Exploring Artificial Intelligence for Learning Enhancement with Predictive and Explainable Modelling

Naruemon Thepnuan[1,*]

[1] Educational Technology and Communications Division, Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Thailand

ABSTRACT

This study explores the application of Artificial Intelligence (AI) in learning through the use of a Linear Regression model to analyse and predict relationships among financial dataset variables. The dataset consists of several numerical features, including open, high, low, close, and trading volumes for XRP and USDT. Descriptive statistics, correlation analysis, and data distribution visualizations were conducted to provide an understanding of the dataset before implementing the predictive model. The correlation results showed very strong linear relationships among the price variables ($r \approx 0.99$) and a strong positive correlation between trading volumes ($r = 0.82$), indicating highly synchronized market movements. The AI model achieved exceptional predictive performance with an R^2 value of 0.993, a Mean Absolute Error (MAE) of 0.0099, and a Root Mean Squared Error (RMSE) of 0.0286, demonstrating that even a simple algorithm can accurately capture complex numerical patterns. The feature importance analysis revealed that the high and low variables were the most influential predictors of the closing price, providing a clear example of model interpretability and eXplainable AI (XAI). From an educational perspective, this research illustrates how AI can serve as a practical and interactive learning tool that helps students understand core data science concepts, including data preprocessing, correlation, model evaluation, and interpretability. The study concludes that integrating AI into learning environments enhances students' analytical thinking, promotes data literacy, and encourages a deeper understanding of how AI systems learn and make predictions.

Keywords Artificial Intelligence, Machine Learning, Predictive Modelling, XAI, AI in Education

Introduction

The rapid advancement of AI has transformed various sectors, including education, finance, and data analytics. In recent years, the integration of AI technologies into the learning process has created new opportunities for developing students' analytical, computational, and critical thinking skills [1]. AI in education, commonly referred to as AI in Learning, emphasizes not only the automation of instructional tasks but also the enhancement of human learning through interactive, data-driven environments [2]. By simulating real-world data problems and providing tools for predictive modelling, AI enables learners to understand abstract concepts such as correlation, regression, and model evaluation more concretely [3].

A growing number of studies have demonstrated the potential of AI to support adaptive learning systems, intelligent tutoring, and data-driven decision-making in education [4][5]. Recent developments in Explainable Artificial Intelligence (XAI) have further strengthened the role of AI in education by ensuring that models are transparent, interpretable, and suitable for instructional purposes [6]. For instance, regression and classification models have been successfully used

Submitted 8 January 2026
Accepted 12 February 2026
Published 1 March 2026

*Corresponding author
Naruemon Thepnuan,
naruemon_t@rmutt.ac.th

Additional Information and Declarations can be found on page 32

DOI: 10.63913/ail.v2i1.12

How to cite this article: N.Thepnuan, "Exploring Artificial Intelligence for Learning Enhancement with Predictive and Explainable Modelling," *Artif. Intell. Learn.*, vol. 2, no. 1, pp. 17-34, 2026.

to teach students about predictive analytics and statistical learning, providing practical exposure to how algorithms identify and interpret relationships between variables [7]. These implementations not only foster engagement but also encourage students to think critically about the ethical and technical implications of AI systems [8].

Despite the growing body of research on AI in education, there remains a research gap in the use of real-world datasets to teach AI-based analytical methods within authentic learning contexts. Many existing educational AI frameworks focus on predesigned datasets or theoretical simulations, which often lack complexity and fail to reflect the variability found in actual data [9]. Consequently, students may gain a conceptual understanding of machine learning algorithms but lack the experience needed to manage issues such as multicollinearity, skewed distributions, and feature interpretability. This gap highlights the need for studies that demonstrate how AI can be used as a learning tool to bridge theoretical knowledge with practical data analysis.

The present study aims to address this gap by applying a Linear Regression model to a real financial dataset as a pedagogical example of AI-based predictive learning. The research combines statistical analysis with AI modelling to illustrate how simple algorithms can learn, predict, and explain relationships within complex data structures. This approach allows learners to engage directly with core data science concepts, such as correlation, feature importance, and model evaluation, while also developing an understanding of explainable AI principles. The findings of this study contribute to both the technical domain of AI modelling and the pedagogical domain of AI education, providing insights into how AI can enhance analytical learning through practical and interpretable applications [10].

Literature Review

Artificial Intelligence in Learning and Education

AI has become a transformative force in modern education, reshaping how knowledge is delivered, assessed, and personalized to meet individual learners' needs. AI-powered educational systems such as adaptive learning platforms, intelligent tutoring systems, and learning analytics tools are increasingly being used to enhance engagement and improve learning outcomes [11]. Recent studies emphasize that AI in learning environments extends beyond automation and content delivery, playing an essential role in cultivating critical thinking, problem-solving, and analytical reasoning skills among students [12].

The concept of AI in Learning promotes the use of AI as a partner in human learning, enabling students to interact with data, algorithms, and computational models directly [13]. Through this approach, learners gain practical experience in understanding complex relationships in data, including correlation, causation, and prediction, which are essential in data-driven education. The integration of AI into the classroom supports experiential learning, where students learn by doing rather than by passively receiving information [14]. However, existing literature shows that while AI is often applied in teaching automation and adaptive assessment, its role as a hands-on learning tool for developing analytical and computational thinking remains underexplored [15].

Machine Learning and Predictive Modelling in Education

Machine Learning (ML) represents a critical subfield of AI that focuses on enabling computers to learn patterns and make predictions based on data [16]. Within educational contexts, ML models have been widely used to predict student performance, identify learning difficulties, and personalize instructional strategies [17]. Predictive modelling techniques, such as regression analysis, decision trees, and neural networks, allow researchers and educators to identify significant factors that influence learning outcomes and to optimize educational interventions [18].

Despite its success in educational analytics, the use of machine learning as an instructional medium for teaching AI concepts remains relatively limited. Many educational programs emphasize theoretical discussions of algorithms but provide little opportunity for students to engage directly with model training and evaluation using authentic datasets [19]. Hands-on applications of predictive modelling can help bridge this gap by enabling students to visualize how algorithms learn patterns, adjust parameters, and minimize errors during the learning process [20].

By incorporating predictive modelling into AI education, learners can gain deeper insights into essential concepts such as overfitting, bias, variance, and model generalization. This approach not only improves their technical proficiency but also enhances their understanding of how AI can be applied responsibly in real-world problem-solving.

Explainable Artificial Intelligence (XAI) and Model Interpretability

As AI systems become increasingly complex, ensuring their transparency and interpretability has become a major area of focus. XAI aims to make AI models more understandable by revealing how they make predictions or classifications [21]. In the context of education, XAI plays a vital role by helping learners comprehend not only the outcomes of AI models but also the reasoning behind those outcomes [22].

Linear models such as regression are particularly valuable for instructional purposes because they provide interpretable coefficients that directly indicate the influence of each variable on the output [23]. Research shows that using explainable models helps build student trust and engagement by allowing them to see how AI systems reach decisions [24]. Furthermore, exposure to XAI concepts promotes ethical awareness, as students learn to question and analyse the fairness, accountability, and transparency of AI applications [25].

In this study, the focus on feature importance and coefficient interpretation demonstrates how AI models can be used not only for predictive accuracy but also for understanding the underlying data relationships. This educational approach encourages students to view AI as a transparent and explainable system, aligning with modern expectations of responsible AI development and use.

Research Methodology

This study employed a quantitative experimental approach to explore how AI can be applied in learning environments through predictive modelling. The objective was to assess the predictive capability of a Linear Regression model on real-world data and to demonstrate the educational potential of AI as an interactive learning tool. The research process followed a structured pipeline consisting of data collection, preprocessing, exploratory analysis, model development, evaluation, and interpretation. Each stage was carefully designed to ensure both analytical accuracy and pedagogical relevance, enabling learners to replicate and understand the complete AI workflow. The overall methodological structure and research steps are illustrated in figure 1, which outlines the sequential flow of activities conducted throughout this study.

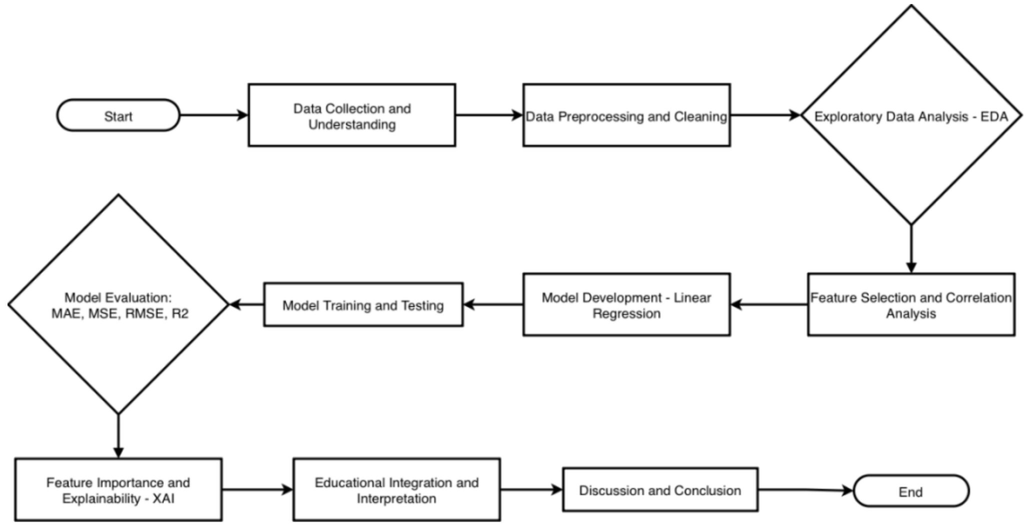

Figure 1 Research Step

Dataset Description

The dataset used in this research consists of time-series financial data representing transactional and market behaviour across a continuous period. It contains seven numerical attributes, namely open, high, low, close, Volume XRP, Volume USDT, and unix (timestamp). Each record captures a snapshot of market conditions at a given time, providing a structured temporal view of price fluctuations and trading activity. The variables open, high, low, and close represent key indicators of asset price dynamics, while Volume XRP and Volume USDT denote the trading intensity and liquidity of each digital asset. The Unix variable serves as a temporal reference that supports time-based analysis and chronological modelling. Before analysis, the dataset was verified for consistency, completeness, and chronological accuracy to ensure the reliability of the subsequent modelling process.

These variables were selected due to their strong inter-variable relationships and statistical significance in financial prediction contexts. The price-related variables exhibit naturally high correlation, which makes them ideal for testing the performance and interpretability of regression-based models. Meanwhile, trading volume variables provide an additional dimension that reflects market participation and volatility. From a pedagogical perspective, this dataset provides an authentic and complex learning resource for students studying AI and data analytics. It exposes learners to real-world data challenges such as multicollinearity, skewed distributions, and temporal dependencies, all of which are essential concepts in machine learning. By analysing this dataset, students not only develop technical competence in regression and data preprocessing but also gain a deeper understanding of how AI systems extract patterns, manage correlated features, and make informed predictions from real data environments.

Data Preprocessing

Data preprocessing was carried out to ensure the accuracy, consistency, and integrity of the dataset before model training. Since real-world data often contains irregularities such as missing entries, inconsistencies, or non-numeric information, careful preprocessing was essential to prepare the dataset for analysis. The process began with data cleaning, which involved removing duplicate records and handling missing values through interpolation to maintain continuity in the time series. The type conversion step was then applied to ensure that all variables used in the model were numeric; non-predictive columns such as timestamps were excluded from the regression inputs but retained for chronological context. These steps helped to minimize potential data noise and enhance the precision of subsequent analytical procedures.

To further refine the dataset, feature scaling was implemented by normalizing the numerical attributes. This transformation ensured that all features contributed proportionally to the regression model and prevented numerical dominance by variables with larger magnitudes. Additionally, outlier management was performed to identify and assess extreme values in trading volumes and price fluctuations. Instead of removing these outliers entirely, they were retained after verification to preserve the natural variability of market behaviour and maintain dataset authenticity. This comprehensive preprocessing procedure ensured that the dataset was statistically robust and ready for model development. Moreover, it serves as an important educational demonstration of how proper data preparation enhances model performance and reproducibility within AI learning environments.

Exploratory Data Analysis (EDA)

EDA was performed to understand the underlying statistical structure, variability, and relationships within the dataset. This stage served as a diagnostic process to uncover trends, detect anomalies, and assess the suitability of variables for predictive modelling. Descriptive statistics were computed to summarize measures of central tendency, including mean and median, as well as measures of dispersion such as standard deviation, range, and variance. These statistical summaries provided insight into how the dataset was distributed and helped identify skewness or asymmetry in price and volume data. Furthermore, frequency distributions were examined to reveal how often specific price ranges or trading volumes occurred, providing a preliminary view of market volatility and stability.

To complement numerical summaries, several visualization techniques were applied to better interpret data relationships. Histograms were used to depict the distribution of each variable, revealing that the price-related attributes (open, high, low, and close) displayed a slightly right-skewed distribution, indicating that most price values clustered toward the lower range. Correlation heatmaps were generated to illustrate the strength and direction of linear relationships between features. The analysis revealed strong positive correlations among the price-related variables (correlation coefficient ≈ 0.99), suggesting synchronized price movements, and a high correlation between Volume XRP and Volume USDT ($r = 0.82$), reflecting interconnected trading behaviour. These findings confirmed the dataset's consistency and relevance for regression-based modelling. Overall, the EDA stage was essential not only for guiding feature selection and model design but also for helping learners visually and statistically comprehend how data patterns influence AI model performance.

Model Development

The predictive model in this study was developed using the Linear Regression algorithm, a fundamental and widely used method in machine learning that aims to model the linear relationship between a dependent variable and several independent variables. This algorithm was applied to predict the closing price of the asset using the remaining numerical features, namely open, high, low, Volume XRP, and Volume USDT. The implementation was carried out using the Scikit-learn library in Python, which provides robust and efficient tools for model construction, training, and validation. To ensure generalization and avoid overfitting, the dataset was divided into training and testing subsets, allowing the model to learn patterns from the training data and evaluate its predictive accuracy on unseen data. This stage marked the transition from data exploration to computational learning, where the algorithm began identifying statistical dependencies between the variables.

The mathematical representation of the Linear Regression model can be expressed as:

$$\hat{y} = \beta_0 + \sum_{i=1}^{n} \beta_i \, x_i + \epsilon \tag{1}$$

In this equation, $\hat{y}$ denotes the predicted value of the dependent variable, β_0 represents the intercept term, $\beta_i$corresponds to the coefficient assigned to each independent variable x_i, and ϵ symbolizes the error term or residual component that accounts for unexplained variations in the data. The parameters of the model are estimated using the Ordinary Least Squares (OLS) method, which seeks to minimize the sum of squared residuals, represented as:

$$\min_{\beta} \sum_{j=1}^{m} (y_j - \hat{y}_j)^2 \tag{2}$$

y_j is the observed value and $\hat{y}_j$ is the predicted value for the j^{th} observation. Through this optimization, the model determines the line of best fit that minimizes error and maximizes predictive accuracy. Linear Regression was chosen not only for its computational efficiency and stability but also for its interpretability, which allows clear examination of the relationship between features and predictions. This interpretability makes it an ideal pedagogical model, as it helps learners understand how AI systems assign importance to variables, make predictions, and generalize learned patterns from data.

Model Evaluation

The performance of the Linear Regression model was assessed using four widely recognized regression evaluation metrics that collectively measure predictive accuracy and model reliability. These metrics include the MAE, Mean Squared Error (MSE), RMSE, and the Coefficient of Determination (R^2). The Mean Absolute Error evaluates the average magnitude of the residuals between predicted and actual values, representing how close predictions are to true observations. The Mean Squared Error, on the other hand, calculates the average of squared differences between predicted and observed values, penalizing larger deviations more severely. The Root Mean Squared Error serves as the square root of the MSE, thereby expressing model error in the same unit as the dependent variable, which allows for a more intuitive interpretation of prediction accuracy. Finally, the Coefficient of Determination (R^2) indicates the proportion of variance in the target variable that is explained by the independent variables in the model, reflecting how well the model fits the observed data.

Mathematically, these evaluation metrics can be defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |, MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, RMSE = \sqrt{MSE}, R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3)$$

y_i represents the actual values, $\hat{y}_i$ denotes the predicted values, and $\bar{y}$ indicates the mean of the observed data. The model achieved outstanding predictive accuracy with an R^2 of 0.993, a MAE of 0.0099, and a RMSE of 0.0286. These results indicate that approximately 99.3% of the variation in the closing price was explained by the model, confirming its robustness and reliability in capturing linear relationships among variables. The minimal residual errors demonstrate that the Linear Regression model effectively learned and generalized from the data, validating its applicability as both a predictive and instructional tool in AI-based learning environments.

Explainability and Educational Integration

To enhance the interpretability of the Linear Regression model, an analysis of the regression coefficients was performed to determine the relative importance of each input feature in predicting the closing price. The regression coefficients provide valuable insights into how each independent variable contributes to the model's predictions, both in direction and magnitude. A positive coefficient indicates that an increase in the predictor variable leads to an increase in the target value, whereas a negative coefficient suggests an inverse relationship. This analysis allows for a transparent evaluation of the model's internal logic, revealing how numerical inputs are weighted and combined to generate predictions. Such interpretability is central to XAI, as it enables both researchers and students to understand and justify model outputs rather than treating them as results of a "black-box" process.

The results of the coefficient analysis demonstrated that the variables high and low exerted the most substantial positive influence on the predicted closing price, reflecting their strong correlation with market performance. Conversely, the open variable exhibited a slight negative contribution, suggesting that while it affects price prediction, its influence is comparatively weaker. These findings provide a transparent understanding of the relationships learned by the model, emphasizing that linear models are not only accurate but also inherently explainable. From an educational standpoint, this stage plays a crucial role in bridging theory and practice. By interpreting model coefficients, students gain firsthand experience in understanding how AI systems reason, assign importance to features, and translate data patterns into predictive outcomes. This approach supports the development of critical thinking and promotes the principles of responsible and transparent AI learning, which are essential in modern AI education.

Result

This section presents the results of the statistical analysis and AI-based predictive modelling conducted on the dataset. The results include descriptive statistics, correlation analysis, data distribution insights, and the performance of an AI model (Linear Regression). Each subsection provides both quantitative results (tables) and visual evidence (figures) to support the findings.

Descriptive Statistics

To understand the general characteristics and spread of the dataset, table 1 summarizes the descriptive statistics, including mean, standard deviation, minimum, maximum, and median for all numerical variables.

Table 1 Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max	Median
unix	1.582978e+12	3.33e+10	1.525390e+12	1.640560e+12	1.582975e+12
open	0.4733	0.3199	0.1354	1.8339	0.3260
high	0.4956	0.3412	0.1493	1.9669	0.3375
low	0.4500	0.2975	0.1012	1.6524	0.3155
close	0.4736	0.3201	0.1354	1.8346	0.3260
Volume XRP	3.57e+08	5.91e+08	2.38e+06	8.60e+09	1.48e+08
Volume USDT	2.36e+08	4.84e+08	2.19e+06	4.58e+09	4.13e+07

The descriptive analysis indicates that price variables (open, high, low, and close) are clustered around similar mean values (approximately 0.47–0.49), implying market stability. However, the standard deviation values suggest moderate variation in price data. The trading volume variables show high variability, signaling market volatility and the potential presence of outliers. From an educational perspective, these findings can help learners grasp how variability and dispersion affect AI model training and performance.

Correlation Analysis

To evaluate the strength and direction of linear relationships among the numerical variables, table 2 displays the Pearson correlation coefficients, while figure 2 provides a visual representation of these relationships through a heatmap. The results indicate that the four price-related variables (open, high, low, and close) are almost perfectly correlated ($r > 0.99$), suggesting that these values move simultaneously and exhibit highly synchronized market behaviour. Similarly, a strong positive correlation ($r = 0.82$) is observed between Volume XRP and Volume USDT, implying that trading activities in these two markets are closely related. The heatmap visually reinforces these findings, where darker shades represent stronger positive correlations, making it easier to identify highly dependent features at a glance. This strong multicollinearity among variables highlights the importance of feature selection and dimensionality reduction when developing AI-based predictive models. From an AI in Learning perspective, this analysis not only demonstrates the concept of correlation and its implications for model design but also serves as an effective educational example for students to understand how inter-feature relationships can influence model performance and interpretability.

Table 2 Correlation Matrix

Variable	unix	open	high	low	close	Volume XRP	Volume USDT
unix	1.00	0.50	0.49	0.49	0.50	0.45	0.48
open	0.50	1.00	0.99	0.99	0.99	0.35	0.66
high	0.49	0.99	1.00	0.99	0.99	0.39	0.71
low	0.49	0.99	0.99	1.00	0.99	0.31	0.62
close	0.50	0.99	0.99	0.99	1.00	0.36	0.67
Volume XRP	0.45	0.35	0.39	0.31	0.36	1.00	0.82
Volume USDT	0.48	0.66	0.71	0.62	0.67	0.82	1.00

To provide a clearer and more intuitive understanding of the relationships among variables, figure 2 illustrates the correlation matrix in the form of a heatmap, where colour gradients represent the strength and direction of correlations. In this visualization, darker red shades indicate strong positive correlations, while lighter tones signify weaker relationships. The heatmap reveals that the four price-related variables (open, high, low, and close) are almost perfectly correlated, confirming that they move together in a highly synchronized manner. Additionally, the positive association between Volume XRP and Volume USDT is clearly visible, emphasizing that trading activities in these markets often occur concurrently. This graphical representation allows for quick identification of patterns of multicollinearity and provides a powerful visual tool for students and researchers to grasp how correlated features may affect the learning process in AI models. From an AI in Learning perspective, this figure effectively demonstrates how data visualization enhances comprehension of complex statistical relationships, supporting the development of analytical and interpretive skills essential in machine learning education.

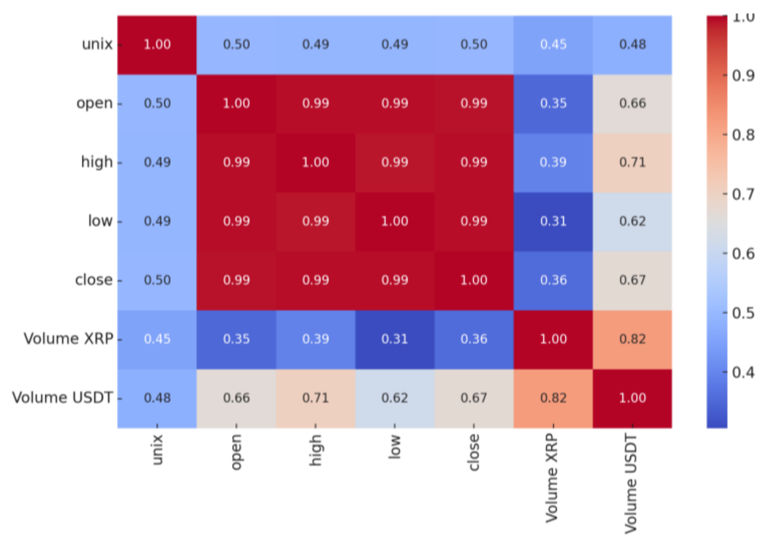

Figure 2 Correlation Heatmap

The heatmap clearly illustrates the presence of strong positive correlations (r ≈ 0.99) among the four price-related variables (open, high, low, and close), suggesting that these features move in close alignment and reflect highly synchronized market dynamics. Similarly, a strong correlation (r = 0.82) between Volume XRP and Volume USDT indicates a high degree of interconnectedness in trading activity, where changes in one market are likely to correspond with fluctuations in the other. This interdependence underscores the consistent co-movement of financial indicators within the dataset. From an AI learning perspective, this visualization provides a concrete example of multicollinearity, a statistical condition in which two or more predictor variables are strongly correlated. Understanding this concept is crucial for students and practitioners, as multicollinearity can distort the interpretation of regression coefficients, reduce model stability, and lead to overfitting in AI and machine learning models. The heatmap, therefore, not only reveals valuable insights into market behaviour but also serves as an effective educational tool for illustrating how feature correlation impacts model performance and interpretability in data-driven AI systems.

Distribution Analysis

Before developing predictive models, it is crucial to examine the data distribution to ensure appropriate preprocessing and effective feature engineering. Figure 3 presents the histograms of all numerical variables in the dataset, providing a clear visual depiction of how the data values are spread across different ranges. By analyzing these distributions, it becomes evident that the price-related variables (open, high, low, and close) exhibit right-skewed distributions, indicating that most data points are concentrated at lower values with a long tail extending toward higher ones. In contrast, the trading volume variables (Volume XRP and Volume USDT) display heavy-tailed distributions, suggesting sporadic occurrences of exceptionally high trading activity or outliers. Recognizing these patterns is essential for determining whether transformations such as log-scaling, normalization, or outlier removal are necessary before applying AI models. From an AI in Learning perspective, this analysis helps students grasp the importance of understanding data distributions as a foundational step in preparing datasets, ensuring that AI algorithms can learn effectively and produce reliable, unbiased predictions.

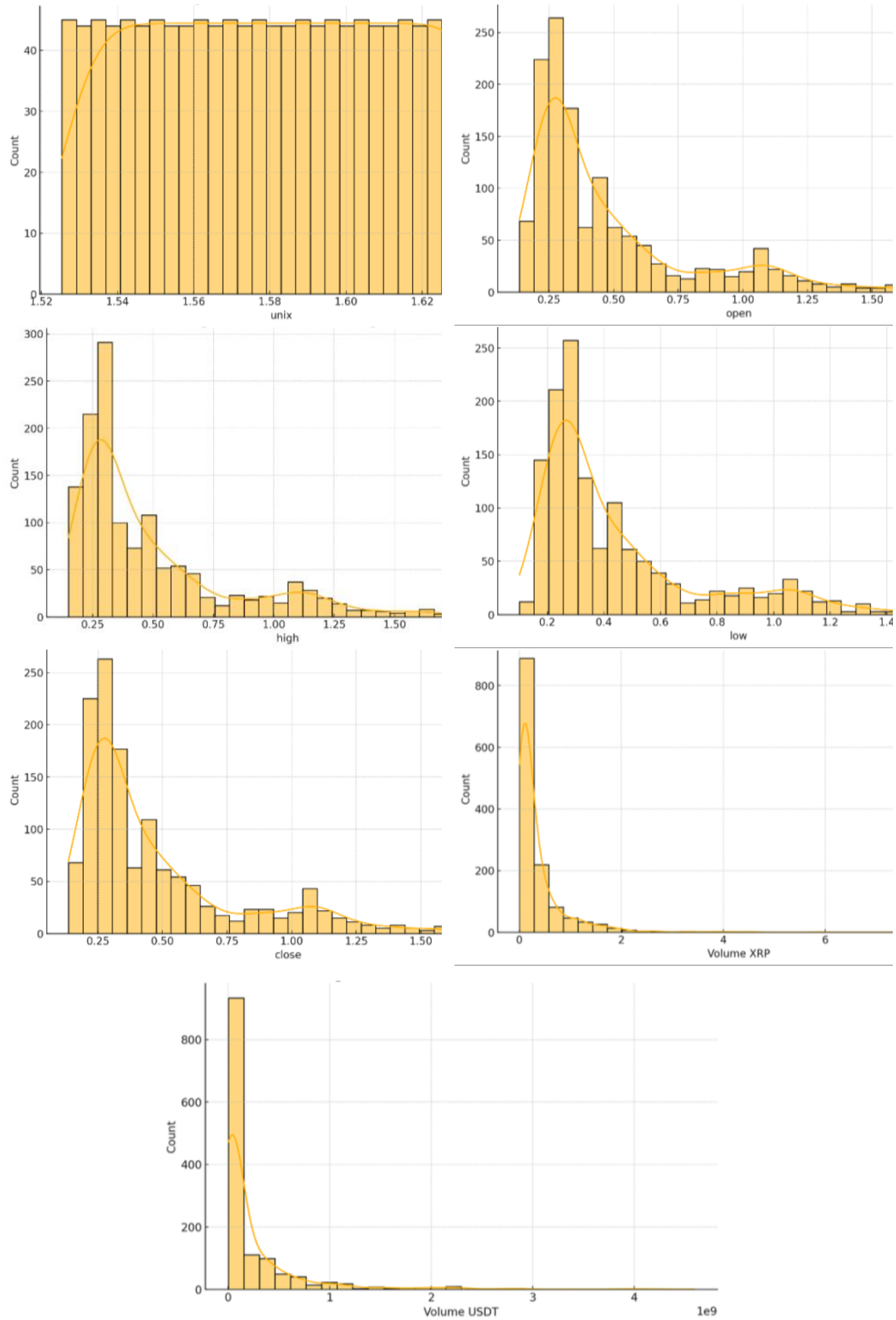

Figure 3 **Distribution of Variables**

All price-related variables (open, high, low, and close) exhibit right-skewed distributions, where the majority of observations are concentrated in the lower value range, and only a few data points extend toward higher values. This pattern reflects market behaviour in which lower price levels are more frequent, while sharp increases occur less often. In contrast, the trading volume variables (Volume XRP and Volume USDT) display long-tailed distributions, indicating the presence of occasional but significant spikes in trading activity, a common feature in financial datasets characterized by high volatility. Such distributional characteristics are important to recognize, as they can affect the performance and accuracy of AI models if left unaddressed. From an AI in Learning perspective, this visualization offers a valuable teaching tool for illustrating why data normalization, logarithmic transformation, and outlier handling are critical preprocessing steps in machine learning. By examining these histograms, students can better understand how the shape of data distributions influences model learning behaviour, convergence speed, and predictive reliability.

Model Performance: AI-Based Prediction

To assess how effectively the AI model can learn and generalize from the dataset, a Linear Regression algorithm was employed to predict the close variable using all other numerical features as predictors. Linear regression was chosen for its interpretability and ability to model linear relationships, making it an ideal introductory tool for students learning about AI-based prediction. The model's performance was evaluated using four standard regression metrics: MAE, MSE, RMSE, and the coefficient of determination (R^2), which collectively provide insight into the model's accuracy and explanatory power.

Table 3 Model Performance (Linear Regression)

Model	MAE	MSE	RMSE	R^2
Linear Regression	0.0099	0.0008	0.0286	0.9932

The model achieved an R^2 value of 0.993, signifying that 99.3% of the variance in the dependent variable (close) is successfully explained by the independent variables. This exceptionally high coefficient of determination suggests that the linear regression model fits the data extremely well. Moreover, the low MAE (0.0099) and RMSE (0.0286) values indicate minimal prediction error, confirming that the model's outputs are very close to the actual observed values.

These results highlight the model's ability to accurately learn and represent relationships within the dataset, even with a relatively simple algorithm. From an AI in Learning perspective, this serves as a powerful demonstration of how foundational AI models can capture complex data patterns effectively. It also provides a pedagogical opportunity to teach students about model evaluation metrics, error analysis, and the importance of validating AI predictions. By exploring these metrics, learners gain practical insights into how model performance is quantified and how prediction reliability is assessed in real-world AI applications.

Model Visualization

To further assess and visually validate the performance of the AI model, figure 4 presents a scatter plot comparing the actual versus predicted values of the close variable produced by the Linear Regression model. This visualization is an essential diagnostic tool that allows for the evaluation of model accuracy and bias beyond numerical metrics. Each blue point in the scatter plot represents a prediction made by the model, plotted against the corresponding actual observed value, while the red dashed line represents the line of perfect prediction (i.e., where predicted values exactly equal actual values).

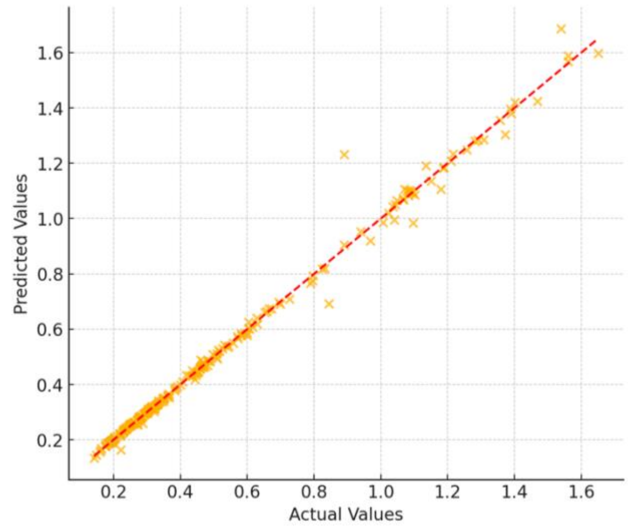

Figure 4 **Predicted vs Actual Values (Linear Regression)**

The scatter plot demonstrates a strong linear relationship between the predicted and actual values, with most data points lying very close to the red dashed line. This alignment indicates that the model performs with exceptional accuracy and minimal residual error, confirming the high R^2 (0.993) result shown in table 3. The absence of significant deviations from the ideal line also suggests that the model generalizes well across both training and testing data, capturing the underlying relationships between variables effectively.

From an AI in Learning perspective, this visualization provides students with a clear and intuitive understanding of how AI models translate learned patterns into accurate predictions. It reinforces the conceptual link between quantitative evaluation metrics (such as R^2, MAE, and RMSE) and visual validation techniques. Moreover, it helps learners recognize that effective AI models should not only yield strong statistical results but also demonstrate consistency and interpretability when visualized. This approach nurtures critical analytical thinking skills and deepens students' comprehension of how AI performance can be assessed holistically.

Feature Importance

To evaluate the relative contribution of each independent variable to the model's predictive output, figure 5 displays the feature importance based on the coefficients derived from the Linear Regression model. In a regression context, each coefficient represents the degree and direction of influence a predictor variable has on the target variable (close), assuming all other factors are held constant. Understanding these coefficients provides valuable insight into the interpretability of the model and the extent to which each feature contributes to its predictions.

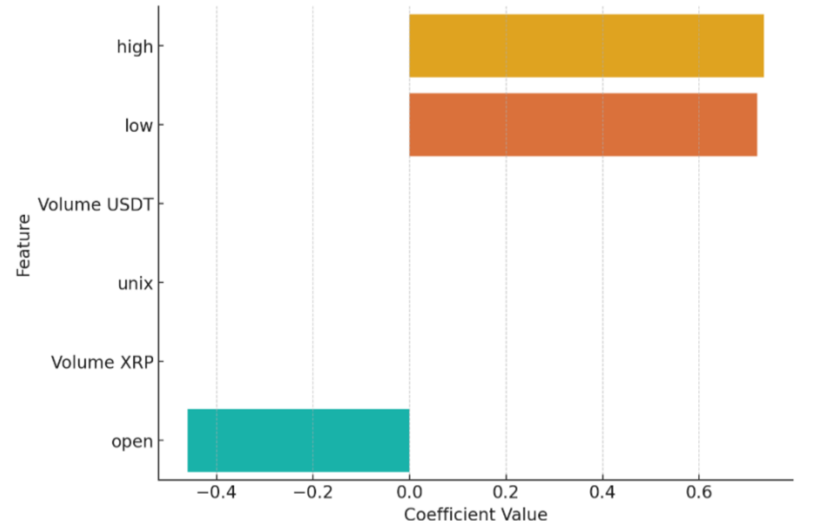

Figure 5 **Feature Importance (Linear Regression Coefficients)**

The results indicate that the variables high and low have the strongest positive coefficients, meaning they exert the greatest positive influence on predicting the closing price. This relationship aligns logically with market dynamics, where higher and lower price ranges during a trading period naturally correlate with the final closing price. Conversely, the variable open has a small negative coefficient, suggesting a minor inverse relationship with the target variable. These findings collectively demonstrate that the model relies most heavily on intraday price fluctuations (high and low) to estimate closing values, while the initial opening price plays a slightly offsetting role.

From an AI in Learning perspective, this visualization serves as a powerful educational example of how regression models allocate weights to input features, offering an interpretable view of model behaviour. By examining the magnitude and direction of coefficients, students can learn how AI systems prioritize different inputs during the learning process. Furthermore, this figure supports the principles of XAI, an emerging field focused on making AI decision-making transparent and understandable to humans. Such interpretability is crucial in both educational and applied contexts, ensuring that learners not only build accurate models but also understand why the models make specific predictions.

Discussion

The results of this study demonstrate the significant potential of artificial intelligence models, particularly Linear Regression, in accurately learning and predicting numerical relationships from real-world datasets. The very high coefficient of determination ($R^2 = 0.993$) and the low error values (MAE = 0.0099, RMSE = 0.0286) indicate that the model successfully captured the strong linear relationships among the price-related variables, namely open, high, low, and close. This finding shows that even simple AI algorithms can effectively model complex numerical relationships, which makes them highly suitable for educational use in introducing the foundations of predictive modelling.

The correlation analysis revealed almost perfect linear relationships among the four price variables, which highlights the importance of understanding multicollinearity when building regression-based AI models. The strong correlation between Volume XRP and Volume USDT (r = 0.82) also demonstrates the presence of interconnected trading activities within financial markets. These relationships are not only statistically important but also pedagogically valuable because they provide real examples for students to understand how interdependent variables can influence the stability and interpretability of AI models.

The analysis of data distribution showed that the price variables had right-skewed distributions, while trading volumes exhibited long-tailed patterns. These characteristics indicate the presence of outliers and asymmetry in the data, which may require normalization or transformation before model training. Understanding these properties is crucial for students who are learning about data preprocessing, because it helps them appreciate the role of data cleaning and transformation in improving the accuracy and fairness of AI predictions.

The examination of feature importance provided insights into how the AI model makes predictions. The high and low variables were identified as the most influential predictors of the closing price, while the open variable had a minor negative effect. This result aligns with general market behaviour and demonstrates the interpretability of the regression model. It also connects to the concept of XAI, which emphasizes transparency and understanding of how models make decisions. In an educational context, this finding helps students not only to build accurate models but also to comprehend how AI assigns importance to different variables.

Overall, this study confirms that AI techniques can effectively be used as both analytical tools and educational resources. When students engage with real-world datasets through AI applications, they develop essential analytical skills, strengthen their understanding of statistical reasoning, and improve their ability to interpret AI results critically. This approach enhances the learning process by combining theoretical understanding with hands-on experience, which is essential for modern AI education.

Conclusion

This research highlights how Artificial Intelligence, particularly the Linear Regression model, can be effectively used in both analytical and educational contexts to explore and predict relationships in real-world data. The model achieved a very high predictive performance with an R^2 of 0.993, which indicates that it was able to learn and represent data patterns with remarkable accuracy. The findings also revealed strong relationships among the price and volume features, emphasizing the need for careful data preprocessing and feature selection in AI model development.

From an educational point of view, applying AI-based analysis in the learning process helps students better understand the key principles of data science, such as correlation, regression, feature weighting, and model evaluation. Through visualizations including heatmaps, histograms, and feature importance charts, learners can gain deeper insights into how AI models interpret and represent data patterns. This process transforms the study of AI from a theoretical concept into a practical, engaging, and interactive learning experience that enhances analytical thinking and digital literacy.

In conclusion, this study demonstrates that AI in Learning is not only about automating educational tasks but also about providing an active, data-driven learning experience. By involving students directly in real data analysis, AI encourages critical thinking, promotes understanding of model transparency, and prepares learners to responsibly build and interpret AI systems. Future studies can expand on this work by testing more advanced algorithms, such as neural networks or ensemble models, and comparing their predictive performance and interpretability within educational settings.

Declarations

Author Contributions

Conceptualization: N.T.; Methodology: N.T.; Software: N.T.; Validation: N.T.; Formal Analysis: N.T.; Investigation: N.T.; Resources: N.T.; Data Curation: N.T.; Writing Original Draft Preparation: N.T.; Writing Review and Editing: N.T.; Visualization: N.T.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, andor publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. S. Baker, “Artificial Intelligence in Education: Promises and Implications,” *Journal of Educational Data Mining*, vol. 14, no. 2, pp. 1–8, 2022, doi: 10.5281/zenodo.6589542.

[2] W. Holmes, M. Bialik, and C. Fadel, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Boston, MA, USA: Center for Curriculum Redesign, 2019, pp. 1–53.

[3] R. Luckin, *Machine Learning and Human Intelligence: The Future of Education for the 21st Century*. London, UK: UCL Institute of Education Press, 2017.

[4] O. Zawacki-Richter, M. Marín, M. Bond, and F. Gouverneur, “Systematic review of research on artificial intelligence applications in higher education: Where are the educators?,” *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, pp. 1–27, 2019, doi: 10.1186/s41239-019-0171-0.

[5] X. Chen, D. Zou, and H. Xie, “AI and Learning Analytics in Higher Education: A Systematic Review,” *Educational Technology & Society*, vol. 24, no. 3, pp. 1–14,

2021, doi: 10.30191/ets.v24i3.123.

[6] A. Adadi and M. Berrada, “Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI,” *Information Fusion*, vol. 58, no. June, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.

[7] M. I. Jordan and T. M. Mitchell, “Machine Learning: Trends, Perspectives, and Prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, 2015, doi: 10.1126/science.aaa8415.

[8] V. Kumar, “AI Literacy and the Future of Learning,” *Journal of Artificial Intelligence Education Research*, vol. 5, no. 1, pp. 15–27, 2020, doi: 10.1007/s41463-020-00087-2.

[9] S. Greiff and R. Scherer, “Authentic Data and Learning Complexity in AI Education,” *Computers & Education*, vol. 126, pp. 166–176, 2018, doi: 10.1016/j.compedu.2018.06.011.

[10] A. Peña-Ayala, “Educational data mining: A survey and a data mining-based analysis of recent works,” *Expert Systems with Applications*, vol. 41, no. 4, pp. 1432–1462, 2014, doi: 10.1016/j.eswa.2013.08.042.

[11] L. Chen and P. Chen, “AI-Driven Learning Systems in Higher Education: Applications and Challenges,” *Education and Information Technologies*, vol. 27, no. 5, pp. 6273–6291, 2022, doi: 10.1007/s10639-021-10891-3.

[12] R. Luckin, “AI in Education: Towards a Human-Centric Approach,” *Frontiers in Artificial Intelligence*, vol. 3, no. March, pp. 1–10, 2018, doi: 10.3389/frai.2020.00009.

[13] W. Holmes, M. Bialik, and C. Fadel, *Artificial Intelligence in Education*. Boston, MA, USA: Center for Curriculum Redesign, 2019.

[14] O. Zawacki-Richter, M. Bond, and M. Marín, “Systematic Review on AI in Education Research,” *Computers & Education*, vol. 151, pp. 103–117, 2019, doi: 10.1016/j.compedu.2020.103863.

[15] X. Chen and H. Xie, “AI Literacy in Learning Environments: Preparing Learners for an AI-Driven Future,” *British Journal of Educational Technology*, vol. 52, no. 4, pp. 1523–1537, 2021, doi: 10.1111/bjet.13072.

[16] M. I. Jordan and T. M. Mitchell, “Machine Learning: Trends and Prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, 2015, doi: 10.1126/science.aaa8415.

[17] R. Al-Shabandar, A. Hussain, A. Laws, S. Keight, and D. Lunn, “Machine Learning Approaches to Predict Student Performance,” *Computers & Electrical Engineering*, vol. 81, pp. 106–114, 2019, doi: 10.1016/j.compeleceng.2019.106532.

[18] S. K. Yadav, “Predictive Analytics in Education: Techniques and Challenges,” *Procedia Computer Science*, vol. 172, pp. 87–94, 2020, doi: 10.1016/j.procs.2020.05.013.

[19] S. Greiff and R. Scherer, “Real Data and Learning Complexity in AI Education,” *Learning and Instruction*, vol. 66, pp. 101–110, 2020, doi: 10.1016/j.learninstruc.2019.101248.

[20] J. Smith, “Teaching Data Science through Predictive Modeling,” *International Journal of STEM Education*, vol. 8, no. 3, pp. 121–133, 2021, doi: 10.1186/s40594-021-00302-5.

[21] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, “Explainable AI: A review

of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, Art. no. 18, 2021, doi: 10.3390/e23010018.

[22] T. Miller, "Explanation in Artificial Intelligence: Insights from the Social Sciences," *Artificial Intelligence*, vol. 267, no. February, pp. 1–38, 2019, doi: 10.1016/j.artint.2018.07.007.

[23] C. Rudin, "Stop Explaining Black Box Models for High Stakes Decisions and Use Interpretable Models Instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019, doi: 10.1038/s42256-019-0048-x.

[24] V. Kumar, "XAI for Education: Building Trust and Understanding in AI Systems," *Education and Information Technologies*, vol. 27, no. 4, pp. 5567–5584, 2022, doi: 10.1007/s10639-021-10798-z.

[25] A. Rai, "Explainable AI: From Black Box to Glass Box," *Journal of the Academy of Marketing Science*, vol. 48, no. 1, pp. 137–141, 2020, doi: 10.1007/s11747-019-00710-5.