Massimiliano FERRARA, PhD (corresponding author)
massimiliano.ferrara@unirc.it
University "Mediterranea" of Reggio Calabria, Italy
ICRIOS - The Invernizzi Centre for Research in Innovation, Organization, Strategy and Entrepreneurship Bocconi University, Milan, Italy

Tiziana CIANO, Researcher
t.ciano@univda.it
University of Aosta Valley, Aosta, Italy
Decisions_Lab, University "Mediterranea" of Reggio Calabria, Italy

Improving the Interpretability of Asset Pricing Models by Explainable AI: A Machine Learning-based Approach

Abstract. *The study examines the integration of machine learning (ML) techniques and explainable artificial intelligence (XAI) for stock price prediction and the interpretation of predictive models. The aim is to improve the accuracy of short-term price forecasts using advanced models like Random Forest and XGBoost, and to utilise XAI tools such as SHAP and LIME to better understand the contribution of each variable in the predictions.*
This approach can be particularly useful for investors interested in sustainability-related securities, such as those with high ESG ratings, as it provides a deeper understanding of market dynamics and allows for more informed and transparent investment decisions. The integration of XAI not only enhances prediction accuracy, but also helps mitigate the risks associated with understanding and trusting machine learning algorithms, ensuring that these can be used with greater awareness and control, especially in a complex and regulated context like finance.

Keywords: *asset pricing, Machine Learning, Explainable Artificial Intelligence, SHAPE and LIME, feature, ESG, interpretability.*

JEL Classification: C61, C38, C53, G17.

1. Motivation of the Study and Literature Review

In recent years, Machine Learning (ML) has revolutionised many sectors, including finance, where it has had a significant impact on asset pricing. Traditional asset pricing models, such as the Capital Asset Pricing Model (CAPM) or the Fama-French three-factor model, are based on linear relationships between risk factors and expected returns (Sharpe, 1964; Fama and French, 1993). These models assume that market behaviour and asset returns can be explained by a limited number of factors and that the relationships are relatively stable and simple. However, real markets are much more complex, characterised by non-linearity, time-varying relationships, and

DOI: 10.24818/18423264/58.4.24.01

vast amounts of data. This is where machine learning becomes a determining factor (Khandani et al.,2010; Gu et al., 2020).

Financial markets are often characterised by complex and non-linear interactions among variables, which traditional asset pricing models struggle to capture effectively. Models like CAPM or Fama-French tend to consider linear and stable relationships, limiting their ability to represent the more sophisticated dynamics of real markets. In contrast, machine learning algorithms (ML), particularly techniques such as decision trees, neural networks, and support vector machines (SVM), provide a more advanced approach to discovering and modelling these complex and non-linear patterns, offering a more accurate view of the markets (Doshi-Velez & Kim, 2017; Biran & Cotton, 2017).

In addition to managing the complexity of relationships among variables, financial markets generate a vast amount of data. This includes prices, volumes, order flows, economic indicators, and even textual data from news or social media. ML algorithms are particularly effective in processing and analysing high-dimensional datasets, managing to identify connections among variables that may be difficult to detect or interpret with traditional econometric methods. The ability to work with heterogeneous and large volumes of data represents an additional advantage of ML over conventional models (Linardatos et al., 2020; Giglio & Xiu, 2021; Yao et al., 2000).

Another key aspect of machine learning in asset pricing is feature selection and dimensionality reduction. Asset returns are influenced by many factors, but not all are relevant or useful to improve predictions. ML techniques, such as Lasso regression or principal component analysis (PCA), allow for the identification of the most significant factors from a wide set of variables, reducing noise, and improving the predictive accuracy of the models. This enables a focus on the most relevant aspects for market dynamics (Fama & French, 1996; Ribeiro et al., 2016).

Predictive strength is another area where ML stands out. Traditional models are often limited in accurately predicting future market movements, whereas ML models, specifically designed for prediction rather than explanation, excel in this area. Algorithms such as random forests, gradient enhancement, and deep learning have demonstrated remarkable superiority in predictive capabilities, especially when used with large datasets that traditional models cannot handle as effectively.

Finally, financial markets are dynamic and constantly evolving. The relationships among variables change over time, and traditional asset pricing models, which often assume stationarity (i.e., stability of relationships over time), can quickly become outdated. ML models, particularly techniques like online learning or reinforcement learning, have the ability to adapt to market changes, remaining relevant even when market conditions shift. This adaptability gives ML a significant advantage, allowing for the development of more flexible and responsive pricing models (Gu et al., 2021; Achituve et al., 2019; Tsang, & Wong, 2019; Heaton et al., 2016).

While machine learning models provide substantial benefits in predicting and managing the intricacies of financial markets, a key criticism is their "black box" characteristic. Often, it remains unclear how a model reaches a specific decision or prediction, which can pose challenges concerning trust, transparency, and regulatory compliance within the financial industry. This is where Explainable AI (XAI) techniques play a vital role (Ertel, 2018; Adadi & Berrada, 2018; Arrieta et al., 2020).

XAI techniques enhance the interpretability of machine learning models, offering transparent explanations of the decision-making processes involved. In asset pricing, this allows a deeper comprehension of the factors that affect asset values and how various variables impact returns. Widely used XAI methods, such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations), deliver both local and global insights into algorithmic decisions (Chen& Guestrin, 2016; Cao, 2021).

These tools are crucial for reconciling the predictive performance of machine learning models with the necessity for clarity and interpretability, particularly in a regulated environment like finance, where justifying and comprehending predictions is essential. Moreover, employing XAI techniques boosts the confidence of analysts and asset managers in machine learning models, thereby facilitating their incorporation into strategic business decisions (Yang et al., 2022; Ye et al., 2024).

This study aims to integrate Explainable AI (XAI) techniques into machine-learning-driven asset pricing models, providing clear and detailed explanations of how each variable influences return predictions. This ensures that investors can trust the forecasts and comprehend the reasoning behind certain decisions. Such an approach could transform the application of machine learning models in financial markets, fostering enhanced transparency and improved risk management.

The paper is organised as follows: Section 1 reviews the existing literature on the applications of machine learning and explainable artificial intelligence. Section 2 outlines the machine learning and XAI models utilised in our analysis, while Section 3 discusses the results obtained from these models. Section 4 examines the outcomes of expanding the feature set within the dataset. Finally, Section 5 offers insight into the conclusions drawn from the study.

2. Models

2.1 Machine learning Models involved

We use the Random Forest and XGBoost models to train a system that makes predictions. Both models belong to the category of ensemble learning models, which combine the predictions of multiple simple models (such as decision trees) to improve performance.

In particular, Random Forest is a machine learning algorithm that relies on a combination of multiple decision trees. Each tree is built using a random sample of the data and a random selection of variables. The final predictions are obtained by averaging the results of all trees, leveraging diversity to achieve more robust predictions.

Meanwhile, XGBoost (Extreme Gradient Boosting) is an advanced algorithm that constructs models sequentially. Unlike Random Forest, where trees are created in parallel, XGBoost builds new trees to correct the errors made by previous trees. This gradual approach allows for continuous improvement of the predictions.

2.2 More on Explainable Artificial Intelligence (XAI): SHAP and LIME

SHAP (Shapley Additive Explanations) is a technique that is based on Shapley values, a concept derived from game theory. The key idea is to assign a "value" or "contribution" to each feature of the model in a fair and consistent manner, determining how much each variable contributes to the final outcome of the model. SHAP provides global explanations (for the entire model) and local explanations (for a single prediction).

SHAP values quantify the contribution of each feature by comparing the prediction that the model makes when a feature is included versus when it is excluded, across all possible subsets of features. More precisely, Shapley values are calculated as the average of the marginal contributions of a feature, considered across all possible orders (Saeed & Omlin, 2023).

Given a complex predictive function f(x) (for example, a Random Forest model or an XGBoost model), the goal is to calculate the contribution of each input variable x_i (such as historical price, volume, etc.) to the overall prediction f(x).

The Shapley values for the variable x_i are calculated as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} (f(S \cup \{i\}) - f(S))$$

where:

- N is the total set of variables (e.g., historical price, volume, volatility, macroeconomic indicators).
- S ⊆ N$\setminus\{i\}$ s a subset of the variables excluding the variable i.
- $f\,(S)$ s the prediction of the model considering only the subset of variables S.
- $f\,(S \cup \{i\})$ is the prediction of the model considering the subset S plus the variable i.
- φ_i is the SHAP value for the variable which measures its contribution to the final prediction.

In the context of asset pricing:
- x_i represents a variable such as historical price, volume, or a macroeconomic indicator (inflation rate, GDP, etc.).
- φ_i quantifies how much each variable contributes to the forecast of the return of a certain asset.

2.3 LIME

LIME (Local Interpretable Model-agnostic Explanations) is a technique that provides local explanations for machine learning models. The LIME approach is based on the idea that even if a model is complex and difficult to interpret globally, it can be locally approximated (i.e., in a small region around a specific prediction) with a simple and interpretable model, such as a linear regression or a decision tree.

LIME constructs an interpretable model that approximates the behaviour of the complex model in a local region, close to the prediction that is to be explained.

LIME approximates the complex predictive function $f(x)$using an interpretable model $g(x)$(for example, a linear regression or a decision tree) in a local region around an input data point x_0 (the asset in question).

The goal is to minimise the difference between the complex function $f(x)$ and the interpretable model $g(x)$ around x_0 giving greater weight to points close to x_0.

$$\hat{g} = arg \min_{g \in G} \mathcal{L}(f, g, \pi_{x_0}) + \Omega(g)$$

where:

$-\mathcal{L}(f, g, \pi_{x_0})$ is the loss (for example, the mean squared error) between the predictions of the complex model and those of the interpretable model g, weighted by the proximity of x to x_0.

$-\pi_{x_0}$ is a weighting function that gives greater importance to points close to x_0 (for example, a Gaussian kernel function centred at x_0).

$-\Omega(g)$ is a measure of the complexity of the interpretable model g, to avoid overly complex models (for example, by limiting the number of variables used in the explanation).

In the context of asset pricing, LIME approximates the behaviour of the complex model (such as a neural network) around a single prediction x_0, for example, the predicted return of a specific asset. The interpretable model $g(x)$ can be a linear regression that uses only a few variables such as historical price, volume, and sentiment.

3. Results and related discussion on

3.1 Dataset

In this work, we analyse the stocks of companies that have a strong commitment to environmental sustainability and are known for having a high ESG (Environmental, Social, and Governance) rating. These stocks represent companies that are contributing to the transition towards a greener economy, focusing on renewable energy, electric vehicles, and technological solutions to reduce environmental impact. In particular:

- Tesla (TSLA): A leader in the production of electric vehicles and the development of renewable energy.
- NextEra Energy (NEE): One of the largest providers of renewable energy in the United States, with a strong commitment to solar and wind energy.
- Vestas (VWDRY): One of the largest manufacturers of wind turbines in the world.
- Enel (ENEL): An Italian company active in renewable energy and committed to environmental sustainability projects.

These stocks are particularly interesting in the context of sustainable finance, where investors are increasingly interested in how ESG factors influence asset returns. The model is designed to predict the short-term return of an asset, which is crucial for evaluating investment opportunities, both for short-term traders and institutional investors.

The objective of this work is to utilise machine learning techniques and Explainable AI to predict the closing price of an asset for the following day and to explain how different variables influence these predictions. Daily data is downloaded from Yahoo Finance, and the reference period is from January 1, 2018, to January 1, 2023. This period covers key moments for the renewable energy and sustainability industry. For example, Tesla has accelerated its production of electric vehicles, NextEra and Enel have invested heavily in renewable energy, and Vestas has continued to lead the wind turbine market.

These data are essential for understanding how these companies have performed in a market increasingly influenced by environmental regulations and the growing demand for sustainable solutions.

3.2 Features

The analysis forecasts the future closing price of each stock, a crucial activity for investors who wish to understand how financial and sustainability factors influence short-term returns. Companies with high ESG ratings may have long-term competitive advantages, but in this study, we seek to understand whether historical factors can help predict short-term price movements. Investors aim to anticipate the performance of these companies to invest in a rapidly growing sector like sustainability. The variables used to predict the price for the following day are: Open, High, Low, Volume.

These variables reflect market behaviour. For instance, trading volume may indicate increasing or decreasing interest in a company like Tesla, which could reflect concerns about the adoption of electric vehicles, or Enel, in response to government policies incentivising renewable energy.

In the context of asset pricing for sustainable companies, these variables represent market sentiment regarding the positive impact of companies on ESG factors, which can lead to price movements.

3.3 Mean Squared Error

The Random Forest model produced predictions that were closer to the actual values compared to the XGBoost model in your specific dataset. In contrast, XGBoost generated predictions with higher average errors, suggesting that the model may not have captured the market dynamics well in the dataset (see Table 1).

Table 1. Mean Squared Error of Models

Model	Mean Squared Error (MSE)
Random Forest	13.864210669656801
XGBoost	20.787944732946826

Source: Authors' processing.

3.4 Explainable Artificial Intelligence (XAI)

After training the machine learning models, we use Explainable AI techniques such as SHAP and LIME to make the predictions interpretable, allowing us to understand the contribution of each variable in the prediction models. In Graph 1, we can observe the impact of the analysed variables. In particular, the "Low" and "High" variables show a strong positive impact on the prediction. The presence of a higher concentration of points towards high SHAP values indicates that an increase in these variables tends to raise price predictions. This suggests a direct and clear relationship between the opening and closing values of the stocks. In contrast, the "Open" variable has a distribution more concentrated around values close to zero, suggesting that, in the context of the Random Forest model, the opening price does not have a significant influence on the closing price predictions. On the other hand, the "Volume" variable may have a wider distribution, indicating a strong positive impact: in other words, high trading volume is associated with higher price predictions.

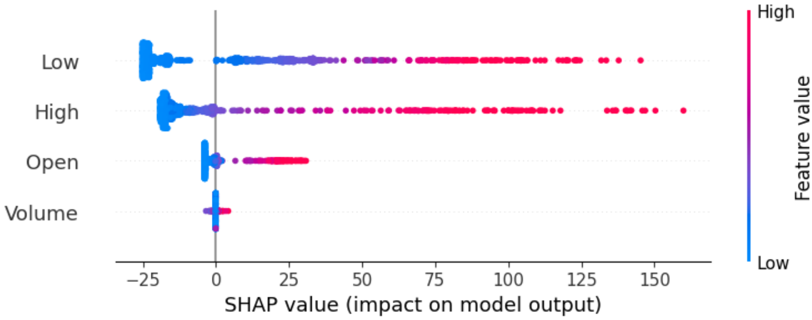

Figure 1. SHAP: Random Forest Model
Source: Authors' processing.

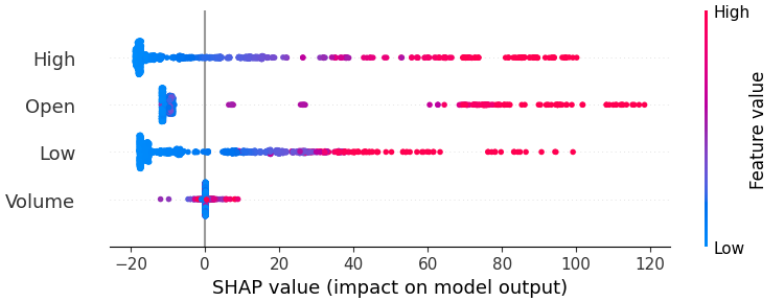

Figure 2. SHAP: XGBoost Model
Source: Authors' processing.

The "Low" variable also demonstrates a positive impact, although its distribution may not be as broad as in the Random Forest model. Finally, if the points related to the "Volume" variable are distributed towards positive SHAP values, this suggests that trading volume remains an important factor even in the Gradient Boosting model, although its distribution may vary compared to what was observed in the Random Forest model.

The interpretation of the SHAP graphs in relation to the models used provides investors with significant insights into the dynamics of the analysed stocks. The importance of the "Low" and "High" variables suggests that these should be closely monitored during the decision-making process. Furthermore, the difference in the importance of the variables between the two models indicates the need to develop a more nuanced investment strategy that takes into account the specific dynamics of the machine learning models used.

Table 2 presents the results of the machine learning models, Random Forest and XGBoost, using LIME. For the Random Forest model, the intercept is approximately 74.53, indicating the expected value of the target variable when all other independent variables are equal to zero. The local prediction for this model is about 8.65, suggesting that for a specific observation, the model predicts a value close to this number. The predicted value ranges from a minimum of 4.20 to a maximum of

340.44, showing quite a wide range of results, which may suggest some variability or uncertainty in the model.

Regarding the analysed variables, the value of "Low" is 5.92 and its impact on the prediction is 35.43, meaning that a lower "Low" value has a positive relationship with the model's prediction. The "High" variable has a value of 5.96 and an impact of 27.45, indicating a positive relationship with the predicted outcome as well. The "Open" variable shows a value of 5.95 with an impact of 2.10, suggesting that the opening price has a limited influence on the prediction. Finally, the "Volume" variable has a value of 180,000.00 and an impact of 0.89, implying that while trading volume has an effect, it is not as significant as other variables.

Table 2. LIME Explanations

Metric	Random Forest	XGBoost
Intercept	74.52509094250547	74.83407583121094
Prediction_local	8.64794522	8.4559072
Predicted_value_min	4.2	-3.28
Predicted_value_max	340.44	331.39
Low_value	5.92	5.92
Low_impact	35.43	23.43
High_value	5.96	5.96
High_impact	27.45	28.02
Open_value	5.95	5.95
Open_impact	2.1	16.13
Volume_value	180000.0	180000.0
Volume_impact	0.89	1.19

Source: Authors' processing.

Moving to the XGBoost model, the intercept is slightly higher, at 74.83, while the local prediction is about 8.46. Here, the predicted value ranges from a minimum of -3.28 to a maximum of 331.39. This range indicates that the XGBoost model has similar predictive potential but also presents a negative prediction, suggesting that there are observations where the model may not effectively capture the dynamics of the data.

Regarding the variables, the value of "High" is 5.96 with an impact of 28.02, similar to the Random Forest model, indicating a positive correlation. The "Low" variable maintains the same value of 5.92, but its impact drops to 23.43.

The "Open" variable has a value of 5.95 with a more significant impact of 16.13, suggesting that in the XGBoost model, the opening price plays a more relevant role. Finally, the "Volume" remains at 180,000.00, but its impact increases to 1.19, indicating that the trading volume has a greater role compared to what was observed in the Random Forest model. Therefore, both models provide similar predictions, but there are differences in the importance of the variables, suggesting that each model may have its own characteristics and dynamics that influence the forecasts.

This type of analysis is crucial for investors seeking to understand which factors affect the outcomes in the context of investments and trading. We believe that using machine learning techniques and Explainable AI on financial data related to companies with high ESG ratings allows us to better understand the dynamics underlying the performance of these assets, providing useful tools for evaluating short-term investment opportunities in sustainability-related sectors.

4. Integration of derived variables to improve asset pricing forecasts

The initial variables used (such as Open, High, Low, Volume) simply represent the data of the current day. These variables are useful, but they provide only a snapshot of the market situation on a specific day. Predicting future prices based solely on these variables has some limitations. Stocks can experience significant fluctuations from day to day due to external factors such as economic news, geopolitical events, or corporate announcements. Without a historical view of the stock's behaviour, it is difficult for the model to identify patterns or trends.

Moreover, the initial variables do not take into account past dynamics or market trends; thus, the model may miss crucial information needed to understand the future direction of prices. Therefore, to improve the predictions of our models, we have considered new features such as:

- MA_50: 50-day moving average of the closing price, which helps capture long-term market trends.
- Volatility_50: 50-day volatility (standard deviation of the price), which indicates the stability or instability of the stock.
- Close_lag_1: Closing price of the previous day, useful for capturing short-term trends.
- Volume_lag_1: Volume of the previous day, which can indicate buying or selling pressure on the stock.

These variables provide the model with a significantly greater amount of information. They enrich the dataset with indicators that consider the trends and recent history of price and volume. Let’s see how each of these derived variables contributes to improving predictions:

Moving averages smooth out daily fluctuations and capture long-term trends, allowing the model to detect market directions that may not be evident in daily data. Volatility measures the variability of prices over a given period, indicating market instability; by integrating it into the model, forecasts can be adjusted for risks, taking more cautious approaches for volatile stocks. Lag variables, such as the price and volume from the previous day, add short-term context, helping identify continuity in daily trends, making the model more capable of capturing short-term patterns (Sadhwani et al., 2021).

All of this leads to increased model complexity and more structured information about the past and present behaviour of the stock. A model with more variables can build more complex relationships among the data, detecting patterns that the previous model (with only the initial variables) could not capture. The model has the flexibility to better adapt to market dynamics, identifying subtler relationships and complex temporal patterns that influence the price for the following day. This allows for potentially more accurate predictions. Table 3 shows the MSE values for the respective machine learning models.

Table 3. Mean Squared Error of Models

Model	Mean Squared Error (MSE)
Random Forest	9.471633136543028
XGBoost	13.644980711221393

Source: Authors 'processing.

The lowest MSE value for Random Forest (9.4716) indicates that this model has predicted the closing prices for the next day more accurately compared to XGBoost, which obtained a higher MSE of 13.6449. We use SHAP values to understand the internal workings of the machine learning models.

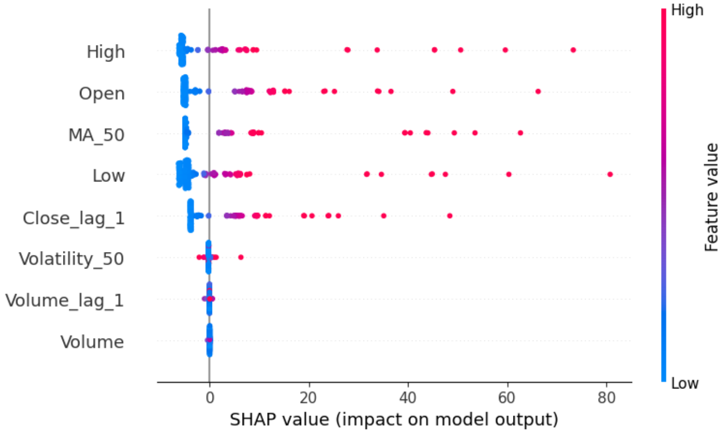

Figure 3. SHAP: Random Forest Model
Source: Authors' processing.

Figures 3 and 4 show how each feature impacts the model's output. For instance, features like "High" and "Open" in Figure 3 have higher SHAP values, indicating that these variables have a significant impact on the predictions. This suggests that variations in these features can greatly influence the predicted outcome.

The dispersion of points in the graphs represents the variability of the impact of each feature. A large dispersion, as seen in "Close_lag_1" and "Volume_lag_1," indicates that their impact on the model can vary significantly depending on specific values. This is particularly interesting when trying to understand risk or opportunity factors in a forecasting context.

Understanding which features have the greatest impact allows for the optimisation of data-driven strategies. For example, if "High" and "Open" are the most influential features for predictions in a financial market, analysts can focus on these factors to improve their operational decisions.

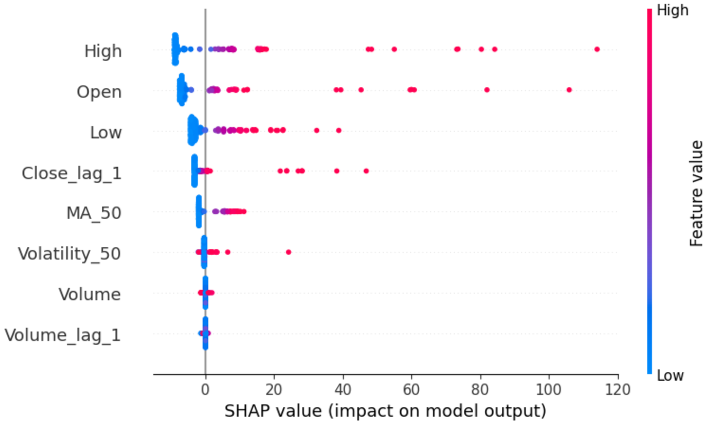

Figure 4. SHAP: XGBoost Model
Source: Authors' processing.

Thus, SHAP values provide a clear view of how each feature influences the predictions of the models, helping to decipher the complexities of machine learning models. This is particularly useful in contexts where decisions need to be based on analyses of complex data. Improved interpretability not only makes models more transparent but also supports the validation of results and trust in the decisions made based on such models.

Tables 4 and 5 provide a detailed view of how the Random Forest and XGBoost models make predictions and which features are most influential.

Table 4. LIME Explanations

Feature	Random Forest (Value)	Random Forest (Contribution)	XGBoost (Value)	XGBoost (Contribution)
High	7.85	9.69	7.85	12.33
MA_50	7.87	8.68	7.87	4.40
Open	7.75	7.80	7.75	9.43
Low	7.69	7.33	7.69	6.38

Feature	Random Forest (Value)	Random Forest (Contribution)	XGBoost (Value)	XGBoost (Contribution)
Close_lag_1	7.75	5.23	7.75	3.49
Volume_lag_1	9164110.00	0.73	9164110.00	0.76
Volatility_50	0.13	0.54	0.13	0.22
Volume	29107003.00	0.02	29107003.00	0.02

Source: Authors' processing.

Table 5. Predictions & Intercepts

Predicted Value	Intercept	Right Value
11.2973	49.8538	7.7691
11.7241	48.2549	7.7517

Source: Authors' processing.

The LIME explanation allows for a more interpretable understanding of the model's decisions, highlighting the importance of each feature for the specific prediction. This approach is crucial to ensure that decisions based on the models are informed and justifiable, especially in critical contexts such as financial or economic forecasting.

5. Conclusions

The conducted study analyses the use of machine learning (ML) and Explainable AI (XAI) techniques for predicting stock prices, focusing on predictive accuracy and the transparency of decisions made by the models. Two ML models were used: Random Forest and XGBoost, aiming to predict the daily closing price of sustainability-related stocks such as Tesla, NextEra Energy, Vestas, and Enel.

The results show that the Random Forest model outperformed XGBoost in terms of accuracy, achieving a lower mean squared error (MSE). To further improve predictions, new derived variables were introduced, such as the 50-day moving average (MA_50), the 50-day volatility (Volatility_50), and "lag" variables like the closing price and volume of the previous day. These new variables enrich the dataset by providing historical context that allows the models to capture more complex market trends and dynamics.

In addition to model construction, the study integrated Explainable AI (XAI) techniques such as SHAP and LIME to explain how each variable affects predictions. SHAP revealed that the "High" and "Open" variables had a significant impact on the predictions of Random Forest, while in the case of XGBoost, "High" remained the dominant variable, but with more weight assigned to the "Open" variable. These results enable a better understanding of the models' functioning, facilitating the decision-making process for investors.

The integration of XAI techniques has made it possible to interpret the behaviour of the models more transparently, thus overcoming one of the main limitations of machine learning, namely the "black box" nature of predictions . This approach proves particularly useful for investors in companies with high ESG ratings, providing a deeper understanding of market dynamics and the variables driving the prices of sustainable stocks. The study concludes that by combining ML and XAI, it is possible not only to improve the accuracy of predictions but also to provide greater transparency in investment decisions.

Acknowledgements: *This work was funded by European Union under the NextGeneration EU Programme within the Plan "PNRR - Missione 4 "Istruzione e Ricerca" - Componente C2 Investimento 1.1 "Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN)" by the Italian Ministry of University and Research (MUR), Project title: "Climate risk and uncertainty: environmental sustainability and asset pricing". Project code "P20225MJW8" (CUP: E53D23016470001), MUR D.D. financing decree n. 1409 of 14/09/2022.*

References

[1] Achituve, I., Kraus, S., Goldberger, J. (2019), *Interpretable online banking fraud detection based on hierarchical attention mechanism*. In 2019 *IEEE 29th international workshop on machine learning for signal processing* (MLSP), IEEE, 1-6.

[2] Adadi, A., Berrada, M. (2018), *Peeking inside the black-box: a survey on explainable artificial intelligence* (XAI). *IEEE access*, 6, 52138-52160.

[3] Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J. et al. (2020), *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*. *Information fusion*, 58, 82-115.

[4] Biran, O., Cotton, C. (2017), *Explanation and justification in machine learning: A survey.* In *IJCAI-17 workshop on explainable AI* (XAI), 8(1), 8-13.

[5] Cao, N. (2021), *Explainable artificial intelligence for customer churning prediction in banking*. In *Proceedings of the 2nd International Conference on Human-Centered Artificial Intelligence* (Computing4Human 2021), Danang, Vietnam, 28-29.

[6] Chen, T., Guestrin, C. (2016), *Xgboost: A scalable tree boosting system*. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, San Francisco California (USA), August 13-17, 2016, 785-794.

[7] Doshi-Velez, F., Kim, B. (2017), *Towards a rigorous science of interpretable machine learning*, arXiv preprint arXiv:1702.08608.

[8] Ertel, W. (2018), *Introduction to Artificial Intelligence*. Springer - Berlin, 1-356.

[9] Fama, E.F., French, K.R. (1996), *Multifactor explanations of asset pricing anomalies*. *The Journal of Finance*, 51(1), 55-84.

[10] Fama, E.F., French, K.R. (1993), *Common risk factors in the returns on stocks and bonds*. *Journal of financial economics*, 33(1), 3-56.

[11] Giglio, S., Xiu, D. (2021), *Asset pricing with omitted factors*. *Journal of Political Economy*, 129(7), 1947-1990.

[12] Gu, S., Kelly, B., Xiu, D. (2020), *Empirical asset pricing via machine learning*. *The Review of Financial Studies*, 33(5), 2223-2273.

[13] Gu, S., Kelly, B.T., Xiu, D. (2021), *Autoencoder asset pricing models. Journal of Econometrics*, 222(1), Part B, 429-450.

[14] Heaton, J.B., Polson, N.G., Witte, J.H. (2016), *Deep learning in finance*, arXiv preprint arXiv:1602.06561.

[15] Khandani, A.E., Kim, A.J., Lo, A.W. (2010), *Consumer credit-risk models via machine-learning algorithms*. *Journal of Banking & Finance*, 34(11), 2767-2787.

[16] Linardatos, P., Papastefanopoulos, V., Kotsiantis, S. (2020), *Explainable AI: A review of machine learning interpretability methods*. *Entropy*, 23(1), 18.

[17] Ribeiro, M.T., Singh, S., Guestrin, C. (2016), *"Why should I trust you?" Explaining the predictions of any classifier*. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, San Francisco California (USA), August 13-17, 2016, 1135-1144.

[18] Sadhwani, A., Giesecke, K., Sirignano, J. (2021), *Deep learning for mortgage risk*. *Journal of Financial Econometrics*, 19(2), 313-368.

[19] Saeed, W., Omlin, C. (2023), *Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities*. *Knowledge-Based Systems*, 263, 110273.

[20] Sharpe, W.F. (1964), *Capital asset prices, A theory of market equilibrium under conditions of risk. The Journal of Finance*, 19(3), 425-442.

[21] Tsang, K.H., Wong, H.Y. (2019), *Deep-learning solution to portfolio selection with serially-dependent returns*. *SIAM Journal on Financial Mathematics*, 11(2), 593-619.

[22] Yang, G., Ye, Q., Xia, J. (2022), *Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond*. *Information Fusion*, 77, 29-52.

[23] Yao, J., Li, Y., Tan, C.L. (2000), *Option price forecasting using neural networks*. *Omega*, 28(4), 455-466.

[24] Ye, J., Goswami, B., Gu, J., Uddin, A., Wang, G. (2024), *From Factor Models to Deep Learning: Machine Learning in Reshaping Empirical Asset Pricing*, arXiv preprint arXiv:2403.06779.