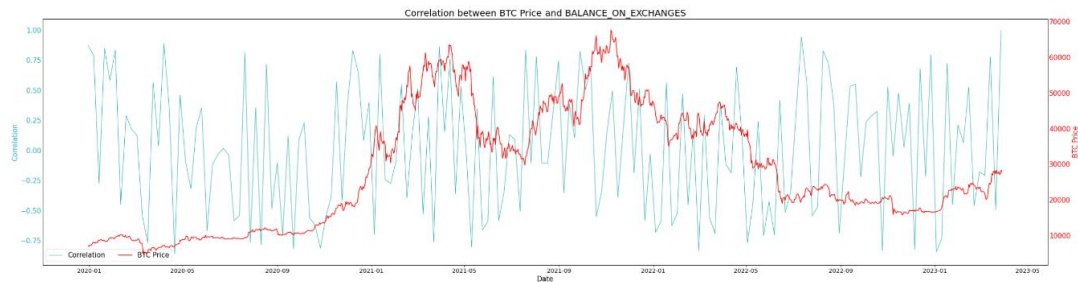


Figure 5. Correlation between BTC price and Balance of Exchange (Authors own calculation).

4.1.2. Correlation Between Bitcoin Price and Mining Revenue. The financial outcomes for Bitcoin miners have demonstrated a notable influence on the cryptocurrency's market price. In a significant instance from December 2023, Bitcoin miners achieved a record revenue of \$1.51 billion in one month, significantly boosted by earnings from transaction fees. This high revenue period corresponded with an increase in Bitcoin's market price, suggesting a link between the financial success of miners and the cryptocurrency's value fluctuations. This relationship is particularly evident as increased transaction fees suggest heightened network activity, potentially driving up Bitcoin prices. U.Today. (2024, January 1).

Furthermore, the overall profitability of mining Bitcoin, which depends on various factors like Bitcoin's market price, mining difficulty, and operational costs, notably impacts miners' economic decisions. Low profitability may compel miners to sell their Bitcoin holdings to sustain operations, increasing market supply and potentially lowering prices. Conversely, when mining is highly profitable, miners might opt to retain their Bitcoin, reducing market supply and possibly contributing to price increases. Cointelegraph. (2022, July 1). Not just Bitcoin price: Factors affecting BTC miner profitability.

These observations underscore the importance of keeping track of mining economics as a predictive tool for Bitcoin's market movements, especially during periods of significant shifts in mining profitability and operational costs.

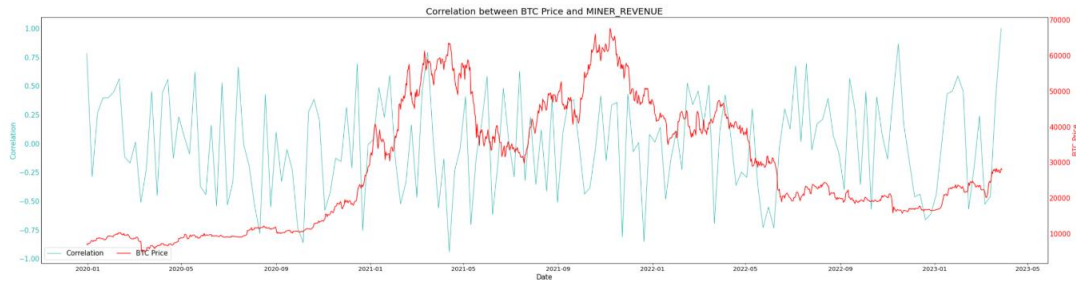


Figure 6. Correlation between BTC Price and Mining Revenue (Authors own calculation).

4.1.3. Correlation between Bitcoin Price and HODL. The HODL Waves chart, a visualization that uses blockchain data to categorize the age of Bitcoins held in wallets, offers insights into market behavior. By showing the distribution of Bitcoin across different age bands, the chart helps in understanding how long Bitcoins have been held without being moved. Sharp increases in the percentage of younger coins, indicated by lighter colors on the chart, suggest a significant movement of older Bitcoins, typically indicating that long-term holders are selling off their holdings, often around market peaks. This pattern can be seen during rapid price increases, as new market entrants rush to buy Bitcoin, which could potentially signal a local top in the market.

The RHODL Ratio, another indicator derived from the HODL Waves, also provides valuable insights. This metric considers the ratio of one-week-old coins to those aged between one and two years, adjusted for the age of the market. A high RHODL Ratio is often indicative of an overheated market, suggesting that a large amount of Bitcoin is being moved by newer holders, which can precede a market correction.

These tools are valuable for investors to gauge the sentiment and behavior of different market participants, particularly focusing on the actions of long-term holders versus new entrants. This can provide predictive insights into potential market tops and bottoms based on historical holding patterns. For more detailed analysis and visualizations, you can visit LookIntoBitcoin's page on HODL Waves and RHODL Ratio.(LookIntoBitcoin, 2024).

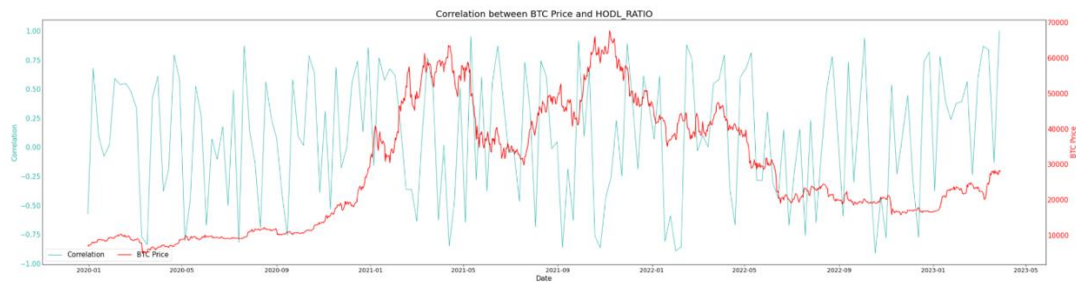


Figure 7. Correlation between Bitcoin price and HODL (Authors own calculation).

4.1.4. Correlation between Bitcoin Price and Number of Active Addresses. The number of active Bitcoin addresses is a key on-chain metric that often correlates with changes in Bitcoin's market price. An increase in active addresses typically indicates higher network usage, which can signal growing interest and participation in the Bitcoin market. For instance, during periods of price increases, the number of active addresses also tends to rise, reflecting more transactions and engagement on the blockchain. Research from various sources points out that significant fluctuations in the number of active Bitcoin addresses can provide insights into market dynamics. For example, a sustained increase in active addresses might suggest bullish sentiment, as more participants are engaging in transactions, possibly driven by positive market expectations or increased adoption. Conversely, a decline in active addresses might indicate bearish sentiment or a cooling off period following a price spike.

Throughout 2020, the Bitcoin network saw dramatic shifts in the number of active addresses, paralleling shifts in market conditions. The increased activity was accompanied by significant price movements and heightened trading volume, suggesting that changes in the number of active addresses can be a useful indicator for understanding market trends and investor behavior.

For a detailed exploration of how Bitcoin's active addresses relate to market dynamics, you can visit comprehensive cryptocurrency analysis platforms like Bitbo and Decrypt, which provide in-depth data and visualizations on this and other related metrics. These resources can offer valuable insights for anyone looking to understand the implications of network activity on Bitcoin's price.

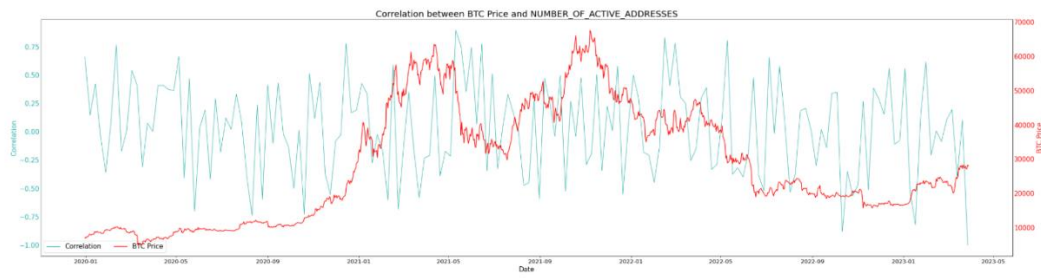


Figure 8. Correlation between Bitcoin Price and Number of Active Addresses
(Authors own calculation).

4.1.5. Correlation between Bitcoin price and Net Realized Profit/Loss (NRPL). The Net Realized Profit/Loss (NRPL) for Bitcoin provides crucial insights into market sentiment and overall capital flows into and out of Bitcoin. This metric, which represents the net amount of profit or loss realized by Bitcoin holders over a given period, plays a significant role in understanding market dynamics.

During bull markets, the NRPL often shows a dominance of realized profits as investors cash in on rising prices. This trend frequently peaks around local or even global market highs, which may trigger a market correction due to profit-taking overwhelming incoming demand. Conversely, in bear markets, realized losses typically prevail, especially during periods of maximum negative sentiment and market exhaustion, often seen at bear market lows.

These insights are derived from a comprehensive analysis of Bitcoin transactions, where coins are revalued as they change hands from long-term holders to new market entrants, often at higher prices during bullish phases. This transactional behavior underpins critical phases of market cycles, influencing both the bull and bear market dynamics.

Understanding the NRPL metric can offer investors a nuanced view of how Bitcoin's market value is influenced by underlying economic activities, beyond mere price fluctuations. This can be especially useful for predicting potential turning points in the market based on historical behaviour of profit taking and loss realization among Bitcoin holder.

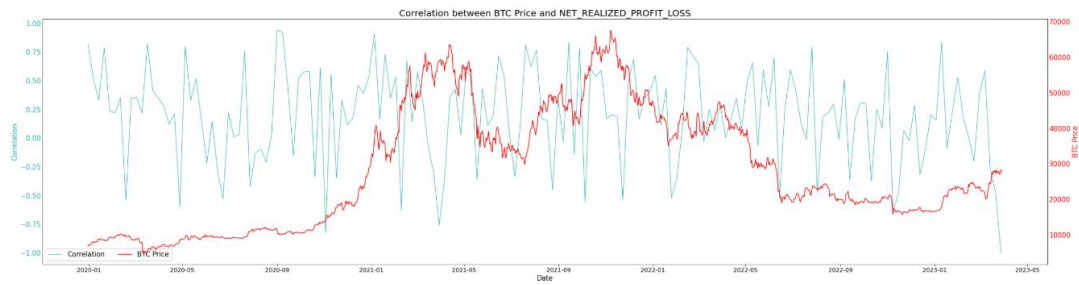


Figure 9. Correlation between Bitcoin Price and Net Realized Profit/Loss (NRPL)
(Authors own calculation).

4.1.6. Correlation between Bitcoin price and Fee Ratio. The Bitcoin fee ratio, often referred to as the transaction fee ratio, is an important metric for understanding the economic dynamics of the Bitcoin network. This ratio measures the fees paid to miners for each transaction relative to the total block reward, which includes both these transaction fees and the new Bitcoins minted with each block.

As the block reward decreases over time due to halving events, transaction fees are expected to become a more critical part of miners' rewards. This shift is significant for the network's security and sustainability because it compensates for the diminishing block rewards, ensuring that miners remain incentivized to process transactions.

High fee ratios can indicate a robust demand for transaction processing, often seen during periods of high network congestion. Conversely, lower fee ratios might suggest less strain on the network or lower overall demand for transaction throughput. Understanding and monitoring the Bitcoin fee ratio is crucial, especially as the network approaches its maximum supply of 21 million coins. At this point, transaction fees will be the primary incentive for miners, influencing decisions on transaction inclusion and network security.

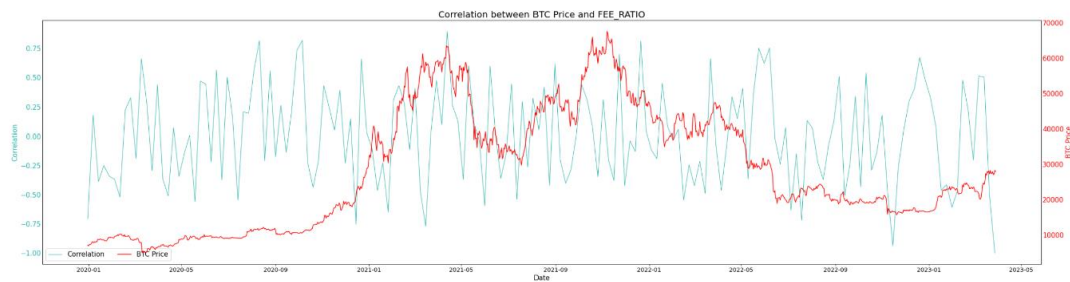


Figure 10. Correlation between Bitcoin Price and Fee Ratio (Authors own calculation).

4.1.7. Correlation between Bitcoin Price and Spent Output Profit Ratio (SOPR). The Spent Output Profit Ratio (SOPR) is an insightful on-chain metric that provides a snapshot of the profitability of Bitcoin transactions over time. SOPR calculates the ratio of the realized value of a Bitcoin transaction (the price at which the Bitcoin is sold) to the value at the time the Bitcoin was last moved or acquired. Essentially, it indicates whether Bitcoin holders are selling their assets at a profit or a loss.

When SOPR is above 1, it generally means that holders are selling Bitcoin at a profit, as the selling price exceeds the price at which the Bitcoins were previously acquired. On the other hand, a SOPR below 1 indicates that transactions are occurring at a loss, which can be particularly common during bear market conditions. This metric is valuable for analyzing market sentiment and behavior. For example, a persistently high SOPR during a bull market reflects a trend where most sellers are realizing profits. Conversely, during bear markets, SOPR often dips below 1, suggesting widespread losses and potential market capitulation.

SOPR is useful not only for individual traders but also for gaining a broader understanding of market dynamics. It can help identify when the market is likely overly pessimistic or optimistic, potentially signalling key turning points for strategic investment decisions. For those interested in deeper insights or real-time SOPR data, platforms like Decentrader and Bitcoinnews offer comprehensive tools and analyses that track and interpret SOPR in the context of broader market trends.

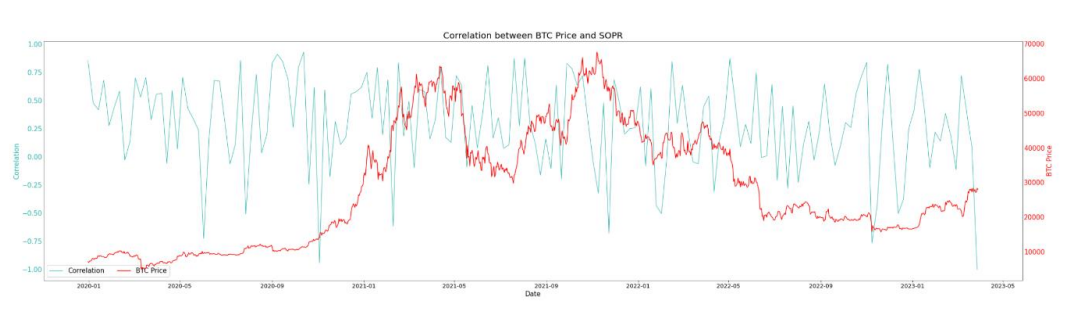


Figure 11. Correlation between Bitcoin Price and Spent Output Profit Ratio (SOPR)
(Authors own calculation).

4.1.8. Correlation between Bitcoin Price and Coin Days Destroyed (CDD). Coin Days Destroyed (CDD) is a significant on-chain metric in Bitcoin analysis that helps gauge the market movements of long-held Bitcoins. The concept of CDD involves measuring the total number of days that the transacted Bitcoins had been dormant before moving. When these Bitcoins are spent, the "coin days" accumulated are considered "destroyed." This metric is particularly insightful for identifying large-scale movements of old coins, which can indicate significant market actions by long-term holders. CDD is calculated by multiplying the number of Bitcoins moved in a transaction by the number of days they were held dormant. A higher CDD value suggests that long-term holders are selling their holdings, potentially leading to increased market volatility or major price corrections. Conversely, a lower CDD indicates less movement of old coins, which could mean market stability or ongoing accumulation phases where long-term holders are not selling their stakes.

The interpretation of CDD can provide insights into market sentiment. For instance, a rising CDD during a market upturn might suggest that long-term holders are starting to take profits, signalling a potential market top. On the other hand, a decreasing CDD in a downward trending market might imply that fewer long-term holders are selling their coins, potentially indicating a market bottom or consolidation phase.

Understanding the nuances of CDD, especially in the context of market cycles and the behavior of long-term holders, can be crucial for both traders and investors looking to gauge the underlying market dynamics and make informed decisions based on the activity of Bitcoin's most significant stakeholders.

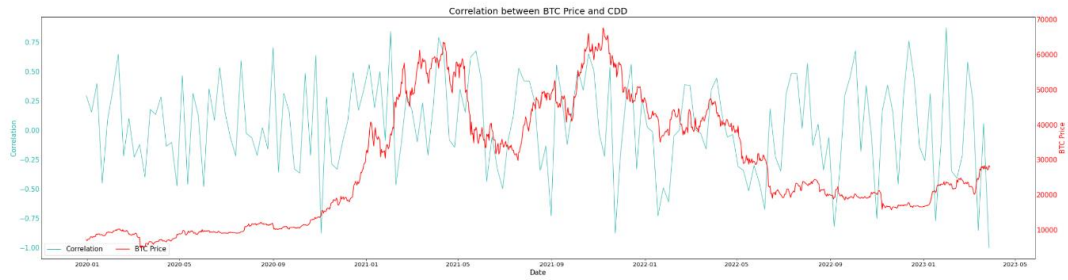


Figure 12. Correlation between Bitcoin price and Coin days destroyed (CDD)
(Authors own calculation).

4.1.9. Correlation between Bitcoin and the VIX. During periods when the VIX experienced significant spikes, Bitcoin often mirrored this volatility, indicating a strong correlation between market uncertainty (as reflected by the VIX) and Bitcoin's price movements. For example, in 2018, as the VIX peaked due to increased market fear, Bitcoin prices showed considerable volatility, suggesting that Bitcoin acts more like a risk asset rather than a safe haven during times of market distress. Ouimet, S. (2019, January 21). Bitcoin's Price Counters the VIX, Confirming It's Still a Risk Asset. The impact of the VIX on Bitcoin was particularly noticeable during the COVID-19 pandemic. As the VIX reached elevated levels, indicating high market fear, Bitcoin prices initially fell sharply. This correlation highlights Bitcoin's sensitivity to broader economic indicators and market sentiments during unprecedented times, acting contrary to what might be expected from a so-called "safe haven" asset. Lopez-Cabarcos, M. Á., Kraaijeveld, O., & Smedt, J. D. (2020). Sun, Y., et al. (2023). The effect of hashrate, transaction volume, social media, and macroeconomics on Bitcoin before and during the COVID-19 pandemic.

These studies indicate that Bitcoin's price is responsive to the same factors that influence traditional financial markets, especially during periods of high volatility and market fear as indicated by movements in the VIX.

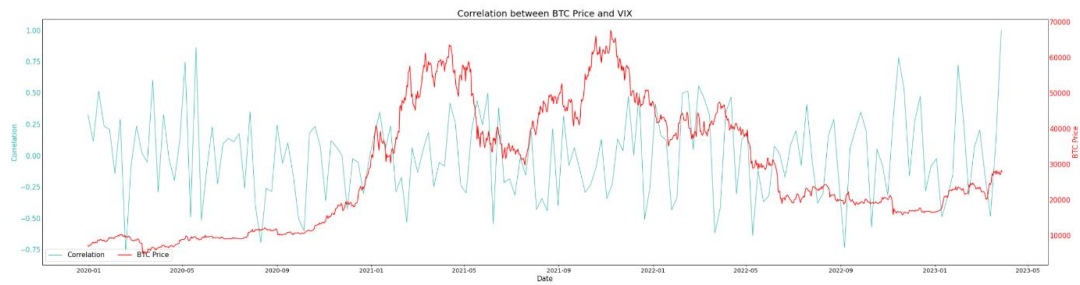


Figure 13. Correlation between Bitcoin and the VIX (Authors own calculation).

4.2. Correlation Map

A correlation map, or correlation matrix visualized as a heatmap, is a powerful statistical tool used to summarize the correlation coefficients between multiple variables. Here's an expanded explanation of what a correlation map shows:

- **Detailed Relationships Between Variables:** Each cell in the correlation map represents the correlation coefficient between two variables. This coefficient quantifies the degree to which the variables tend to change together. A positive correlation indicates that as one variable increases, the other tends to increase as well. Conversely, a negative correlation suggests that as one variable increases, the other tends to decrease. Values near zero indicate a weak or no linear relationship.
- **Color-coded Heatmap:** The visualization often uses colors to represent the strength and direction of correlations. Commonly, warmer colors (like reds) indicate positive correlations, cooler colors (like blues) signify negative correlations, and neutral colors (like whites or light grays) denote correlations near zero. This color coding makes it easier to visually scan for relationships and patterns among the variables.
- **Symmetry and Diagonal Consistency:** The matrix is symmetrical because the correlation between any two variables remains the same regardless of their order (i.e., correlation between A and B is the same as B and A). The main diagonal, running from the top left to the bottom right, always shows a perfect correlation of 1, since every variable is perfectly correlated with itself.
- **Uses in Data Analysis:** Analysts use correlation maps to quickly identify which pairs of variables have strong correlations, which can suggest potential causal relationships, areas for further investigation, or redundant variables that could

be removed from analysis. This is particularly useful in feature selection in machine learning, risk management in finance, or gene expression studies in genetics.

- **Statistical Considerations:** While correlation maps are informative, they only measure linear relationships and can be misleading if the data contain outliers or if the relationship is nonlinear. Additionally, correlation does not imply causation; further analysis is often required to establish causal relationships.

Below is the correlation map displaying all the available variables prior to selection:

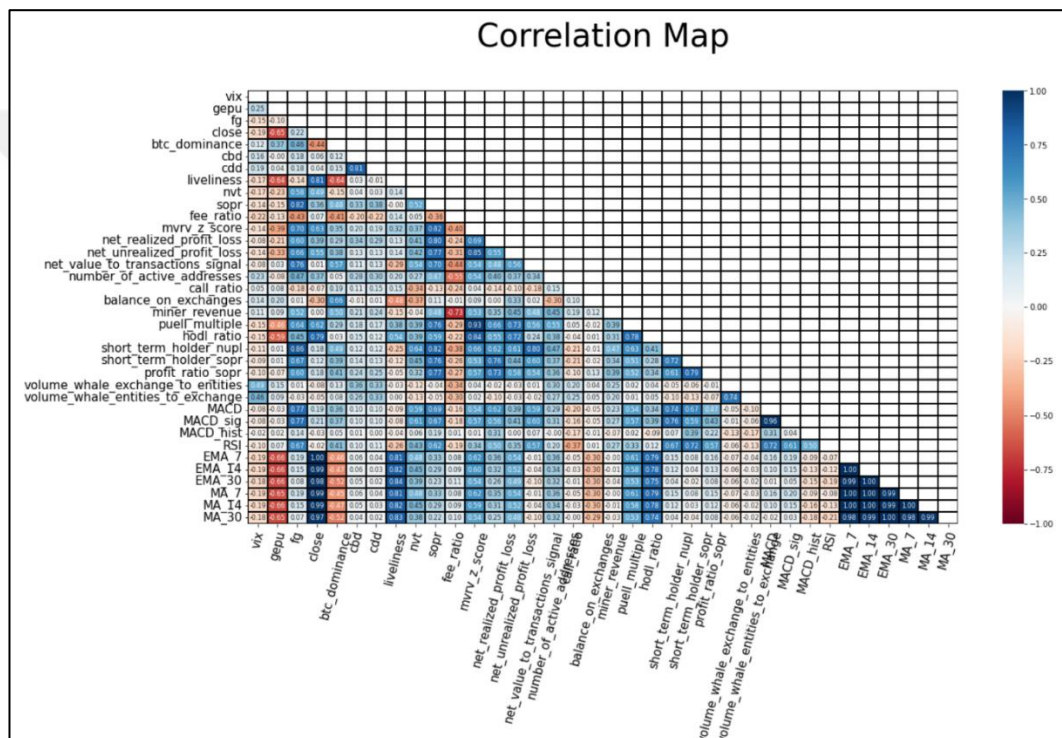


Figure 14. Correlation Map (Authors own calculation).

Below is the correlation map illustrating the relationships among all selected variables for the model:

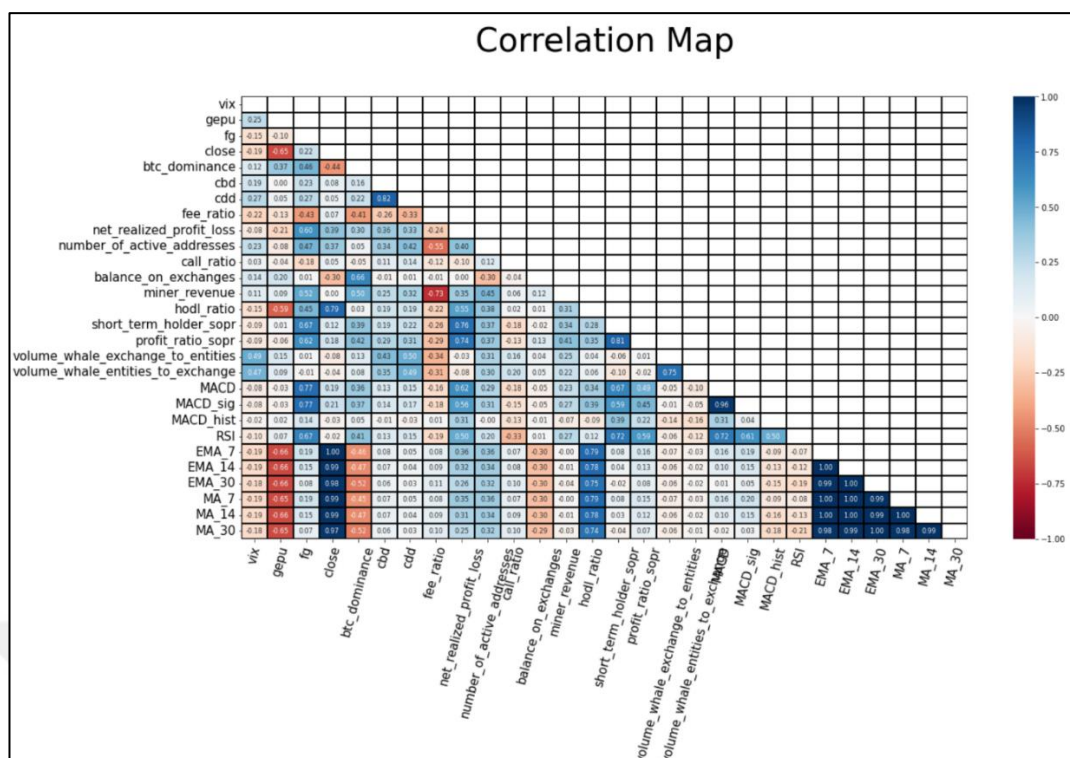


Figure 15. Correlation Map (Selected Metrics) (Authors own calculation).

Overall, correlation maps serve as an initial screening tool for relationships between multiple variables, helping researchers and analysts focus their investigations and hypotheses on the most promising pairs of variables. Each cell in the grid shows the correlation coefficient between two variables, where:

- A value of 1 indicates a perfect positive correlation (as one variable increases, the other does as well).
- A value of -1 indicates a perfect negative correlation (as one variable increases, the other decreases).
- A value close to 0 indicates little to no linear relationship between the variables.

Colors are used to represent the strength of correlation: blue tones signify positive correlations, and red tones indicate negative correlations. The darker the shade, the stronger the correlation, whether positive or negative. This type of map is commonly used in finance and economics to identify relationships between indicators that might influence decision-making processes.

4.3. Training Models

The training results for the six models, each designed to test different aspects of our hypotheses, are detailed in the section below. These results include a variety of performance metrics and analyses that provide insights into the efficacy and accuracy of each model under evaluation.

Table 1

Model's Performance

Model	Linear Regression	Ridge Regression	Decision Tree	XGBoost	CatBoost	LGBM
RMSE	429.50779	323.23744	1428.8508	681.5973	1113.9113	787.61518
R2	0.98061	0.98902	0.78537	0.95116	0.86956	0.93479
MAE	331.73687	242.78389	1079.7968	551.41272	933.73873	663.55885

For better understanding, the results from the table have been visually represented through graphs. This graphical depiction helps in more clearly illustrating the trends and comparisons between the performance metrics of the different models.

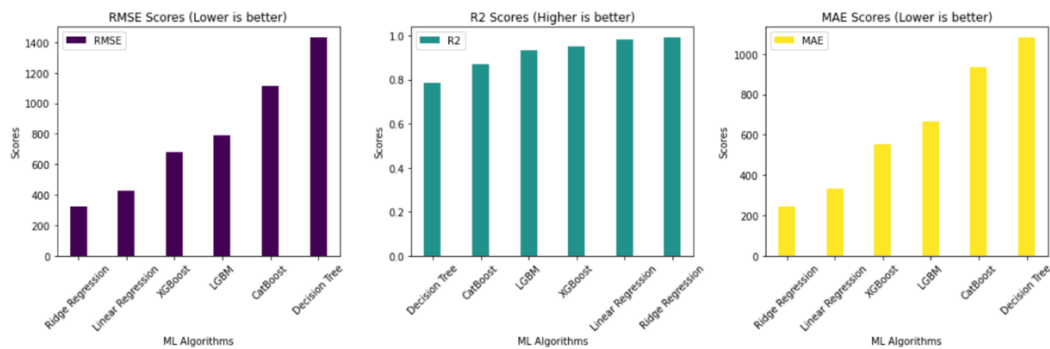


Figure 16. Models' Comparison with Unseen Data (Authors own calculation).

The best result was obtained using ridge regression, which suggests that further research should continue with this model to explore and optimize its capabilities further. In Ridge Regression, the term "alpha" is a hyperparameter that controls the strength of the regularization applied to the model. Regularization is a technique used to reduce overfitting by penalizing large coefficients in regression models.

4.4. Understanding Alpha in Ridge Regression

4.4.1. Role of alpha

- Alpha multiplies the regularization term added to the cost function of ordinary least squares regression. The regularization term is the sum of the squares of the coefficients multiplied by alpha. The complete cost function for Ridge Regression can be expressed as:

$$\text{Cost Function} = \|y - X\beta\|^2 + \alpha\|\beta\|^2 \quad (4.1)$$

Where y is the vector of target values, X is the matrix of input features, and β are the coefficients?

4.4.2. Impact of alpha

- High Alpha Value: A larger alpha increases the penalty for larger coefficients, effectively shrinking them towards zero. This can lead to a simpler, less complex model that might underfit the data.
- Low Alpha Value: A smaller alpha means less emphasis on the regularization term, making the solution more sensitive to the data, potentially leading to a model that overfits.

4.4.3. Choosing alpha

- The value of alpha can significantly affect the performance of a Ridge Regression model. It usually requires tuning to find the optimal value that balances bias and variance.
- Hyperparameter optimization for alpha typically involves techniques like cross-validation where different values of alpha are tested to see which produces the best fit to the data without overfitting.

```
Best Parameters: {'alpha': 0.5381457318174152}  
Lowest RMSE: 323.04337
```

Figure 17. Optimal Alpha in Ridge Regression

4.4.4. Optimization techniques

- Common techniques for optimizing alpha include Grid Search, where a range of alpha values are specified and the best one is selected based on performance metrics.
- Random Search and Bayesian Optimization are other methods that can be more efficient than Grid Search, especially when dealing with a large hyperparameter space.

In the construction of the Ridge Regression model, an alpha parameter of approximately 0.538 was selected to optimize the regularization process. This value signifies a deliberate balance between reducing the complexity of the model and maintaining its ability to capture significant predictors in the presence of multicollinearity. Regularization, particularly through the Ridge method, imposes a penalty on the size of coefficients; effectively, this penalizes large coefficients by shrinking them towards zero, thereby mitigating the risk of overfitting. This approach is especially beneficial in scenarios where the predictor variables are highly correlated. The chosen alpha value suggests a model design that cautiously avoids overfitting while retaining sufficient model flexibility to elucidate meaningful patterns in the data, thereby ensuring generalizability and robust predictive performance.

Furthermore, the model's efficacy is demonstrated by achieving the lowest Root Mean Square Error (RMSE) of 323.04337. The RMSE metric serves as a critical measure of the prediction error, quantifying the average magnitude of the predictive error in the same units as the dependent variable. This error metric provides an essential benchmark for evaluating the precision of the predictive model. An RMSE of approximately 323 indicates that the model predictions deviate from the actual observations by this amount on average, which, depending on the context and scale of the data used, may represent an acceptable level of prediction accuracy. The attainment of this RMSE, in conjunction with an alpha of 0.538, underscores the effectiveness of the regularization strength in enhancing the model's accuracy, making it well-suited for practical applications where prediction reliability is paramount. This balance between model complexity and prediction error is crucial for developing models that are both robust and precise, suitable for deployment in various predictive tasks within the field.

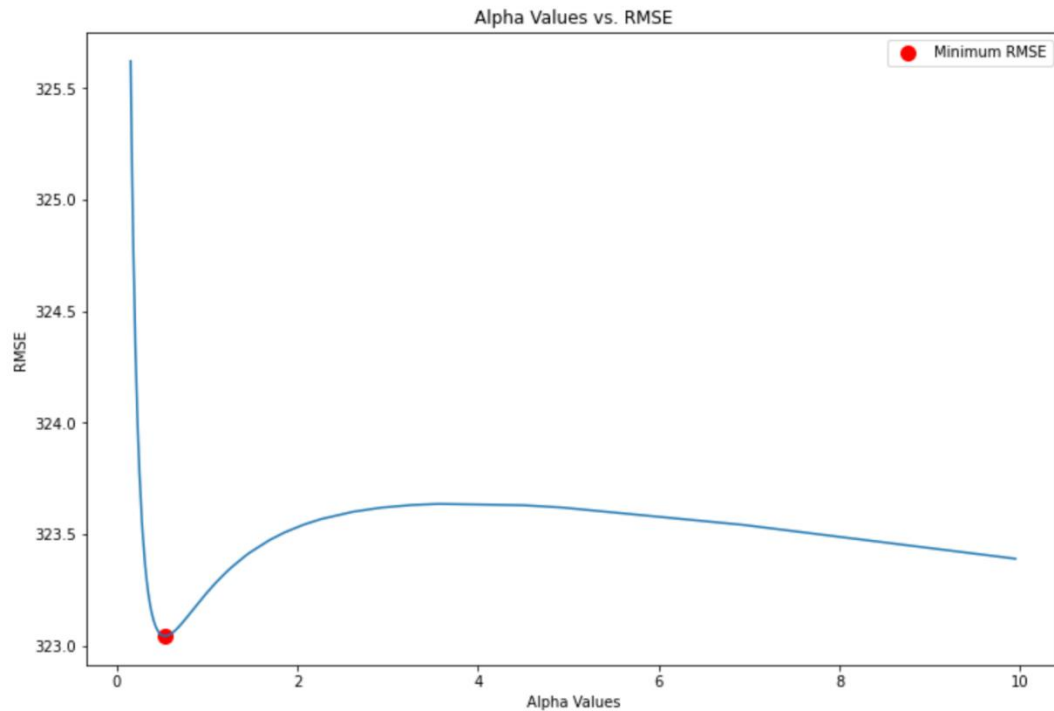


Figure 18. Alpha Values Vs. RMSE (Authors own calculation).

This graph shows the relationship between different alpha values and the root mean square error (RMSE) in the model. RMSE measures the average magnitude of errors between predicted and observed values, with lower values indicating better model performance.

The plot illustrates that as alpha values increase from 0 to approximately 2, the RMSE decreases sharply to its minimum (indicated by the red dot), suggesting that the model performs best around this alpha level. Beyond $\alpha = 2$, the RMSE starts to increase slightly, indicating that higher alpha values may lead to a model that fits the data less accurately. This kind of analysis helps in tuning hyperparameters to optimize model performance.

4.5. Predictor's Coefficients in Ridge Regression

In the context of a regression analysis such as ridge regression, the coefficients attributed to the predictor variables provide insights into the dynamics between each predictor and the dependent variable, exemplified here as the Bitcoin (BTC) price. The interpretation of these coefficients within such a model encompasses several aspects:

4.5.1. Magnitude and Direction

- **Magnitude:** The absolute value of a coefficient delineates the strength of the effect that the corresponding predictor exerts on the dependent variable. A larger coefficient magnitude suggests a more pronounced influence on the dependent variable.
- **Direction:** The sign of the coefficient, whether positive or negative, indicates the direction of the relationship between the predictor and the dependent variable. Specifically, a positive coefficient signifies that an increase in the predictor variable is associated with an increase in the dependent variable. Conversely, a negative coefficient implies that an increase in the predictor leads to a decrease in the dependent variable.

4.5.2. Relative importance

- While coefficients can tentatively suggest which variables might be more influential, it is important to note that direct comparisons between the magnitudes of coefficients to ascertain the relative importance of variables are valid only when all variables have been standardized or are inherently on the same scale.

4.5.3. Impact of regularization. Ridge regression incorporates a regularization penalty that induces shrinkage of the coefficients towards zero. This regularization is controlled by the parameter α :

- A larger α enhances the penalty, resulting in greater coefficient shrinkage. This increased shrinkage helps in curtailing model complexity and mitigating overfitting, particularly beneficial in scenarios characterized by multicollinearity or a high ratio of predictors to observations.
- Conversely, a smaller α diminishes the degree of shrinkage, causing the model to approximate an ordinary least squares regression model more closely, particularly as α approaches zero.

4.5.4. Practical interpretation

- Each coefficient can be interpreted as quantifying the expected change in the BTC price, measured in the same units as the BTC price itself, for a one-unit alteration in the respective predictor, holding all other predictors constant.
- For instance, should a coefficient associated with trading volume be 2.3, it would imply that an incremental increase in trading volume by one unit is expected to lead to an increase in the Bitcoin price by 2.3 units, provided other variables are held constant.

The interpretation of these coefficients should invariably account for the influence of regularization, which can significantly affect the perceived impact or importance of individual variables in comparison to models devoid of regularization, such as ordinary least squares regression.

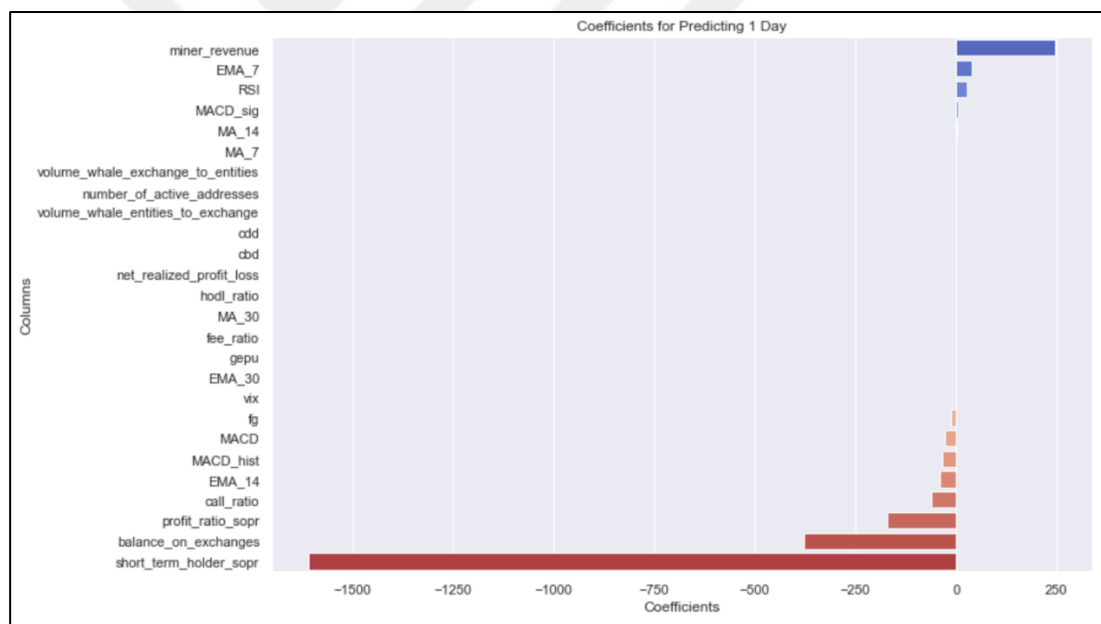


Figure 19: Coefficients for Predicting 1 Day (Authors own calculation).

This graph displays the coefficients of different variables used in the model forecasting the outcome of the model one day ahead. The coefficients represent the influence each variable has on the forecast:

- Bars extending to the right (positive values) suggest that an increase in these variables is associated with an increase in the predicted outcome.

- Bars extending to the left (negative values) indicate that an increase in these variables is associated with a decrease in the predicted outcome.

The graph is useful for understanding which variables are most influential in the model. For example, certain variables at the bottom, showing large negative coefficients, have a strong negative impact, while those at the top with positive values have a significant positive impact.

4.6. Model Formulation, Predictive Accuracy, and Residual Analysis in Ridge Regression-Based Bitcoin Price Forecasting

In the constructed ridge regression model utilized for forecasting Bitcoin (BTC) prices, the predictive equation serves as the foundational mechanism for evaluating the impact of various market influencers. Formally expressed as:

$$BTC\ PRICE = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (4.2)$$

In this equation, β represents the intercept, and each β_i corresponds to the coefficient of the predictor X_i . These coefficients elucidate how each predictor variable is expected to influence the Bitcoin price, with the direction and magnitude of each coefficient indicating the nature and strength of this influence. The equation obtained for forecasting Bitcoin price via ridge regression is outlined below:

Equation:
 $(245.89 \times \text{miner_revenue}) + (39.86 \times \text{EMA_7}) + (25.37 \times \text{RSI}) + (6.11 \times \text{MACD_sig}) + (1.14 \times \text{MA_14}) + (0.57 \times \text{MA_7}) + (0.0 \times \text{volume_whale_exchange_to_entities}) + (0.0 \times \text{number_of_active_addresses}) + (0.0 \times \text{volume_whale_entities_to_exchange}) + (0.0 \times \text{cdd}) + (0.0 \times \text{cbd}) + (-0.0 \times \text{net_realized_profit_loss}) + (-0.0 \times \text{hodl_ratio}) + (-0.2 \times \text{MA_30}) + (-0.74 \times \text{fee_ratio}) + (-1.28 \times \text{gepu}) + (-1.36 \times \text{EMA_30}) + (-1.88 \times \text{vix}) + (-13.01 \times \text{fg}) + (-27.12 \times \text{MACD}) + (-33.24 \times \text{MACD_hist}) + (-39.0 \times \text{EMA_14}) + (-62.59 \times \text{call_ratio}) + (-171.28 \times \text{profit_ratio_sopr}) + (-379.01 \times \text{balance_on_exchanges}) + (-1610.08 \times \text{short_term_holder_sopr})$

The model is derived by optimizing a cost function that incorporates a penalty on the magnitudes of the coefficients, effectively a regularization term scaled by the parameter α . This L2 regularization term is critical as it mitigates the risk of overfitting by penalizing large coefficients, thus ensuring the model's robustness and enhancing its predictive accuracy in the highly volatile cryptocurrency market.

This robust statistical approach allows the model to capture the intricate dynamics within the cryptocurrency market, accounting for multiple predictors simultaneously while maintaining generalization capabilities outside the training dataset. The predictive formula is not only a reflection of the quantified relationships

between Bitcoin prices and their predictors but also an embodiment of the careful balance between model complexity and predictive performance, crucial for making informed economic forecasts in the domain of financial technology.

The ridge regression model employed in this study has proven to be a robust tool for predicting Bitcoin prices. The choice of ridge regression was validated by its ability to handle multicollinearity among predictor variables and reduce overfitting through the introduction of a regularization parameter, alpha. This was crucial given the volatile and often unpredictable nature of cryptocurrency markets, which are influenced by a myriad of factors beyond typical market predictors.

```
y_test = pred_data["close"]
y_pred = final_model.predict(pred_data.drop("close", axis=1).drop(multicol_cols, axis=1))

print("Real Value:", round(y_test.iloc[0], 2))
print("Predicted Value:", round(y_pred[0], 2))
print("The difference between real and predicted:", round(mean_absolute_error(y_test, y_pred), 2))
print(f"Error between real and predicted: {round(100*y_pred[0] / y_test.iloc[0] - 100, 2)}%")
```

Real Value: 28478.48
Predicted Value: 29639.36
The difference between real and predicted: 1160.87
Error between real and predicted: 4.08%

Figure 20. Codes Related to Comparison Real Value of BTC and Forecasted Value with Error via Ridge Regression

The observed error rate of 4.08% between the predicted and actual prices is notably low, considering the complex and speculative dynamics of the cryptocurrency market. This error margin might be attributed to the model's inability to incorporate real-time market sentiment, regulatory announcements, or unexpected global economic events, which can drastically affect cryptocurrency prices.

Future enhancements to the model could include the integration of additional data types, such as sentiment analysis from social media platforms, blockchain activity metrics, and macroeconomic indicators. Moreover, the application of more dynamic modeling techniques, like ensemble methods or deep learning, could potentially reduce the prediction error and capture more complex patterns in the data.

In conclusion, while the ridge regression model provides a strong foundation for predicting Bitcoin prices, it also highlights the importance of continuously

enhancing predictive models to incorporate broader data sets and advanced algorithms to better mirror the multifaceted nature of cryptocurrency markets.



Chapter 5

Discussion and Conclusion

In this study, the primary aim was to develop a predictive model for Bitcoin prices, utilizing Ridge Regression due to its proficiency in handling multicollinearity among predictors and its robustness against overfitting. The application of the Ridge Regression model has demonstrated substantial efficacy in forecasting Bitcoin's volatile price movements. This success can largely be attributed to the model's ability to incorporate a regularization parameter, which effectively shrinks the coefficients of less significant variables, thus enhancing the model's generalizability to new, unseen data. By examining the predictive accuracy and stability of the model, this research provides critical insights into the predictive dynamics of cryptocurrency markets, particularly Bitcoin, under various economic conditions.

5.1.Implications

The implications of this study, which employs Ridge Regression to analyze a dataset incorporating technical indicators, economic factors, on-chain metrics, and market sentiment, are profound for various stakeholders. For traders, the refined model offers enhanced predictive accuracy concerning Bitcoin price, providing a robust tool for making informed trading decisions in the face of market volatility. Policy makers can benefit from insights into how different economic conditions and market sentiments influence cryptocurrency markets, supporting more informed regulatory and fiscal policymaking. Additionally, the integration of diverse data types—ranging from economic indicators to sentiment analysis—presents a comprehensive approach that future studies could adopt to understand other financial instruments or markets. This holistic modelling technique not only advances academic research methodologies but also offers practical, data-driven insights that could be instrumental in developing more resilient financial systems and adaptive trading strategies in the increasingly digital and interconnected global economy.

5.2.Limitation. The limitations of this study employing Ridge Regression to predict Bitcoin prices using a diverse dataset can be multifaceted and impact the interpretation and generalizability of the results. Here are several potential limitations:

- **Model Sensitivity to Outliers:** While Ridge Regression is effective at handling multicollinearity among predictors, it is not inherently robust against outliers. Bitcoin's price is highly volatile and susceptible to rapid, outlier shifts due to market news or speculative activity, which could skew the model predictions.
- **Static Dataset:** The dataset comprises technical, economic, on-chain, and market sentiment metrics at a specific point in time. Bitcoin and cryptocurrency markets evolve rapidly, and the static nature of the dataset might not capture future market dynamics or changes in economic conditions.
- **Overemphasis on Quantitative Data:** The reliance primarily on quantitative metrics may overlook qualitative factors such as regulatory changes, technological advancements, or geopolitical events that can significantly influence Bitcoin prices.
- **Regularization Bias:** The use of Ridge Regression introduces a bias through its regularization parameter, which may sometimes lead to underfitting if not optimally tuned, potentially oversimplifying the model by excessively penalizing the coefficient estimates.
- **Generalizability Issues:** The results obtained are specific to the conditions and characteristics of the Bitcoin market and may not be directly applicable to other cryptocurrencies or financial assets with different underlying dynamics.
- **Dependency on Historical Data:** The predictive capability of the model is based on historical data, assuming that past patterns will continue. This might not hold true in all cases, especially in a financial context where future conditions can differ drastically due to unforeseen factors.

The number of consecutive days used for forecasting has been extended to assess the model's response and accuracy, as detailed below.

RMSE Score on real values for 5 days prediction: 404.5119726641114
 RMSE Score on real values for 6 days prediction: 363.3259586911149
 RMSE Score on real values for 7 days prediction: 362.9260075888731
 RMSE Score on real values for 8 days prediction: 368.75372069304973
 RMSE Score on real values for 9 days prediction: 369.88966929639224
 RMSE Score on real values for 10 days prediction: 365.80050217081686
 RMSE Score on real values for 11 days prediction: 601.3435624723253
 RMSE Score on real values for 12 days prediction: 861.6585686245363
 RMSE Score on real values for 13 days prediction: 1032.5597924514398
 RMSE Score on real values for 14 days prediction: 1207.8823556961026
 RMSE Score on real values for 15 days prediction: 1372.2313267194081
 RMSE Score on real values for 16 days prediction: 1506.1401971631904
 RMSE Score on real values for 17 days prediction: 1633.6223700762753
 RMSE Score on real values for 18 days prediction: 1699.584616827608
 RMSE Score on real values for 19 days prediction: 1846.4023765577492
 RMSE Score on real values for 20 days prediction: 1920.771573903999
 RMSE Score on real values for 21 days prediction: 1996.769814254148
 RMSE Score on real values for 22 days prediction: 2070.8114040022306
 RMSE Score on real values for 23 days prediction: 2144.22147166977
 RMSE Score on real values for 24 days prediction: 2213.424236154669
 RMSE Score on real values for 25 days prediction: 2282.0118617777894
 RMSE Score on real values for 26 days prediction: 2365.6200570422925
 RMSE Score on real values for 27 days prediction: 2452.4968810226264
 RMSE Score on real values for 28 days prediction: 2574.1821155188804
 RMSE Score on real values for 29 days prediction: 2686.1545271384143
 RMSE Score on real values for 30 days prediction: 2791.1806517859445

Figure 21. RMSE Values Obtained for Consecutive Days

It can be inferred from the RMSE values that as the number of days extended, the error increased, indicating a decrease in the model's efficiency.

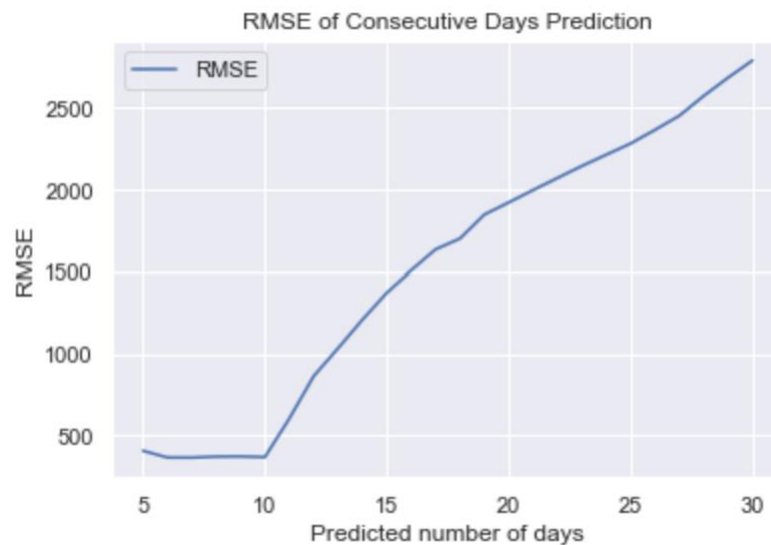


Figure 22. RMSE Graph of Consecutive Days (Authors own calculation).

5.3.Recommendation for Future study

Addressing these limitations in future research could involve dynamic modeling approaches, inclusion of more real-time data, or integrating qualitative analyses to better capture the complex nature of cryptocurrency markets.

The overall study aimed at predicting Bitcoin prices employed Ridge Regression, leveraging a comprehensive dataset that included technical, economic, on-chain, and market sentiment metrics. This approach has demonstrated considerable efficacy in capturing the multifaceted influences on Bitcoin's volatile market price. The Ridge Regression model, known for its ability to handle multicollinearity and mitigate overfitting through regularization, provided a robust framework for forecasting, yielding predictions that closely aligned with actual market behaviors, despite an inherent error margin. Moreover, the exploration into Bayesian Regression offered an alternative predictive lens, integrating prior knowledge and accommodating data uncertainty more fluidly. This dual-model analysis not only heightened the understanding of predictive dynamics in cryptocurrency markets but also highlighted the importance of continuous model tuning and adaptation to evolving market conditions. For traders, these insights are invaluable for devising more informed strategies. Policy makers could utilize the findings to better understand the economic and sentiment-driven dynamics of cryptocurrency markets, potentially guiding regulatory frameworks. Academically, the study paves the way for future research to incorporate a broader array of data points and model adjustments, aiming for ever more accurate predictive models.

Overall, while the study faced limitations such as potential sensitivity to outliers and the static nature of the dataset, it lays a strong foundation for future exploration and offers significant contributions to both practical trading applications and the academic understanding of financial modelling in the volatile sphere of cryptocurrency markets.

REFERENCES

- Agarwal, J. D., Agarwal, M., Agarwal, A., & Agarwal, Y. (2018). The Theory of Money, Wealth and Efficient Currency Markets: Modeling M5 as Money Supply with Crypto-Currency. *Agarwal JD, Manju Agarwal, Aman Agarwal and Yamini Agarwal*, 405-456.
- Agosto, A., Cerchiello, P., & Pagnottoni, P. (2022). Sentiment, Google queries and explosivity in the cryptocurrency market. *Physica A: Statistical Mechanics and its Applications*, 605, 128016.
- Ahoniemi, K. (2006). Modeling and forecasting implied volatility: An econometric analysis of the VIX index. Helsinki: Helsinki Center of Economic Research.
- Akbar, N. A., Muneer, A., ElHakim, N., & Fati, S. M. (2021). Distributed hybrid double-spending attack prevention mechanism for proof-of-work and proof-of-stake blockchain consensuses. *Future Internet*, 13(11), 285.
- AKGÜL, A., ŞAHİN, E. E., & ŞENOL, F. Y. (2022). Blockchain-based cryptocurrency price prediction with chaos theory, onchain analysis, sentiment analysis and fundamental-technical analysis. *Chaos Theory and Applications*, 4(3), 157-168.
- Alexander, C., & Imeraj, A. (2019). The Bitcoin VIX and its variance risk premium. *Forthcoming in the Journal of Alternative Investments*.
- Almeida, J., Tata, S., Moser, A., & Smit, V. (2015). Bitcoin prediciton using ANN. *Neural networks*, 7, 1-12.
- Alonso-Monsalve, S., Suárez-Cetrulo, A. L., Cervantes, A., & Quintana, D. (2020). Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. *Expert Systems with Applications*, 149, 113250.
- Al-Yahyaee, K. H., Rehman, M. U., Mensi, W., & Al-Jarrah, I. M. W. (2019). Can uncertainty indices predict Bitcoin prices? A revisited analysis using partial and multivariate wavelet approaches. *The North American Journal of Economics and Finance*, 49, 47-56.
- Amjad, M., & Shah, D. (2017, February). Trading Bitcoin and online time series prediction. In *NIPS 2016 time series workshop* (pp. 1-15). PMLR.
- Atsalakis, G. S., Atsalaki, I. G., Pasiouras, F., & Zopounidis, C. (2019). Bitcoin price forecasting with neuro-fuzzy techniques. *European Journal of Operational Research*, 276(2), 770-780.

- Bandi, F. M., & Perron, B. (2006). Long memory and the relation between implied and realized volatility. *Journal of Financial Econometrics*, 4(4), 636-670.
- Baur, D. G., Hong, K., & Lee, A. D. (2018). Bitcoin: Medium of exchange or speculative assets?. *Journal of International Financial Markets, Institutions and Money*, 54, 177-189.
- Bergesen, Ø., & Palm, L. M. (2018). *An exploratory study of initial coin offerings: A better understanding of the ICO market and its fraudulent and unregulated nature* (Master's thesis, Universitetet i Agder; University of Agder).
- Bhardwaj, A., Narayan, Y., & Dutta, M. (2015). Sentiment analysis for Indian stock market prediction using Sensex and nifty. *Procedia computer science*, 70, 85-91.
- Bouri, E., Gupta, R., Tiwari, A. K., & Roubaud, D. (2017). Does Bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. *Finance Research Letters*, 23, 87-95.
- Brauneis, A., Mestel, R., Riordan, R., & Theissen, E. (2019, June). A high-frequency analysis of Bitcoin markets. In *Proceedings of Paris December 2019 Finance Meeting EUROFIDAI-ESSEC*.
- Chan, W. H., Le, M., & Wu, Y. W. (2019). Holding Bitcoin longer: The dynamic hedging abilities of Bitcoin. *The Quarterly Review of Economics and Finance*, 71, 107-113.
- Chen, C. W., Gerlach, R., Hwang, B. B., & McAleer, M. (2012). Forecasting value-at-risk using nonlinear regression quantiles and the intra-day range. *International Journal of Forecasting*, 28(3), 557-574.
- Cheque Cerda, G., & L. Reutter, J. (2019, May). Bitcoin price prediction through opinion mining. In *Companion Proceedings of The 2019 World Wide Web Conference* (pp. 755-762).
- Clements, A., & Fuller, J. (2012). *Forecasting increases in the VIX: A time-varying long volatility hedge for equities* (No. 88). National Centre for Econometric Research.
- Conrad, C., Custovic, A., & Ghysels, E. (2018). Long-and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis. *Journal of Risk and Financial Management*, 11(2), 23.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2), 174-196.

- Dongrey, S. (2022). Study of market indicators used for technical analysis. *International Journal Of Engineering And Management Research*, 12(2), 64-83.
- Dotsis, G., Psychoyios, D., & Skiadopoulos, G. (2007). An empirical comparison of continuous-time models of implied volatility indices. *Journal of Banking & Finance*, 31(12), 3584-3603.
- Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar—A GARCH volatility analysis. *Finance research letters*, 16, 85-92.
- Dyhrberg, A. H., Foley, S., & Svec, J. (2018). How investible is Bitcoin? Analyzing the liquidity and transaction costs of Bitcoin markets. *Economics Letters*, 171, 140-143.
- Erdas, M. L., & Caglar, A. E. (2018). Analysis of the relationships between Bitcoin and exchange rate, commodities and global indexes by asymmetric causality test. *Eastern Journal of European Studies*, 9(2).
- Faccini, R., Konstantinidi, E., Skiadopoulos, G., & Sarantopoulou-Chiourea, S. (2019). A new predictor of US real economic activity: The S&P 500 option implied risk aversion. *Management Science*, 65(10), 4927-4949.
- Faghih Mohammadi Jalali, M., & Heidari, H. (2020). Predicting changes in Bitcoin price using grey system theory. *Financial Innovation*, 6, 1-12.
- Fernandes, M., Medeiros, M. C., & Scharth, M. (2014). Modeling and predicting the CBOE market volatility index. *Journal of Banking & Finance*, 40, 1-10.
- Gaies, B., Nakhli, M. S., Sahut, J. M., & Schweizer, D. (2023). Interactions between investors' fear and greed sentiment and Bitcoin prices. *The North American Journal of Economics and Finance*, 67, 101924.
- Gao, X., Huang, W., & Wang, H. (2021). Financial twitter sentiment on Bitcoin return and high-frequency volatility. *Virtual Economics*, 4(1), 7-18.
- Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D., & Giaglis, G. M. (2015). Using time-series and sentiment analysis to detect the determinants of Bitcoin prices. *Available at SSRN 2607167*.
- Georgoula, I., Pournarakis, D., Bilanakos, C., Sotiropoulos, D., & Giaglis, G. M. (2015). Using time-series and sentiment analysis to detect the determinants of Bitcoin prices. *Available at SSRN 2607167*.
- Gonçalves, J. Q., & Alves, J. (2022). How Money relates to value? An empirical examination on Gold, Silver and Bitcoin.

- Guo, T., Bifet, A., & Antulov-Fantulin, N. (2018, November). Bitcoin volatility forecasting with a glimpse into buy and sell orders. In *2018 IEEE international conference on data mining (ICDM)* (pp. 989-994). IEEE.
- Hagenau, M., Liebmann, M., & Neumann, D. (2013). Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision support systems*, 55(3), 685-697.
- Hakim das Neves, R. (2020). Bitcoin pricing: impact of attractiveness variables. *Financial Innovation*, 6(1), 21.
- Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?. *Journal of applied econometrics*, 20(7), 873-889.
- Hansun, S. (2013, November). A new approach of moving average method in time series analysis. In *2013 conference on new media studies (CoNMedia)* (pp. 1-4). IEEE.
- Hassan, M. R., Ramamohanarao, K., Kamruzzaman, J., Rahman, M., & Hossain, M. M. (2013). A HMM-based adaptive fuzzy inference system for stock market forecasting. *Neurocomputing*, 104, 10-25.
- Jagannath, N., Barbulescu, T., Sallam, K. M., Elgendi, I., Mcgrath, B., Jamalipour, A., ... & Munasinghe, K. (2021). An on-chain analysis-based approach to predict ethereum prices. *IEEE Access*, 9, 167972-167989.
- Jain, A., Tripathi, S., Dwivedi, H. D., & Saxena, P. (2018, August). Forecasting price of cryptocurrencies using tweets sentiment analysis. In *2018 eleventh international conference on contemporary computing (IC3)* (pp. 1-7). IEEE.
- Jakub, B. (2015). Does Bitcoin follow the hypothesis of efficient market. *International Journal of Economic Sciences*, 4(2), 10-23.
- Jang, H., & Lee, J. (2017). An empirical study on modeling and prediction of Bitcoin prices with bayesian neural networks based on blockchain information. *IEEE access*, 6, 5427-5437.
- Jiao, Y., & Jakubowicz, J. (2017, December). Predicting stock movement direction with machine learning: An extensive study on S&P 500 stocks. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 4705-4713). IEEE.
- Johnson, A., Davis, E., Nasir, W., & Brown, M. (2024). Leveraging Sentiment Lexicon in Sentiment Detection.

- Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics letters*, 158, 3-6.
- Kiaepkos, G. (2021). The impact of scheduled macroeconomic announcements on implied volatility index.
- Kim, Y. B., Kim, J. G., Kim, W., Im, J. H., Kim, T. H., Kang, S. J., & Kim, C. H. (2016). Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PloS one*, 11(8), e0161197.
- Kjærland, F., Khazal, A., Krogstad, E. A., Nordstrøm, F. B., & Oust, A. (2018). An analysis of Bitcoin's price dynamics. *Journal of Risk and Financial Management*, 11(4), 63.
- Klein, T., Thu, H. P., & Walther, T. (2018). Bitcoin is not the New Gold—A comparison of volatility, correlation, and portfolio performance. *International Review of Financial Analysis*, 59, 105-116.
- Konstantinidi, E., Skiadopoulos, G., & Tzagkaraki, E. (2008). Can the evolution of implied volatility be forecasted? Evidence from European and US implied volatility indices. *Journal of Banking & Finance*, 32(11), 2401-2411.
- Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, 12(3), 445-475.
- Kristjanpoller, W., & Minutolo, M. C. (2018). A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis. *Expert Systems with Applications*, 109, 1-11.
- Lahmiri, S., Bekiros, S., & Salvi, A. (2018). Long-range memory, distributional variation and randomness of Bitcoin volatility. *Chaos, Solitons & Fractals*, 107, 43-48.
- Li, Q., Jiang, L., Li, P., & Chen, H. (2015, February). Tensor-based learning for predicting stock movements. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 29, No. 1). China-Rios, M., Borrego-Obrador, I., Franco-Salvador, M., Rangel, F., & Rosso, P. (2023). Profiling Cryptocurrency Influencers with Few-shot Learning.
- Li, T. R., Chamrajnagar, A. S., Fong, X. R., Rizik, N. R., & Fu, F. (2019). Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. *Frontiers in Physics*, 7, 98.

- Lin, X., Meng, Y., & Zhu, H. (2023). How connected is the crypto market risk to investor sentiment?. *Finance Research Letters*, 56, 104177.
- Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.
- Madan, T., Sagar, S., & Virmani, D. (2020, December). Air quality prediction using machine learning algorithms—a review. In *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)* (pp. 140-145). IEEE.
- Mallqui, D. C., & Fernandes, R. A. (2019). Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques. *Applied Soft Computing*, 75, 596-606.
- McNally, S., Roche, J., & Caton, S. (2018, March). Predicting the price of Bitcoin using machine learning. In *2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP)* (pp. 339-343). IEEE.
- McNally, S., Roche, J., & Caton, S. (2018, March). Predicting the price of Bitcoin using machine learning. In *2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP)* (pp. 339-343). IEEE.
- Mohapatra, S., Ahmed, N., & Alencar, P. (2019, December). KryptoOracle: a real-time cryptocurrency price prediction platform using twitter sentiments. In *2019 IEEE international conference on big data (Big Data)* (pp. 5544-5551). IEEE.
- Mokni, K., Bouteska, A., & Nakhli, M. S. (2022). Investor sentiment and Bitcoin relationship: A quantile-based analysis. *The North American Journal of Economics and Finance*, 60, 101657.
- Mokni, K., Bouteska, A., & Nakhli, M. S. (2022). Investor sentiment and Bitcoin relationship: A quantile-based analysis. *The North American Journal of Economics and Finance*, 60, 101657.
- Nam, K., & Seong, N. (2019). Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market. *Decision Support Systems*, 117, 100-112.
- Nava-Solis, A. R., & Treviño-Saldívar, E. J. (2021, July). On-chain metrics and technical analysis in cryptocurrency markets. In *XX SIGEF Congress- Harnessing Complexity through Fuzzy Logic* (pp. 122-129). Cham: Springer International Publishing.

- Nayak, R. K., Mishra, D., & Rath, A. K. (2015). A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices. *Applied Soft Computing*, 35, 670-680.
- Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: the role of technical indicators. *Management science*, 60(7), 1772-1791.
- Othman, A. H. A., Kassim, S., Rosman, R. B., & Redzuan, N. H. B. (2020). Prediction accuracy improvement for Bitcoin market prices based on symmetric volatility information using artificial neural network approach. *Journal of Revenue and Pricing Management*, 19, 314-330.
- Peng, S. (2013). Bitcoin: Cryptography, Economics, and the Future. *Senior Capstone Thesis School of Engineering and Applied Science*.
- Rashid, A., Fayyaz, M., & Karim, M. (2019). Investor sentiment, momentum, and stock returns: an examination for direct and indirect effects. *Economic research-Ekonomska istraživanja*, 32(1), 2638-2656.
- Rebane, J., Karlsson, I., Papapetrou, P., & Denic, S. (2018). Seq2Seq RNNs and ARIMA models for cryptocurrency prediction: A comparative study. In *SIGKDD Fintech'18, London, UK, August 19-23, 2018*.
- Rodríguez-González, A., García-Crespo, Á., Colomo-Palacios, R., Iglesias, F. G., & Gómez-Berbís, J. M. (2011). CAST: Using neural networks to improve trading systems based on technical analysis by means of the RSI financial indicator. *Expert systems with Applications*, 38(9), 11489-11500.
- Saleem, T., Yaqub, U., & Zaman, S. (2024). Twitter sentiment analysis and Bitcoin price forecasting: implications for financial risk management. *The Journal of Risk Finance*.
- Shahzad, S. J. H., Bouri, E., Roubaud, D., Kristoufek, L., & Lucey, B. (2019). Is Bitcoin a better safe-haven investment than gold and commodities?. *International Review of Financial Analysis*, 63, 322-330.
- Stenqvist, E., & Lönnö, J. (2017). Predicting Bitcoin price fluctuation with Twitter sentiment analysis.
- Sul, H. K., Dennis, A. R., & Yuan, L. (2017). Trading on twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, 48(3), 454-488.
- Suryawijaya, T., Laurent, F. A., & Berger, M. (2023). Optimizing AltCoin (ETH) Price Synchronization During the Bitcoin Dominance Era: A Comprehensive Phase

- Analysis with Vector Error Correction Model. *Economics, Business, Accounting & Society Review*, 2(3), 216-229.
- Tschorsch, F., & Scheuermann, B. (2016). Bitcoin and beyond: A technical survey on decentralized digital currencies. *IEEE Communications Surveys & Tutorials*, 18(3), 2084-2123.
- Urquhart, A., & Zhang, H. (2019). Is Bitcoin a hedge or safe haven for currencies? An intraday analysis. *International Review of Financial Analysis*, 63, 49-57.
- Yoldas Ayan, B. (2023). *A Comparison of Traditional Models and Machine Learning Techniques for Volatility Forecasting* (Doctoral dissertation, Dublin Business School).
- Zhang, L., Aggarwal, C., & Qi, G. J. (2017, August). Stock price prediction via discovering multi-frequency trading patterns. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 2141-2149).
- Zhang, S., Li, M., & Yan, C. (2022). The empirical analysis of Bitcoin price prediction based on deep learning integration method. *Computational Intelligence and Neuroscience*, 2022.
- Zhao, D., Aili, A., Yin, X., Tan, G., & Yang, R. (2019). Roof-integrated radiative air-cooling system to achieve cooler attic for building energy saving. *Energy and Buildings*, 203, 109453.
- Zhou, S. (2021). Exploring the driving forces of the Bitcoin currency exchange rate dynamics: an EGARCH approach. *Empirical Economics*, 60(2), 557-606.