Article

Not peer-reviewed version

Assessing Machine Learning Performance in Financial Forecasting and AI-Driven Customer Service Systems

Gabriela Vasileva , Dilyana Karova , Mariyan Milev * , Penko Mitev

Posted Date: 15 April 2026

doi: 10.20944/preprints202604.1107.v1

Keywords: cryptocurrency market; forecasting; AI; machine learning; regression models

Article

Assessing Machine Learning Performance in Financial Forecasting and AI-Driven Customer Service Systems

Gabriela Vasileva [1], Dilyana Karova [2], Mariyan Milev [2,*] and Penko Mitev [3]

[1] Varna Free University, Varna, Bulgaria
[2] Department of Statistics and Econometrics, Faculty of Economics and Business Administration, 1113, 125 Tsarigradsko Shosse Blvd., bl. 3, Sofia University "St.Kliment Ohridski&", 39, 1504 Sofia, 15 Tsar Osvoboditel Blvd., Bulgaria
[3] Faculty of Mechanical Engineering, Mechanical and Instrument Engineering, Technical University of Sofia, Branch Plovdiv, 4000 Plovdiv, Bulgaria
* Correspondence: m.milev@feb.uni-sofia.bg

Abstract

This study examines the multifaceted application of machine learning and artificial intelligence (AI) in two key, dynamically developing sectors: cryptocurrency market capitalisation forecasting and customer service optimisation. An analysis of the effectiveness of various regression models (Linear, Lasso, and Decision Tree Regression) in predicting the market capitalisation of 3 leading cryptocurrencies shows that a model's success is highly dependent on the specific characteristics of the asset. While linear models achieve exceptional accuracy (R^2>0.99) for most major and liquid cryptocurrencies, nonlinear approaches like Decision Tree Regression prove superior for assets with more complex and nonlinear market dynamics, highlighting the need for a flexible approach to model selection. In parallel, the study analyses the implementation of AI in customer service, reviewing chat communication data with the AI assistant "Naomi" (January 26–February 8, 2025). The AI "Naomi" demonstrated high overall effectiveness in chat communication, resolving over 60% of inquiries. However, a significant number of unresolved chats due to customer inactivity or AI limitations indicate areas for further optimisation. In conclusion, the effective application of AI and machine learning requires a strategic approach tailored to the specific field. The key to success lies in careful model selection, prioritising technical reliability, and continuous adaptation and optimisation based on empirical data and a deep understanding of AI's limitations.

Keywords: cryptocurrency market; forecasting; AI; machine learning; regression models

---

1. Introduction

Artificial intelligence is everywhere: in our phones, cars, shopping experiences, healthcare, banking, media, education [18], etc. With the increasing pace of digital transformation, artificial intelligence (AI) is solidifying its position as a strategic tool for improving business processes, decision-making, and customer service optimisation. No wonder corporate directors, CEOs, vice presidents, managers, team leaders, entrepreneurs, investors, coaches, and policy makers are anxiously racing to learn about AI: they all realise it is about to fundamentally change their businesses, by Agrawal A. [2].

Changing form of the business terms and work forces, the way of doing business by using new technologies will have serious impacts on the daily business life and deriving from these on countries and on world economics by Dirican C. [6], while the development of algorithms for machine learning, natural language processing, and predictive analysis offers opportunities for automation and precision that were recently unattainable for human resources in real time.

 doi:10.20944/preprints202604.1107.v1

Two key areas where AI demonstrates its potential are: forecasting dynamic and nonlinear economic systems like the cryptocurrency market, and automating customer support through intelligent assistants in multichannel environments. This study combines an analytical and empirical approach to examine the effectiveness of AI in two contrasting, yet conceptually related, contexts: processing and predicting cryptocurrency market values using regression models and implementing AI in customer service through chatbot communication.

The first part of the study shows the real-world behaviour of AI in an operational business environment through chat communication with an intelligent assistant, "Naomi," in the service sector. The data covers a period where key metrics such as the percentage of successfully resolved cases, the percentage of technical errors, and the interaction between AI and customer engagement are tracked.

The second part focuses on applying classical and tree-based regression models (Linear Regression, Lasso, and Decision Tree Regression) to forecast the market capitalisation of cryptocurrencies using historical data from CoinMarketCap. The main goal is to compare the predictive power of different algorithms and identify significant explanatory variables in this highly volatile market context.

The combination of these two research lines aims not only to demonstrate the wide spectrum of AI's applicability but also to highlight the need for a specific approach in designing, training, and implementing intelligent systems. By simultaneously considering market predictability and effectiveness in communication processes, this article offers a holistic view of the challenges and opportunities associated with the real-world use of AI in various economic sectors.

2. Data and Methods

The approach used in this research focuses on data collection and the application of analytical techniques to assess the impact of artificial intelligence on the cryptocurrency market and businesses in the cleaning and landscaping services sector.

The methodology comprises two main stages: collecting data from real companies and the cryptocurrency market, and applying various analytical techniques to forecast trends and assess the impact of AI. A mathematical model was created to measure linear and simple dependencies for analysis and evaluation.

2.1. Data Collection for Chatbot "Naomi"

The first part of the information will be collected from the company Fantastic Services, a company in the service sector in the United Kingdom, where the effectiveness of implementing artificial intelligence in business processes is analysed. The analysis is based on empirical data collected from the implementation of an AI assistant, "Naomi," in chat communication over a specific period of study with exact dates: 26.01.2025 - 08.02.2025.

The inclusion criteria for this study were all chat interactions with clients from the United Kingdom that occurred within the specified two-week period. No explicit exclusion criteria were provided, but the analysis was restricted to chat data, and other forms of client communication or work processes were not included in the dataset.

Quantitative data includes: the daily total number of chats with customers from the UK, the number of chats resolved by the AI (with or without generating a task), the number of unresolved chats (due to the client or due to AI limitations), and the corresponding percentage ratios of these categories to the total number of chats. The number of generated tasks is also tracked.

2.2. Mathematical Model for "Naomi's" Effectiveness

This model describes and allows for the forecasting of the main outcomes of chat interactions with the AI assistant "Naomi," based on the provided data.

Key Variables:

For each observed day D:

- CD: Total number of chats.
- RND, D: Number of chats resolved by Naomi without generating a task.
- RTD, D: Number of chats resolved by Naomi with at least one task generated.
- UCD, D: Number of unresolved chats due to the client.
- UNCD, D: Total number of unresolved chats (including UCD, D and those unresolved due to Naomi's limitations).
- TD: Total number of generated tasks.

Key Proportions (Outcome Probabilities):

For a specific day "D," we calculate the proportions of the different outcomes relative to the total number of chats "CD":

- Proportion of chats resolved without a task (PRND, D): PRND, D=CDRND, D
- Proportion of chats resolved with a task (PRTD, D): PRTD, D=CDRTD, D
- Proportion of unresolved chats due to the client (PUCD, D): PUCD, D=CDUCD, D
- Proportion of total unresolved chats (PUNCD, D): PUNCD, D=CDUNCD, DTask Generation Coefficient:
- Average number of tasks per resolved chat with a task (αD): αD=RTD, DTD (for RTD, D>0)

This coefficient shows the average number of tasks generated for each successfully resolved chat that requires a task.

Application and Forecasting of Future Results:

To forecast results for a future period, we use the average values of the calculated proportions and the coefficient for the entire observed specific period (denoted with a bar above, e.g., PRN¯).

If the expected total number of chats for the future period is Cforecast, then:

- Expected number of chats resolved without a task: E[RND]=PRN¯×Cforecast
- Expected number of chats resolved with a task: E[RTD]=PRT¯×Cforecast
- Expected number of unresolved chats due to the client: E[UCD]=PUC¯×Cforecast
- Expected total number of unresolved chats: E[UNCD]=PUNC¯×Cforecast
- Expected total number of generated tasks: E[T]=E[RTD]×α¯

The model provides a simple but effective framework for understanding the distribution of Naomi's chat outcomes and allows for quick calculation of key performance metrics and basic forecasts. Its main limitation is that it assumes static probabilities and does not account for dynamic changes over time or external factors.

2.3. Data Statistics and Visualisation of Naomi Chatbot

In Figure 1, between January 26 and February 8, 2025, 65–83% of client chats were successfully resolved, while unresolved cases ranged from 17% to 31%. The number of tasks generated per day varied between 20 and 71, with the majority of successful cases requiring at least one task.

In this context, AI can perform not only simple tasks but also complicated tasks as long as these tasks are based on predetermined rules or standards that are highly standardised in a mass of codifiable data by Benhamou S. [4], as demonstrated in the work of the AI assistant "Naomi", which showed high overall effectiveness in chat communication.

On average for the period, the total percentage of resolved chats (with or without a task) consistently exceeded 60%, often reaching more than 70-80%. Approximately 64% of all chats were resolved with at least one task generated, highlighting Naomi's ability to take action. However, a significant share of chats remained unresolved: an average of 10.7% due to the client's abandonment and about 25.6% overall unresolved (including both client- and AI-initiated failures). This indicates that, despite being highly effective, Naomi encounters limitations with more complex inquiries or insufficient client engagement.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 April 2026 doi:10.20944/preprints202604.1107.v1

- General statistics for the period between 26.01 - 08.02.2025

DATE	Working hours	UK clients chats only	Resolved by Naomi with NO task	Resolved by Naomi with NO task - %	Resolved by Naomi with min 1 task/ Client profile task	Resolved by Naomi with min 1 task - %	All Resolved chats with or without a task - %	All resolved chats with or without a task + chats unresolved due to client	Unresolved due to client	Unresolved due to Naomi	All Unresolved chats (by Naomi + by cust) - %	Actual tasks generated
26.01.2025	8h+	34	4	11.76%	18	52.94%	64.71%	73.53%	3	3	17.65%	20
27.01.2025	8h+	89	9	10.11%	57	64.04%	74.16%	86.52%	11	5	17.98%	60
28.01.2025	8h+	81	6	7.41%	37	45.68%	53.09%	64.20%	9	5	17.28%	39
29.01.2025	8h+	67	7	10.45%	44	65.67%	76.12%	94.03%	12	6	26.87%	61
30.01.2025	8h+	65	4	6.15%	49	75.38%	81.54%	90.77%	6	5	16.92%	54
31.01.2025	8h+	76	10	13.16%	55	72.37%	85.53%	94.74%	7	9	21.05%	71
01.02.2025	8h+	74	8	10.81%	44	59.46%	70.27%	82.43%	9	12	28.38%	48
02.02.2025	8h+	53	12	22.64%	29	54.72%	77.36%	90.57%	7	2	16.98%	37
03.02.2025	8h+	61	6	9.84%	40	65.57%	75.41%	81.97%	4	11	24.59%	54
04.02.2025	8h+	79	2	2.53%	55	69.62%	72.15%	88.61%	13	9	27.85%	55
05.02.2025	8h+	61	9	14.75%	37	60.66%	75.41%	83.61%	5	9	22.95%	39
06.02.2025	8h+	47	4	8.51%	32	68.09%	76.60%	82.98%	3	8	23.40%	44
07.02.2025	8h+	72	4	5.56%	46	63.89%	69.44%	87.50%	13	9	30.56%	54
08.02.2025	8h+	47	8	17.02%	31	65.96%	82.98%	87.23%	2	7	19.15%	34

Figure 1. Statistics for the period from January 26 to February 8, 2025.

Key Conclusions and Future Directions:

AI and generative AI are transforming industries, but true reinvention requires more than adoption - it demands deep integration into strategy, processes and decision-making [1]. Generative AI's application in customer support includes analytics to anticipate, deflect, and address potential customer issues; chatbots to expand digital self-service offerings and automate interactions; algorithms to connect customers with the most appropriate representative; and knowledge assistant tools that help agents act more efficiently [3].

But in that research, the analysis of Generative AI implementation data revealed mixed results. The AI assistant "Naomi" shows strong performance in a chat environment, successfully resolving a large volume of customer inquiries and automating task generation. For future optimization, it is recommended to: conduct a detailed analysis of the unresolved chats with "Naomi" to identify the types of problems that exceed its capabilities; reduce the number of chats unresolved due to a lack of client engagement; and develop more precise metrics, especially for "Hit Rate," to ensure a clearer interpretation of the results.

2.4. Description of the Cryptocurrency Dataset and Its Applicability in the Study

This study uses a dataset that serves as a basis for building, testing, and comparing different predictive models for forecasting cryptocurrency prices. This includes daily historical trading and pricing information for crypto assets, extracted from CoinMarketCap. This is a leading platform for real-time market information. The dataset is publicly available on the Kaggle platform, provided by Rajkumar S., under the project "Cryptocurrency Historical Prices" [10].

It contains individual CSV files for each of the top cryptocurrencies by market capitalisation.

Study Period and Scope

The analysis covers three separate periods for different cryptocurrencies. The price data is shown in US dollars (USD), reflecting the global nature of cryptocurrency markets.

- Bitcoin (BTC) & Ethereum (ETH): The study period for both assets is from May 2017 to March 2021.
- Cosmos (ATOM): The study period is from March 2019 to May 2021.

Unit of Analysis and Metrics

The primary unit of analysis is the cryptocurrency asset, with a focus on three specific coins: Bitcoin, Ethereum, and Cosmos.

The key metrics analysed are:

- Price: Represented by a candlestick chart, which shows the opening, closing, high, and low prices for each trading period.

- Volume: The total volume of trading for each asset, represented by the bar chart in the lower panel of each figure. The volume is measured in units of the respective cryptocurrency (e.g., BTC, ETH, ATOM).

2.5. Data Files and Structure of the Cryptocurrency

Each cryptocurrency is stored in a separate .csv file named using the pattern coin_<CoinName>.csv. For example:

- coin_Bitcoin.csv
- coin_Ethereum.csv
- coin_Cardano.cvs
- The files are named using the format coin_<Name>.csv (e.g., coin_Bitcoin.csv, coin_Ethereum.csv, coin_Cardano.cvs) and contain the following columns:

Column	Description
Date	Date of observation (YYYY-MM-DD)
Open	Opening price for the day
High	Highest price for the day
Low	Lowest price for the day
Close	Closing price for the day
Volume	Volume of traded assets in USD
Marketcap	Total market capitalisation of the cryptocurrency in USD

All columns are either numeric or date-type, except for Name and Symbol, which are categorical identifiers dropped during preprocessing.

The quantitative data here represent the numerical variables used as input variables (predictors) or as target indicators in the models for forecasting cryptocurrency market capitalisation. These include the asset's opening, high, low, and closing prices for the corresponding period, trading volume, and market capitalisation itself. Additionally, metrics such as the **Mean Absolute Error (MAE)** and the **coefficient of determination (R^2)** are used to evaluate model performance. Different evaluation metrics are included in this study to measure the accuracy of the time series forecasting models. They give a complete explanation of how each of the models is in line with the accurate sales data. Prominent ones include Mean Absolute Error (MAE), and R-squared (R^2). Each of these metrics provides insight into the model's accuracy and reliability in Naskinova, Kolev et al. [9].

Qualitative data includes the type of cryptocurrency (e.g., Bitcoin, Ethereum), asset categories (such as Layer 1 or DeFi tokens), market cycle phases (accumulation, markup, correction), and the chosen forecasting models.

2.6. Data Preprocessing of the Cryptocurrency

Imputation refers to the process of replacing missing or incomplete values in the dataset with statistically or algorithmically derived estimates. In the context of cryptocurrency market data, imputation plays a crucial role, as gaps may arise due to interruptions in trading, inconsistencies across exchanges, or incomplete historical records.

The proper imputation ensures that predictive models, such as regression or decision tree approaches, are not biased by missing information and can operate on a coherent and continuous dataset. Depending on the characteristics of the data, the imputation involves simple techniques (e.g., mean or median substitution), time-series-oriented methods (e.g., forward or backwards filling, interpolation), or advanced machine learning approaches designed to capture nonlinear dependencies.

2.7. Applicability and Methodological Framework of the Cryptocurrency

The data is used to train and compare three different regression analysis models: **Linear Regression**, **Lasso Regression**, and **Decision Tree Regression**. Each of these models has its specific characteristics.

Predictive models are built for each cryptocurrency, with the closing price (**Close**) as the primary target variable. The independent variables used include **Open**, **High**, **Low**, **Volume**, and **Marketcap**. Models are evaluated by splitting the data into training and test sets and calculating standard regression metrics: **Mean Squared Error (MSE)**, **Mean Absolute Error (MAE)**, and the **coefficient of determination (R^2) as** in Naskinova, Kolev et al. [9].

2.8. Creating a Mathematical Model

Econometric models suitable for measuring linear or simple dependencies will be used to determine how the increase in AI investments affects profitability.

The economic effect will be evaluated, where:

$$Y = \alpha + \beta X$$

Here, Y is the effect (e.g., profitability), X are the variables (e.g., investments, time), α and β are the coefficients.

Linear Regression

In the context of cryptocurrency price prediction, linear regression can be used to analyse the relationship between factors like daily trading volume, opening price, or market capitalisation and the target variable - usually the asset's closing price.

Mathematical Formulation

In the one-dimensional case (univariate regression), the model is expressed as:

$$\hat{y} = \beta_0 + \beta_1 x,$$

With the standard denotations: y is the predicted value (e.g., closing price), x is the input variable (e.g., trading volume), β_0 is the intercept, and β_1 is the slope coefficient.

In the multidimensional case (multivariate regression):

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n,$$

or in vectorised form: $Y = X\beta$, where X is the matrix of input features (size n×p), β is the vector of regression coefficients, and $\hat{y}$ is the vector of predicted values.

The goal is to find the values of β that minimise the difference between the predicted and actual values. This is accomplished using the Ordinary Least Squares (OLS) method, which minimises the following function:

$$min_{\beta_i} \sum_i^n (y_i - \hat{y}_i)^2$$

Linear regression offers high interpretability and low computational complexity, making it applicable as a baseline model for forecasting in dynamic environments like the cryptocurrency market. The resulting metric values indicate the degree of correspondence between predicted and actual values and serve as a quantitative assessment of the model's predictive capability.

2.9. Decision Tree Model for Forecasting Cryptocurrency Market Values

The tree is built through recursive binary splitting, where at each level, the variable and threshold value that led to the smallest loss (variance or MSE) in the child nodes are chosen. The process continues until predefined stopping conditions are met, such as a maximum tree depth or a minimum number of examples in a leaf node.

Formally, for a split S, the goal is to minimise:

$$Loss(S) = \frac{|S_1|}{|D|} Var(S_1) + \frac{|S_2|}{|D|} Var(S_2),$$

where S_1 and S_2 are the results of splitting the data D by a given variable and threshold.

In the context of cryptocurrencies, this model stands out for its ability to capture nonlinear dependencies and threshold effects, which are frequently observed in dynamic markets.

2.10. Lasso (Least Absolute Shrinkage and Selection Operator)

Lasso is a modification of linear regression that introduces L1-regularisation to constrain the values of the coefficients. This allows for both a reduction in the risk of overfitting and an automatic selection of significant variables by setting the coefficients of less important ones to zero.

Mathematical Formulation

The objective of Lasso is to minimise the following loss function:

$$min_{\beta_i} \sum_i^n (y_i - X_i\beta)^2 + \lambda \sum_j^p |\beta_i|,$$

where y_i are the observed values, X_i is the matrix of input variables, β are the model coefficients, and λ is the regularisation parameter, which controls the severity of the penalty (the larger λ is, the more coefficients are set to zero).

In the context of crypto markets, Lasso is especially useful when dealing with multiple correlated indicators (volume, volatility, market capitalisation, etc.). The model automatically selects the most informative variables, thereby improving its ability to generalise to new data.

3. Data Statistics Visualisation of Cryptocurrency Volatility and Trading Patterns

3.1. Bitcoin

Figure 2 presents a **candlestick chart** such as [14] showing the historical price and trading volume movements for the cryptocurrency Bitcoin (BTC) from May 2017 to June 2021. The chart consists of two main sections:

Price Dynamics (Top Section):

Bitcoin went through several distinct macro-cycles. The first major bull market occurred in late 2017, when the price peaked just under USD 20,000, followed by a sharp drop and an extended consolidation phase from 2018 - 2019, with prices mainly ranging from $3,000 - $10,000 USD. A new upward trend began in mid-2020, which evolved into a parabolic rally, reaching an all-time high of over USD 60,000 in April 2021. This was followed by a sharp correction to the $30,000 - USD 40,000 range, illustrating Bitcoin's characteristic "boom and bust" cyclical volatility, often accompanied by events such as halving, institutional entry, and macroeconomic factors.

Trading Volume (Bottom Section):

Volume remained relatively stable but low during 2017 and 2018, with a noticeable increase from 2019 onward. In 2020, there was a gradual rise in baseline volume, with distinct peaks in early 2021 that coincided with Bitcoin's price surge. A particularly anomalous volume spike was observed around the first quarter of 2021, likely due to a single high-volume event like a large-scale liquidation, an exchange collapse, or coordinated institutional activity. The overall trend of increasing volume reflects the market's maturity, growing institutional participation, and improved trading infrastructure.

Bitcoin's complex price behaviour and volume dynamics make it an ideal candidate for multi-feature machine learning models [15], such as time-series transformer-based models with external macroeconomic inputs (e.g., interest rates, inflation data, institutional flows, stock market returns [16,20]). Its historical cycles also make it suitable for market regime-switching models or long/short-term LSTM networks with attention, which can identify phases of accumulation, markup, distribution and differences such as [17].

From an investment strategy perspective, Bitcoin serves as an archetype for predictive models in the crypto markets, providing a benchmark for volatility and a leading indicator of market sentiment and structural changes in investment paradigms for digital assets.

 doi:10.20944/preprints202604.1107.v1

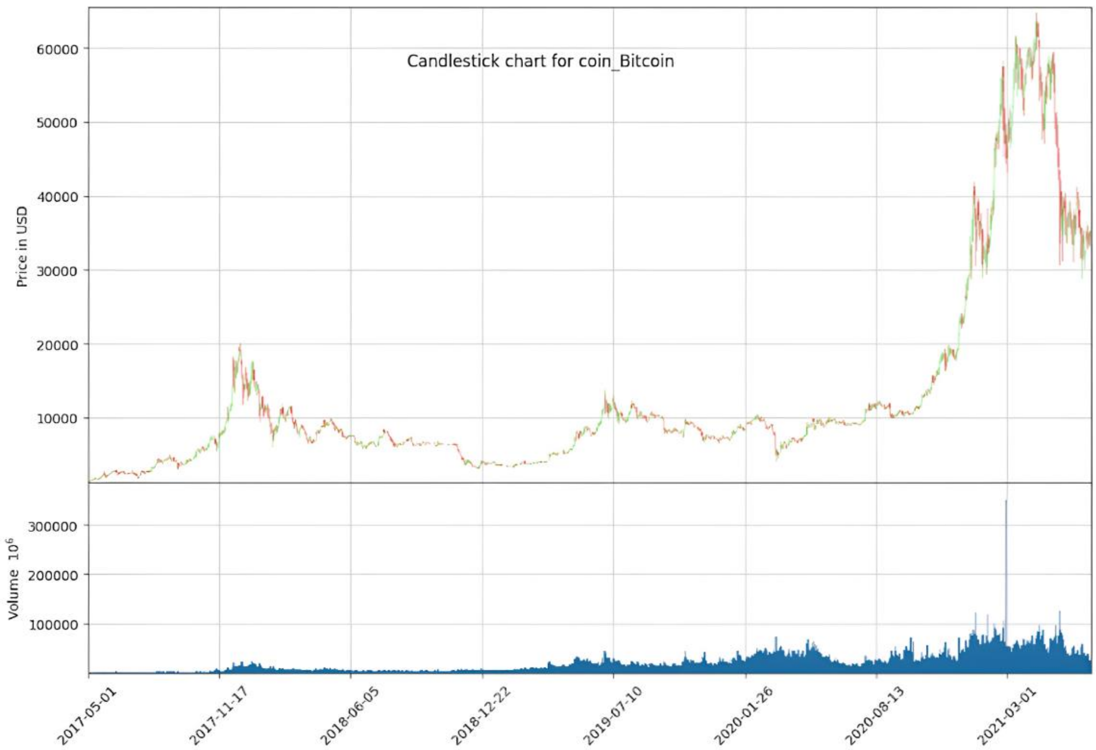

Figure 2. Bitcoin Cryptocurrency Volatility and Trading Patterns.

3.2. Cosmos

Figure 3 presents a **candlestick chart** showing the historical price and trading volume movements for the cryptocurrency Cosmos (ATOM) from March 2019 to July 2021. The chart includes two main sections:

Price Dynamics (Top Section)

Cosmos demonstrated a moderately volatile but gradually developing price structure during its initial trading period. From 2019 to mid-2020, ATOM fluctuated in the $2 to USD 6 range, showing low volatility and contained behaviour. A clear upward breakout occurred in late 2020, accelerating sharply in early 2021, with the price reaching a peak of nearly USD 30 in May 2021. This was followed by a sharp correction to around $10–USD 12 before moderate recovery attempts began. The price trend reflects the increasing utility of Cosmos in the field of interoperability and cross-chain infrastructure, which attracted speculative and institutional interest during broader market rallies.

Trading Volume (Bottom Section):

Trading volume progressively increased throughout the observed period. While volume was low and stable from 2019 to early 2020, a significant increase began in the second half of 2020, accompanying the price rise. In early 2021, a jump in volume was observed, coinciding with the parabolic price movement, with peaks occurring in the February–May 2021 period. These volume bursts are directly related to the price peaks and indicate intense market participation and phases of speculative trading. After the peak, volume dropped but remained significantly above historical baseline levels.

Cosmos represents a smooth transition from a consolidation phase to a dynamic driven by high volatility and momentum, which makes it ideal for time-series classification and trend-based forecasting models, such as RNN networks with attention mechanisms or transformer-based architectures. The distinct relationship between price and volume supports the inclusion of volume data as a leading indicator in predictive models.

From an investment strategy perspective, Cosmos's trajectory highlights the importance of recognising the narratives around infrastructure cryptocurrencies and volume-driven breakouts as signals for rebalancing and dynamic portfolio allocation in rapidly developing crypto sectors.

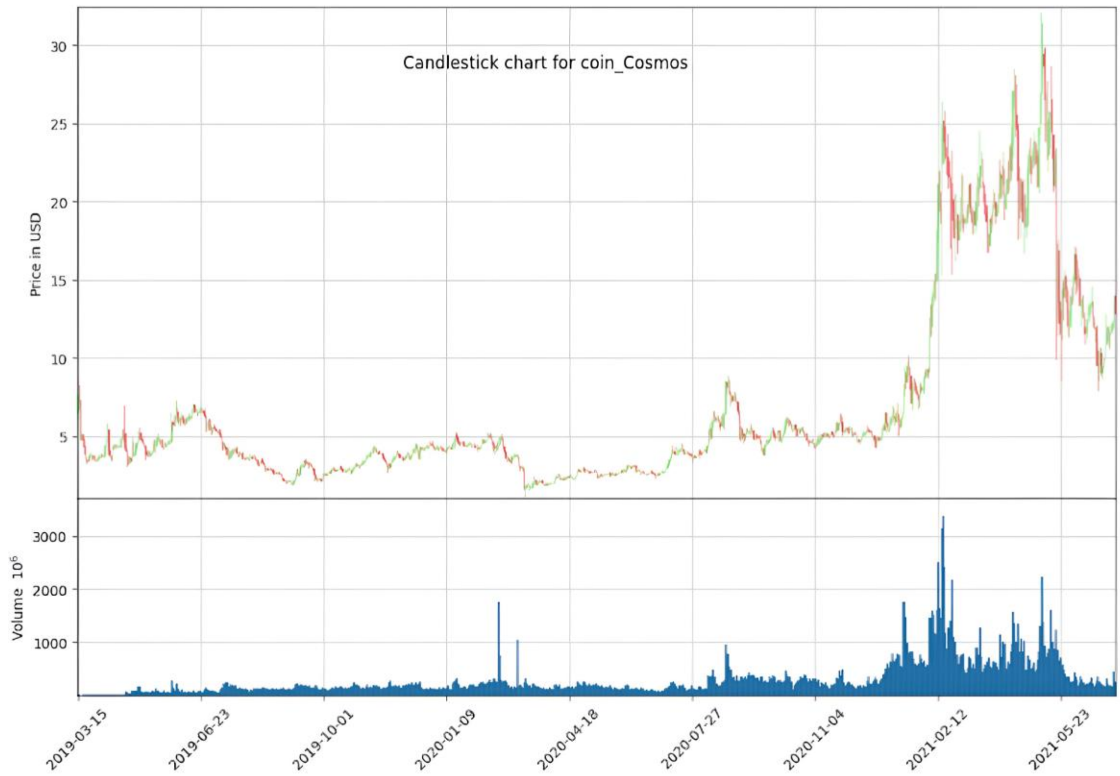

Figure 3. Cosmos Cryptocurrency Volatility and Trading Patterns.

3.3. Ethereum

The candlestick chart (Figure 4) presents the historical price and trading volume movements for the cryptocurrency Ethereum (ETH) for the period from May 2017 to June 2021. The chart consists of two main sections:

Price Dynamics (Top Section):

Ethereum went through several major bull cycles, starting with its initial surge in late 2017 and early 2018, when the price reached approximately USD 1,400. This was followed by a prolonged correction throughout 2018 and 2019, with the price stabilising below USD 300. After a relatively calm period, a significant upward trend began in mid-2020, which accelerated sharply in early 2021. ETH reached an all-time high of around USD 4,300 in May 2021, after which it underwent a sharp correction, returning to the $2,000–USD 2,500 range.

These pronounced fluctuations reflect both the general market cycles in the crypto sector and the increasing use of Ethereum as a foundational layer for DeFi and smart contracts.

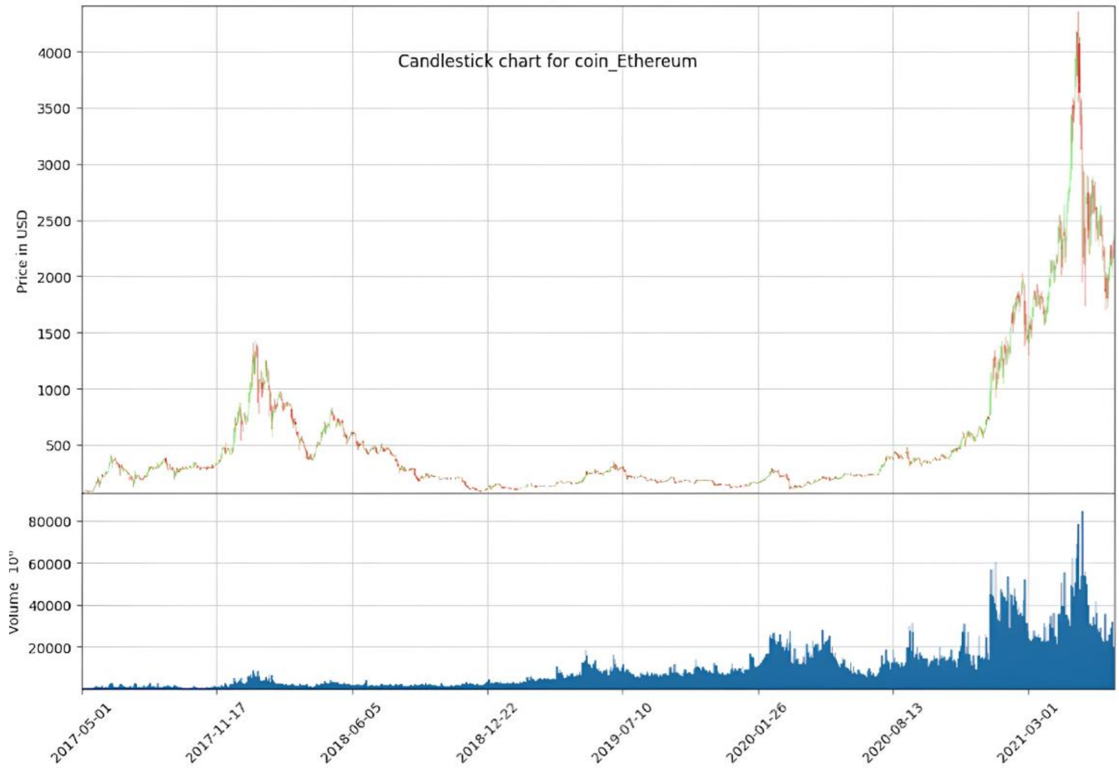

Figure 4. Ethereum Cryptocurrency Volatility and Trading Patterns.

Trading Volume (Bottom Section):

Trading volume also demonstrated a cyclical increase. In the early trading periods (2017–2019), the volume was relatively low and stable. From mid-2020 onward, a noticeable and sustained increase in volume was observed, peaking in the first and second quarters of 2021, parallel to the dramatic price increase. This surge coincides with Ethereum's growing utility in decentralised applications, increasing institutional participation, and speculative activity from retail investors. The volume peaks closely track the major price movements, highlighting the strong connection between liquidity and market sentiment.

Ethereum's price and volume behaviour clearly show that it is a highly volatile, trend-sensitive asset [20], which makes it suitable for advanced machine learning models such as:

- Transformer-based time series predictors
- Volatility clustering models (e.g., GARCH variants)
- Attention-enhanced LSTM networks

The clear relationship between volume and price dynamics supports multivariate strategies with volume as a predictor, which improves forecast accuracy.

From an investment strategy perspective, Ethereum's historical trajectory underscores the importance of predicting macro-trends and integrating dynamic liquidity analysis into algorithmic trading and portfolio management in decentralised markets.

4. Results

4.1. Bitcoin

The predictive effectiveness for Bitcoin's market capitalisation was evaluated using Linear Regression, Lasso Regression, and Decision Tree Regression (Table 1).

Table 1. Model Performance Comparison for Predicting Bitcoin's Market Capitalisation.

Model	Accuracy (R^2)	MSR (MAE)	RSME	R^2 Score

 doi:10.20944/preprints202604.1107.v1

Linear Regression	0.99952	2,866,891,169.60	2,866,891,170.50 2	0.99952
Lasso Regression	0.99940	3,138,647,243.02	3,138,647,244.12 1	0.99940
Decision Tree	0.99870	4,546,167,840.66	4,546,167,841.56 4	0.99870

Linear Regression

The Linear Regression model achieved the highest accuracy with an **R value of 0.99952**. This indicates that the model captures almost all the variability in the market capitalisation data. The **Mean Absolute Error (MAE) was approximately 2.87 billion**, the lowest among the models tested, which reflects high precision and consistency in its predictions.

Lasso Regression

Lasso Regression also produced strong results, with an **R^2 value of 0.99940** and an **MAE of about 3.14 billion**. While its performance was slightly lower than that of standard linear regression, the difference is negligible. This suggests that the regularisation in the Lasso model has a minimal effect and does not substantially compromise accuracy on this dataset.

Decision Tree Regression

The Decision Tree model achieved a lower **R^2 score of 0.99870**, indicating strong explanatory power, although it does not reach the level of the linear models. More notably, the **MAE was approximately 4.55 billion**, significantly higher than that of the linear models. This shows that while the Decision Tree captures the general pattern, it is more prone to larger individual errors in its predictions.

Comparative Analysis and Conclusions

Both Linear and Lasso Regression outperformed the Decision Tree model in both accuracy and predictive error. The results indicate that Bitcoin's market capitalisation is best modelled by a linear relationship between the variables. The small difference between the two linear models suggests that Lasso regularisation does not provide a significant advantage. The significantly higher MAE value for the Decision Tree, despite a high R2, suggests greater variability and lower reliability for precise forecasts in this context.

4.2. Cosmos

Linear Regression Results

The linear model (Table 2) achieved a coefficient of determination of **R^2=0.98631**, indicating that over 98% of the variation in Cosmos's market capitalisation can be explained by linear relationships with the input variables. The model's **Mean Absolute Error (MAE) was approximately 106 million USD**, which suggests a moderate level of predictive accuracy.

Table 2. Model Performance Comparison for Predicting Cosmos's Market Capitalisation.

Model	Accuracy (R^2)	MSR (MAE)	RSME	R^2 Score
Linear Regression	0.98631	106,217,303.22	106,217,312.126	0.98631
Lasso Regression	0.98637	106,097,408.62	106,097,421.625	0.98637
Decision Tree	0.99585	54,207,043.91	54,207,049.131	0.99585

Comparative Analysis of Regression Models for Forecasting Cosmos (ATOM) Market Capitalisation.

Lasso Regression Results

The Lasso regression model, which applies L1 regularisation to constrain coefficients and reduce potential overfitting, showed almost identical results to standard linear regression. The achieved values were **R^2=0.98637 and an MAE of ≈ 106 million**, which suggests that in this case, regularisation neither significantly improved nor worsened the quality of the forecast.

Decision Tree Regression Results

The Decision Tree Regression model demonstrated the highest predictive effectiveness among the three models tested. With a reported value of **R2=0.99585 and an MAE of ≈ 54 million USD**, the model showed both higher accuracy and a significantly lower forecast error compared to the linear approaches.

Summary and Conclusions

Although all three models showed a high degree of predictability, the results highlight the superiority of Decision Tree Regression in modelling Cosmos's market capitalisation. The lower error and higher R^2 value indicate that the relationships between the predictors and the target variable contain significant nonlinear dependencies that the linear models fail to capture effectively.

In practice, this means that for Cosmos, regression models capable of capturing nonlinear structures - such as decision trees - offer higher reliability in forecasting. Linear and Lasso regression remain stable and easy-to-interpret solutions, but when more complex patterns exist, as is the case with Cosmos, a tree-based model is the preferred approach.

4.3. Ethereum

Linear Regression Results

The Linear Regression model (Table 3) achieved a coefficient of determination of **R^2=0.99719** and a **Mean Absolute Error (MAE) of ≈ 2.53 billion USD**. These values reflect high explanatory power and a relatively low average forecasting error, making the model suitable for tasks with a linear data structure.

Table 3. Model Performance Comparison for Predicting Ethereum's Market Capitalisation.

Model	Accuracy (R^2)	MSR (MAE)	RMSE	R^2 Score
Linear Regression	0.99719	2,533,139,425.19	2,533,139,505.11	0.99719
Lasso Regression	0.99714	2,544,171,355.85	2,544,171,395.62	0.99714
Decision Tree	0.99709	2,191,188,515.03	2,191,188,545.43	0.99709

Comparative Analysis of Regression Models for Forecasting Ethereum (ETH) Market Capitalisation.

Lasso Regression Results

Lasso Regression, which applies L1 regularisation to limit model complexity, showed almost identical results to standard linear regression. The recorded values were **R^2=0.99714 and an MAE of ≈ 2.54 billion USD**, suggesting that the impact of regularisation in this context is negligible.

Decision Tree Regression Results

The Decision Tree model demonstrated the lowest mean absolute error of approximately **2.19 billion USD**, although its **R^2 value (0.99709)** was slightly lower than that of the linear models. This suggests that the model captures certain nonlinear dependencies, which lead to improved accuracy in individual forecasts.

Comparative Analysis and Conclusions

All three models demonstrate a high degree of predictability, with **R^2 values above 0.997** and **MAE falling within a narrow range between 2.19 and 2.54 billion USD**. The differences between

 doi:10.20944/preprints202604.1107.v1

the models are minimal: Linear and Lasso regression show almost identical accuracy, while the Decision Tree regression achieves a slightly lower error despite having less precise explanatory power.

These results indicate that Ethereum's market capitalisation can be effectively modelled using both linear and nonlinear approaches. The choice of model should be based primarily on the need for interpretability, a preference for regularisation, or sensitivity to local data dependencies, rather than on significant differences in forecasting accuracy. The latter case is the model performance presented in [12] or in the one of [13] where the lattice rule depends on the choice of the generator vectors.

Key Findings Across Different Cryptocurrencies

The empirical results show that the effectiveness of the models is not uniform for all cryptocurrencies, but several general patterns have emerged. For most assets—including leading coins like Bitcoin and Ethereum—both Linear Regression and Lasso Regression achieve very high coefficients of determination (R^2), typically above 0.99, accompanied by low Mean Absolute Error (MAE) values. This indicates that for many of the most well-known and liquid cryptocurrencies, market capitalisation is primarily determined by features that have a strong linear relationship with the target variable. In these cases, the predictions of linear models are stable and highly accurate.

At the same time, the study found an exception where nonlinear models, and specifically the Decision Tree, yielded better results. For a particular currency like Cosmos, the Decision Tree achieved higher R^2 values and lower MAE compared to the linear alternatives. The finding suggests that the underlying data for these assets contain meaningful non-linearities, which are better captured by tree-based models. Such patterns may reflect unique structural characteristics, market behaviours, or exogenous events that affect specific cryptocurrencies by G. Vasileva [11].

5. Implications for Practice and Future Research

The results of this study have clear practical implications for the use of machine learning in cryptocurrency price forecasting, as well as for the implementation of AI in the service sector.

First, the analysis of cryptocurrency price forecasting shows that there is no universal model that is optimal for all currencies; rather, the applicability of a given model depends heavily on the specific asset. Practitioners are therefore advised to conduct preliminary exploratory data analysis and comparative model testing, tailored to the specific characteristics of each cryptocurrency, before selecting a particular forecasting approach. The strong performance of linear models for most currencies indicates that simple and easily interpretable regression techniques can deliver top-tier results in many practical contexts - especially when transparency and analysis are key [19,21]. At the same time, the superior performance of Decision Tree Regression for certain currencies underscores the value of incorporating nonlinear models into the selection process, particularly for assets with unstable or fluctuating market behaviour.

Second, the results from implementing AI in customer service reveal important points for its practical introduction. Despite the high overall success rate of "Naomi" in the chat, the significant number of unresolved cases, due to both the client and AI limitations, emphasises that it is not a universal solution. Practitioners must invest in mechanisms to identify these "edge cases" and ensure seamless workflows for employees to handle complex or non-standard problems.

Ultimately, this research lays the groundwork for future studies to explore more advanced machine learning techniques and ensemble models, such as Random Forests, Gradient Boosting, or neural networks, which could further enhance forecasting accuracy by capturing more complex dependencies in cryptocurrency data. Regarding the implementation of AI in services, future research could focus on developing dynamic models for AI adaptation to changing conditions, real-time automatic bug detection and diagnostics, as well as optimising workflows with the help of AI for maximum efficiency and customer satisfaction.

6. Conclusion

One thing is very clear: artificial intelligence is going to change our world forever. And the change is likely to be more profound than most people realise today [8] such as B. Marr and M. Ward. AI is rapidly making inroads in business. Its swift adoption means that questions on both the opportunities and risks are at the forefront of most discussions… As we've shown, AI gives people powerful tools to do more, in essence, to perform with superhuman capability [5], by T. H. Davenport and J. Kirby.

AI is considered to have the potential to instigate a fourth industrial revolution and is dramatically changing people's patterns of interaction and economic activities see Lu C. H. in [7]. This study extensively examines the application of machine learning in two distinct but rapidly developing areas: cryptocurrency market capitalisation forecasting and the implementation of artificial intelligence in customer service. Through a systematic analysis of quantitative and qualitative data, key insights were derived regarding the effectiveness, limitations, and strategic implications of AI systems.

In the context of AI in the service sector, the key to reliable and effective implementation lies in the careful evaluation of AI functionality for each channel, the prioritisation of stability and reliability, and continuous optimisation based on empirical data and an understanding of the AI's limitations. These insights provide a solid foundation for practitioners and researchers who wish to apply machine learning in the rapidly evolving field of crypto-analysis and service automation.

The findings of this thesis have clear implications for the application of machine learning to cryptocurrency price prediction. First, they demonstrate that no single model is optimal for all currencies; rather, model suitability is highly asset-dependent. Practitioners are therefore encouraged to conduct exploratory data analysis and model benchmarking tailored to the characteristics of each cryptocurrency before settling on a predictive approach [16].

Both linear and nonlinear regression models can be extremely effective tools for forecasting cryptocurrency market capitalisation, but their success depends on the specific characteristics of each digital asset. While linear models are sufficient for most leading currencies, Decision Tree Regression is indispensable for assets with more complex and nonlinear market dynamics.

The potential for AI to transform business is unprecedented, yet an urgent and growing challenge lies ahead. Companies are at a crossroads in its implementation. While businesses implement systems - from machine learning to deep learning - some will continue to see modest productivity improvements in the short term, but these results will eventually wane. Other companies, however, will achieve revolutionary performance improvements by developing innovative, game-changing approaches.

Acknowledgments: The presentation and dissemination of these research results are supported by National Science Fund Project КП-06-Н85-7/05.12.2024 "Significance and Potential Risks of the Fast Integration of Artificial Intelligence (AI) Technologies into the Economy and Financial Sector".

Appendix A

```
# -*- coding: utf-8 -*-
"""Cryptocurrency Lasso Regression, DECISION TREE, RANDOM FOREST
Automatically generated by Colab.
Original file is located at
    https://colab.research.google.com/drive/1Jr_PE9K5W2LKPZBGflRmw7V9zAmUuM4L
"""
!pip install -q kaggle
!pip install -q mplfinance
!pip install -q prophet
!pip install prophet
```

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 April 2026 doi:10.20944/preprints202604.1107.v1


```
import pandas as pd
import zipfile
import pandas as pd
from google.colab import drive
drive.mount('/content/drive')
# files.upload()
! mkdir -p ~/.config/kaggle
!cp /content/drive/MyDrive/kaggle/kaggle.json ~/.config/kaggle/kaggle.json
!chmod 600 ~/.config/kaggle/kaggle.json
!ls ~/.config/kaggle
from kaggle.api.kaggle_api_extended import KaggleApi
api = KaggleApi()
api.authenticate()
api.model_list_cli()
import kaggle
import os
from datetime import datetime
# Get the current date in the desired format (YYYYMMDD)
current_date = datetime.now().strftime('%Y%m%d')
# Define the data folder path
data_folder = f'/content/drive/MyDrive/{current_date}/data_forecasting'
# Create the directory if it doesn't exist
os.makedirs(data_folder, exist_ok=True)
# Verify the folder creation
print(f'Data folder created at: {data_folder}')
# 1. Install the Kaggle CLI (if you haven't already):
# pip install kaggle
# 2. Make sure you have your Kaggle API token file placed at ~/.kaggle/kaggle.json
# (See https://github.com/Kaggle/kaggle-api#api-credentials)
from kaggle.api.kaggle_api_extended import KaggleApi
import os
import zipfile
# --- Configuration ---
DATASET = "sudalairajkumar/cryptocurrencypricehistory"
DATA_FOLDER = "data/crypto_price_history"
UNZIP = True
# 3. Initialize API
api = KaggleApi()
api.authenticate()
# 4. Ensure target folder exists
os.makedirs(DATA_FOLDER, exist_ok=True)
# 5. Download the dataset
api.dataset_download_files(
dataset=DATASET,
path=DATA_FOLDER,
unzip=False # we'll unzip manually below
)
zip_path = os.path.join(DATA_FOLDER, DATASET.split("/")[-1] + ".zip")
# 6. (Optional) Unzip the archive into the same folder
if UNZIP and os.path.exists(zip_path):
with zipfile.ZipFile(zip_path, 'r') as z:
```

doi:10.20944/preprints202604.1107.v1


```
z.extractall(DATA_FOLDER)
# Remove the zip file if you don't need it
os.remove(zip_path)
print(f"Dataset downloaded to '{DATA_FOLDER}'")
input_dir_path = "./data/crypto_price_history"
INPUT_FILE = "coin_Bitcoin.csv"
bitcoin=pd.read_csv(f"{input_dir_path}/{INPUT_FILE}")
"""# DATA PREPROCESSING:"""
pd.set_option('display.max_columns', None)
bitcoin.head()
bitcoin.shape
bitcoin.Name.unique()
bitcoin.Symbol.unique()
bitcoin.drop(["Name"],axis=1, inplace=True)
bitcoin.drop(["SNo"],axis=1, inplace=True)
bitcoin.drop(["Symbol"],axis=1, inplace=True)
bitcoin.columns
bitcoin.info()
bitcoin.describe()
import datetime as dt
bitcoin["Date"]=pd.to_datetime(bitcoin["Date"])
bitcoin['Date_year'] = bitcoin["Date"].dt.year
bitcoin['Date_month'] = bitcoin["Date"].dt.month
bitcoin['Date_day'] = bitcoin["Date"].dt.day
bitcoin['Date_hour'] = bitcoin["Date"].dt.hour
bitcoin['Date_minute'] = bitcoin["Date"].dt.minute
bitcoin['Date_seconde'] = bitcoin["Date"].dt.second
bitcoin.drop(["Date"], axis=1, inplace=True)
bitcoin.head()
X=bitcoin.drop(["Marketcap"], axis=1)
Y=bitcoin["Marketcap"]
X.head()
"""# Modeling"""
from sklearn.linear_model import LinearRegression
LR=LinearRegression()
from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest=train_test_split(X,Y,test_size=0.2, random_state=42)
LR.fit(xtrain,ytrain)
ypred=LR.predict(xtest)
print(pd.DataFrame({'real_value':ytest,'pred_value':ypred}))
from sklearn.metrics import mean_absolute_error, r2_score
print("Linear Regression MAE:", mean_absolute_error(ytest, ypred))
print("Linear Regression RMSE:", root_square_mean_error(ytest, ypred))
print("Linear Regression R^2:", r2_score(ytest, ypred))
"""## Lasso Regression"""
from sklearn.linear_model import Lasso
Ls=Lasso(random_state=42)
Ls.fit(xtrain,ytrain)
ypred=Ls.predict(xtest)
print("Lasso Regression MAE:", mean_absolute_error(ytest, ypred))
print("Lasso Regression RMSE:", root_square_mean_error(ytest, ypred))
```

 doi:10.20944/preprints202604.1107.v1


```
print("Lasso Regression R^2:", r2_score(ytest, ypred))
"""## DECISION TREE Regressor"""
from sklearn.tree import DecisionTreeRegressor
mytre=DecisionTreeRegressor(max_depth=1, random_state=42)
mytre.fit(xtrain,ytrain)
ypredd=mytre.predict(xtest)
print("Decision Tree (depth=1) MAE:", mean_absolute_error(ytest, ypredd))
print("Decision Tree (depth=1) RMSE:", root_mean_square_error(ytest, ypredd))
print("Decision Tree (depth=1) R^2:", r2_score(ytest, ypredd))
mytre=DecisionTreeRegressor(max_depth=5, random_state=42)
mytre.fit(xtrain,ytrain)
ypredd=mytre.predict(xtest)
print("Decision Tree (depth=5) MAE:", mean_absolute_error(ytest, ypredd))
print("Decision Tree (depth=5) RMSE:", root_mean_square_error(ytest, ypredd))
print("Decision Tree (depth=5) R^2:", r2_score(ytest, ypredd))
"""## RANDOM FOREST Regressor"""
from sklearn.ensemble import RandomForestRegressor
# A solid baseline RF configuration; feel free to tune n_estimators/max_depth later
rf = RandomForestRegressor(
n_estimators=300,
max_depth=None, # let trees grow; can cap for speed/generalization
min_samples_split=2,
min_samples_leaf=1,
n_jobs=-1,
random_state=42
)
rf.fit(xtrain, ytrain)
y_rf = rf.predict(xtest)
print("Random Forest MAE:", mean_absolute_error(ytest, y_rf))
print("Random Forest RMSE:", root_mean_square_error(ytest, y_rf))
print("Random Forest R^2:", r2_score(ytest, y_rf))
# Optional: feature importances to see what the RF paid attention to
feat_imp = pd.Series(rf.feature_importances_, index=X.columns).sort_values(ascending=False)
print("Top 10 Feature Importances:")
print(feat_imp.head(10))
```

References

1. Accenture, "Artificial Intelligence Summary Index," 2023. [Online]. Available: https://www.accenture.com/mu-en/insights/artificial-intelligence-summary-index.
2. Agrawal A., Prediction Machines: The Simple Economics of Artificial Intelligence. Harvard Business Review Press, 2018.
3. Bain & Company, "Five Functions Where AI Is Already Delivering Tech Report," 2024. [Online]. Available: https://www.bain.com/insights/five-functions-where-ai-is-already-delivering-tech-report-2024/.
4. Benhamou S., "Artificial Intelligence and the Future of Work," *Revue d'économie industrielle*, vol. 169, 2020.
5. Davenport T. H., Kirby J., Human + Machine: Reimagining Work in the Age of AI. Harvard Business Review Press, 2018.
6. Dirican C., "The Impacts of Robotics, Artificial Intelligence on Business and Economics," *Procedia - Social and Behavioral Sciences*, vol. 195, pp. 564–573, 2015. https://doi.org/10.1016/j.sbspro.2015.06.134.
7. Lu C.-H., "The impact of artificial intelligence on economic growth and welfare," *Journal of Macroeconomics*, vol. 69, p. 103342, Sept. 2021. https://doi.org/10.1016/j.jmacro.2021.103342.

8. Marr B., Ward M., Artificial Intelligence in Practice: How 50 Successful Companies Used AI and Machine Learning to Solve Problems. Wiley, 2018.
9. Naskinova I., Kolev M., Lazarova M., "Forecasting Strategies in Retail: Utilizing Advanced Machine Learning Methods while Safeguarding Privacy," *Journal of Physics: Conference Series*, vol. 2910, no. 1, p. 012008, 2024. DOI: 10.1088/1742-6596/2910/1/012008.
10. Rajkumar S. S., "Cryptocurrency Historical Price Dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory.
11. Vasileva G., Predicting Cryptocurrency Prices Using Machine Learning: Data Analysis and Modelling. Varna Free University, Varna, 2025.
12. Nikolay N., Lomev B., Netova D., Spasov K., Markov Regime-Switching Modeling of Passenger Flow in a Suburban Area Based on Anonymized Telecommunications Data, vol.1492 LNNSPages: 423 – 439, DOI: 10.1007/978-3-031-96775-7_36.
13. Todorov V., Dimov I., DImitrov Y., Efficient quasi-Monte Carlo methods for multiple integrals in option pricing, AIP Conference ProceedingsConference Paper, 2018, DOI: 10.1063/1.5064950.
14. Moser S., Brauneis. A., Intraday price forecasts using candlestick patterns in cryptocurrency markets, 1. April, 2026, International Review of Economics & Finance, https://doi.org/10.1016/j.iref.2026.105158
15. Yotov K., Hadzhikoleva S., Hadzhikolev E., A Study on Global Cyclicality of the S&P 500 Index Using a Hybrid Model of Recurrent Neural Networks and Fourier Transformations, 24th International Symposium INFOTEH-JAHORINA, INFOTEH 2025 Proceedings, 2025, DOI: 10.1109/INFOTEH64129.2025.10959219 .
16. Bogdanova B., Stancheva-Todorova E., ML-based predictive modelling of stock market returns, AIP Conference Proceedings, 2021, DOI: 10.1063/5.0042805.
17. Ivanov I., Kabaivanov S., Bogdanova, B., Stock market recovery from the 2008 financial crisis: The differences across Europe, Research in International Business and Finance, 2016, 37, pp. 360–374, DOI: 10.1016/j.ribaf.2016.01.006
18. Stancheva-Todorova E., Bogdanova B., Enhancing investors' decision-making-An interdisciplinary AI-based case study for accounting students, AIP Conference Proceedings, 2021, 2333, 150007, DOI: 10.1063/5.0041612.
19. Atanasi Tashev, Yordan Stoyanov; Analysis of the CNG impact on some of the combustion parameters of dual-fuel compression ignition engines. AIP Conf. Proc. 24 April 2024; 3078 (1): 050008. https://doi.org/10.1063/5.0208335
20. Kovacheva, M. V., Dynamic pricing in data-based marketing. AIP Conf. Proc. 24 April 2024; 3078 (1): 070006. https://doi.org/10.1063/5.0208310
21. Delova I., Borisov T., Mirchev Y., Raychev R., Numerical Modeling and Analysis of Fatigue Failure in 42CrMo4 Steel Pivot Bolts at Different Heat Treatments. Eng. Proc. 2025, 100, 52. https://doi.org/10.3390/engproc2025100052.