*Journal of Statistics & Management Systems*
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 29 (2026), No. 10, pp. 1151–1160
DOI : 10.47974/JSMS-1754

# AI-augmented ESG investing : Measuring and predicting sustainability impact in capital markets

Mohd Shahid Ali [§]
*School of Management*
*IILM University*
*Gurugram 122003*
*Haryana*
*India*

Ritika Bajaj Waghray [†]
*Department of Finance*
*Vignana Jyothi Institute of Management*
*Hyderabad 500118*
*Telangana*
*India*

Afzalur Rahman *
*School of Business*
*Woxsen University*
*Hyderabad 502345*
*Telangana*
*India*

M. A. Honest Mohideen [‡]
*Department of Commerce*
*Motilal Nehru College*
*University of Delhi*
*Benito Juarez Marg*
*South Campus*
*New Delhi 110021*
*India*

[§] *ORCID-ID:* **0009-0004-9749-8366**
[†] *ORCID-ID:* **0000-0001-6897-5654**
* *ORCID-ID:* **0000-0002-1561-561X**
[‡] *ORCID-ID:* **0009-0002-9660-1419**

[§] *E-mail:* **mohd.shahid@iilm.edu**
[†] *E-mail:* **Dr.ritikawaghray@vjim.edu.in**
* *E-mail:* **afzalur.rahman@woxsen.edu.in (Corresponding Author)**
[‡] *E-mail:* **honest.mohideen@gmail.com**

© Taru Publications

Refia Wiquar [@]
*Sharda School of Business Studies*
*Sharda University*
*Greater Noida 201310*
*Uttar Pradesh*
*India*

Mohsina Hayat [#]
*School of Commerce*
*Jain (Deemed to be University)*
*Bengaluru 560069*
*Karnataka*
*India*

**Abstract**

This paper examines whether artificial intelligence (AI) improves the information that environmental, social and governance (ESG) ratings provide to investors. Sustainalytics ESG risk ratings, market indicators and natural language processing (NLP) features from business descriptions are combined for 429 S&P 500 firms (81,843 firm-months, 2010–2026). Random forest, XGBoost and LSTM models tested out of sample (January 2024 – September 2026) identify firms entering high-volatility months far more accurately than ESG ratings alone (AUC 0.82–0.84 versus 0.50; $p < 0.001$). However, no model predicts returns better than the ratings-only benchmark, and AI portfolio Sharpe ratios do not differ significantly from that of a best-in-class ESG portfolio (0.81 versus 0.84). AI therefore adds most value as a forward-looking risk-monitoring tool rather than as a source of excess returns.

***Subject Classification:*** *62J05, 62H30, 68T07, 91B76, 91G15, 62M45, 91G10.*

***Keywords:*** *Artificial intelligence, ESG investing, Machine learning, Portfolio performance, Sustainability impact, Capital markets.*

## 1. Introduction

ESG (environmental, social and governance) investing has entered the mainstream of capital markets, and ESG considerations are now primary drivers of investment decisions [1]. Yet ESG ratings rest on subjective assessments and lagged disclosures, rating agencies diverge so

[@] *ORCID-ID:* **0009-0006-1801-6547**
[#] *ORCID-ID:* **0000-0003-1274-8713**

[@] *E-mail:* **refia.wiquar@sharda.ac.in**
[#] *E-mail:* **mohsina.hayat@jainuniversity.ac.in**

much that the same firm can receive conflicting scores [2], and greenwashing undermines the reliability of the data [3]. Studies linking ESG scores to portfolio performance find that firms with better sustainability records have lower risk and greater resilience in downturns [4], but this evidence is largely retrospective [5], and few studies use machine learning to predict sustainability-related outcomes [6], although it has proved effective in asset pricing [7]. This study makes two contributions. First, it introduces a novel AI-based ESG framework that produces dynamic, forward-looking and contextually relevant signals by integrating ESG ratings, market data and text. Second, it compares this framework with static ratings on real capital-market data, showing where AI-based ESG signals improve decision-making and risk management, and where they do not.

## 2. Literature review

ESG-related assets under management have reached historic levels, and firms with better ESG performance tend to show lower volatility and more long-term value [8]; institutional investors are increasingly considering ESG issues as part of their fiduciary duty and responsible investing [9, 10]. However, ESG ratings disagree substantially [11] and rely on proprietary methods that are rarely disclosed [12]. Machine-learning models such as gradient-boosted trees [13] scale across firms and can be updated continuously, and cross-checking multiple independent sources can reduce bias and greenwashing [14]. Yet comparative evidence on AI's contribution to ESG-guided investment remains scarce, and such gaps necessitate frameworks that test it systematically.

## 3. Methodology

This study combines ESG risk ratings, market data and company business descriptions. ESG data are the Sustainalytics ESG Risk Ratings (total, environmental, social, governance and controversy scores) of S&P 500 firms compiled by Dugar [15]; higher scores indicate greater unmanaged risk. Adjusted daily prices and volumes come from an open S&P 500 price repository [16], and the risk-free rate is the 3-month Treasury-bill rate [17]. The sample covers N = 429 firms in 11 sectors and 81,843 firm-months (features June 2010 – August 2026; targets one month ahead). It is split chronologically into training (to November 2022; 62,719 observations), validation (December 2022 – November 2023; 5,136) and

test sets (returns January 2024 – September 2026; 13,988 observations in 33 months). Because the ratings are a single 2023 snapshot, the test period begins after their publication, so there is no look-ahead in the ESG data.

Features comprise (i) the ESG scores; (ii) market variables – momentum (1, 3, 6 and 12–2 months), current and 12-month volatility, beta, maximum daily return, dollar volume, Amihud illiquidity, price and sector dummies; and (iii) NLP features – ten latent semantic topics from TF-IDF vectors of business descriptions and the shares of 'brown' and 'green' lexicon terms. Features are ranked cross-sectionally to [−0.5, 0.5] [7]. Targets for month t+1 are the return, its direction (above the median) and high volatility (top tercile). The traditional benchmark is a linear or logistic regression on ESG features only. The AI-augmented models are Random Forest [18], XGBoost [13] and an LSTM [19] that combines 12-month market sequences with the static features; a market-only XGBoost is an ablation. Hyperparameters (Random Forest depth {6, 10}, leaf size {100, 400}; XGBoost depth {3, 5}, learning rate {0.03, 0.1}; LSTM units {32, 64}; early stopping) were tuned on the validation year before refitting. Accuracy is measured by RMSE, MAE, out-of-sample $R^2$ and AUC, with Diebold and Mariano [20] and Newey and West [21] tests; equally weighted top-quintile portfolios are compared with the Memmel test [22]. Data and Python code are supplied as supplementary material.

## 4. Results and analysis

a. *Descriptive Statistics*

Table 1 reports descriptive statistics. The mean total ESG risk score is 21.5 (SD 6.9), with sector means from 13.1 (real estate) to 32.3 (energy); these differences motivate the sector controls. Monthly returns average 1.37% (SD 8.24%).

**Table 1**

**Descriptive Statistics of Key Variables**

| Variable | N | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Total ESG risk score | 429 | 21.54 | 6.90 | 7.10 | 41.70 |
| Environmental risk score | 429 | 5.75 | 5.09 | 0.00 | 25.00 |
| Social risk score | 429 | 9.07 | 3.66 | 0.80 | 22.50 |
| Governance risk score | 429 | 6.72 | 2.21 | 3.00 | 19.40 |
| Controversy score (0–5) | 402 | 2.01 | 0.79 | 1.00 | 5.00 |

*Contd...*

| | | | | | |
|---|---|---|---|---|---|
| Green-lexicon intensity (%) | 429 | 0.33 | 0.87 | 0.00 | 8.41 |
| Brown-lexicon intensity (%) | 429 | 0.98 | 2.67 | 0.00 | 18.95 |
| Monthly return (%) | 81,843 | 1.37 | 8.24 | –83.23 | 213.83 |
| Monthly volatility, annualised (%) | 81,843 | 26.62 | 15.43 | 1.63 | 355.95 |
| Momentum t–12 to t–2 (%) | 81,843 | 15.72 | 31.79 | –86.92 | 959.18 |
| Market beta (24 m) | 81,843 | 0.998 | 0.546 | –1.614 | 7.083 |
| Log dollar volume | 81,843 | 19.08 | 1.06 | 14.83 | 24.82 |

*Note:* ESG and lexicon variables are firm-level (controversy score: 402 firms); market variables are firm-months, June 2010 – August 2026. Sources: [15, 16].

b. *Predictive Performance of AI Models*

Table 2 and Figure 1 report out-of-sample accuracy. Next-month returns are almost unpredictable: RMSE ranges from 0.0828 to 0.0836, $R^2_{oos}$ from –0.7% to 1.1%, the LSTM matches the ratings-only benchmark (Diebold and Mariano p = 0.87), the tree models are slightly less accurate ($p \leq 0.012$), and direction AUCs are 0.49–0.50. In contrast, the AUC for

**Table 2**

**Out-of-Sample Predictive Performance (January 2024 – September 2026)**

| Model | RMSE | MAE | $R^2_{oos}$ (%) | AUC direction | AUC high vol. | DM stat. (p) | ΔAUC high vol. (t) |
|---|---|---|---|---|---|---|---|
| Traditional ESG (linear/ logit) | 0.0828 | 0.0632 | 1.12 | 0.497 | 0.498 | – | – |
| Random Forest (AI-ESG) | 0.0832 | 0.0636 | 0.09 | 0.491 | 0.817 | –2.65 (0.012) | +0.319 (26.8) |
| XGBoost (AI-ESG) | 0.0836 | 0.0639 | –0.75 | 0.490 | 0.831 | –3.30 (0.002) | +0.333 (37.3) |
| LSTM (AI-ESG) | 0.0828 | 0.0632 | 1.08 | 0.500 | 0.839 | –0.16 (0.874) | +0.341 (31.6) |
| XGBoost (market only) | 0.0829 | 0.0632 | 0.98 | 0.501 | 0.826 | –0.87 (0.389) | +0.328 (30.6) |

*Note:* 13,988 firm-months, 33 test months. RMSE, MAE, $R^2_{oos}$: next-month returns; AUC direction: return above the median; AUC high vol.: top-tercile volatility. DM: Diebold and Mariano statistic vs the traditional model (negative = less accurate). ΔAUC: mean monthly gain, Newey–West t.

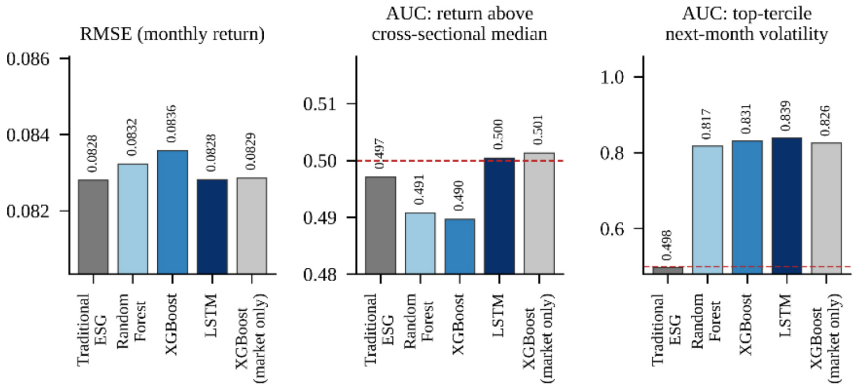



**Figure 1**

**Out-of-sample model performance, January 2024 – September 2026 (dashed line: random classifier)**

next-month high volatility rises from 0.498 for the benchmark to 0.817–0.839 for the AI models ($t \geq 26.8$, $p < 0.001$). Most of this gain comes from market data (market-only XGBoost: 0.826), but ESG and textual features add significantly for XGBoost (+0.005, $p = 0.017$) and the LSTM (+0.013, $p < 0.001$).

**Table 3**

**Back-Tested Portfolio Performance (January 2024 – September 2026)**

| Portfolio | Ann. return (%) | Ann. vol. (%) | Sharpe ratio | Max draw-down (%) | Turn-over (%) | Sharpe test vs Trad. z (p) |
|---|---|---|---|---|---|---|
| Equal-weighted universe | 12.50 | 11.32 | 0.72 | –8.58 | 0.0 | –0.59 (0.55) |
| Traditional ESG (best-in-class) | 15.03 | 12.50 | 0.84 | –8.89 | 0.0 | – |
| AI-ESG: Random Forest | 10.72 | 13.46 | 0.51 | –12.90 | 28.9 | –1.19 (0.23) |
| AI-ESG: XGBoost | 11.93 | 14.08 | 0.57 | –13.47 | 32.6 | –1.01 (0.31) |
| AI-ESG: LSTM | 16.43 | 14.87 | 0.81 | –12.22 | 16.1 | –0.12 (0.90) |
| AI-augmented ESG composite | 7.70 | 11.87 | 0.32 | –7.10 | 31.5 | –1.35 (0.18) |

*Note:* No transaction costs; risk-free rate from [17]. Sharpe test: Memmel [22] vs the traditional ESG portfolio. Sortino ratios and CVaR are in the supplementary file.

c. *Portfolio Simulation*

Table 3 and Figure 2 report the back-tests (33 monthly rebalances, equally weighted top quintiles). The best-in-class ESG portfolio earned 15.0% a year (Sharpe ratio 0.84), ahead of the universe (12.5%; 0.72). The LSTM portfolio earned 16.4% with higher volatility, so its Sharpe ratio (0.81) is indistinguishable from the benchmark ($p = 0.90$); the Random Forest and XGBoost portfolios underperformed (0.51 and 0.57). The composite portfolio, which favours high predicted return and low predicted volatility, had the smallest drawdown (–7.1% versus –8.9%) but the lowest return (7.7%). No Sharpe difference is significant, and AI portfolios required 16–33% monthly turnover.

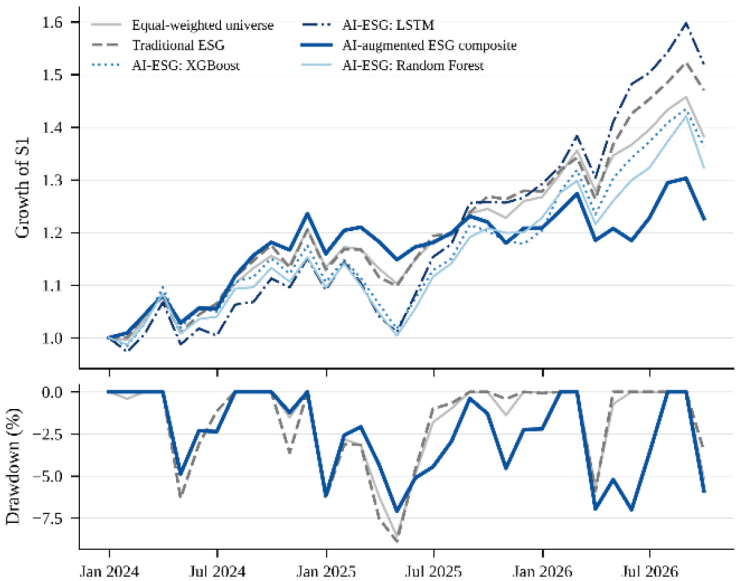



**Figure 2**

**Back-tested cumulative performance (top) and drawdowns (bottom), January 2024 – September 2026.**

d. *Case Study Analysis*

The case study compares 20 fossil-fuel firms with 12 renewable- and clean-energy firms. Ratings separate the groups (mean ESG risk 32.3 versus 28.0; Figure 3 a) but cannot indicate when risk materialises; fossil-fuel firms spent 38.8% of firm-months in the top volatility tercile, against

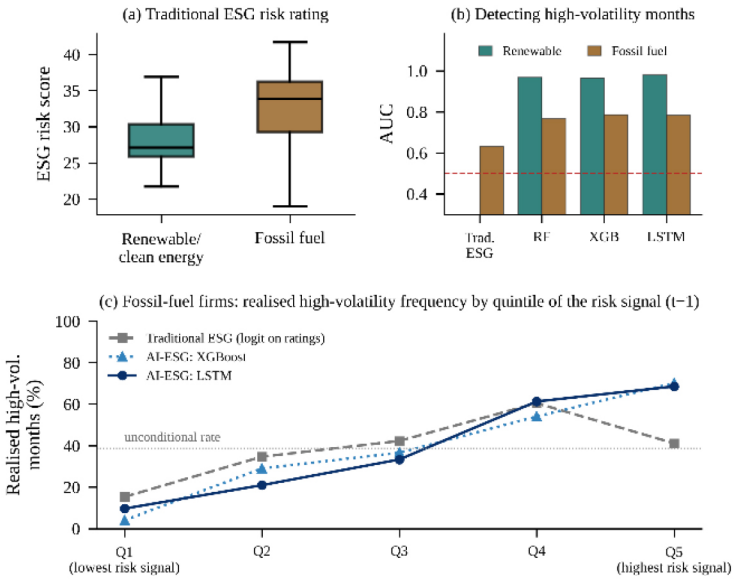



**Figure 3**

**Case study – renewable vs fossil-fuel firms: (a) ESG risk scores, (b) AUC for high-volatility months, (c) high-volatility frequency by risk-signal quintile for fossil-fuel firms**

25.0%. For fossil-fuel firms, the AI models identified high-volatility months better (AUC 0.77–0.79) than the ratings (0.63) (Figure 3 b): 68.5% of months after the top LSTM-signal quintile were high-volatility months, versus 9.7% after the bottom quintile (ratings: 41.1% and 15.3%; Figure 3 c). For renewable firms the AI AUCs were 0.97–0.98, whereas the ratings ranked firms in the wrong direction (AUC 0.29). The small groups call for caution.

## 5. Discussion

AI improves the forward-looking risk information extracted from ESG data but not the prediction of short-horizon returns. Combining ratings with market and NLP features raised the high-volatility AUC from 0.50 to 0.84, turning static ratings into time-varying risk measures for portfolio monitoring and regulatory oversight [8, 14]. The absence of return predictability and the indistinguishable Sharpe ratios caution against claims of AI-driven excess returns, consistent with the mixed evidence on ESG and performance [5, 11]. Limitations are the U.S. large-

cap sample, the single rating snapshot, the use of business descriptions instead of news, the 33-month test period and the limited interpretability of the LSTM.

## 6. Conclusion

AI enhances ESG investing chiefly by improving the prediction of sustainability-related risk rather than returns or Sharpe ratios. Future research should use time-varying ratings from several providers, news and sustainability-report text, samples that include emerging markets, and explainable AI to improve transparency.

## References

[1] A. Amel-Zadeh and G. Serafeim, "Why and how investors use ESG information: Evidence from a global survey," *Financ. Anal. J.*, vol. 74, no. 3, pp. 87–103 (2018).

[2] F. Berg, J. F. Kölbel, and R. Rigobon, "Aggregate confusion: The divergence of ESG ratings," *Rev. Finance*, vol. 26, no. 6, pp. 1315–1344 (2022).

[3] M. A. Delmas and V. C. Burbano, "The drivers of greenwashing," *Calif. Manage. Rev.*, vol. 54, no. 1, pp. 64–87 (2011).

[4] K. V. Lins, H. Servaes, and A. Tamayo, "Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis," *J. Finance,* vol. 72, no. 4, pp. 1785–1824 (2017).

[5] A. Banu, A. Rahman, and M. Murari, "A study on financial performance of Nifty 50 Indian companies with strong corporate governance," *Journal of Information and Optimization Sciences*, vol. 46, no. 8, pp. 2403–2412 (2025), doi: 10.47974/JIOS-2055.

[6] V. D'Amato, R. D'Ecclesia, and S. Levantesi, "ESG score prediction through random forest algorithm," *Comput. Manag. Sci.*, vol. 19, pp. 347–373 (2022), doi: 10.1007/s10287-021-00419-3.

[7] S. Gu, B. Kelly, and D. Xiu, "Empirical asset pricing via machine learning," *Rev. Financ. Stud.*, vol. 33, no. 5, pp. 2223–2273 (2020).

[8] R. G. Eccles, I. Ioannou, and G. Serafeim, "The impact of corporate sustainability on organizational processes and performance," *Manage. Sci.*, vol. 60, no. 11, pp. 2835–2857 (2014).

[9] P. Krueger, Z. Sautner, and L. T. Starks, "The importance of climate risks for institutional investors," *Rev. Financ. Stud.*, vol. 33, no. 3, pp. 1067–1111 (2020).

[10] F. Rizwan and A. M. A. Jenita, "An analysis of impact of CSR practices on investment behaviour," *J. Stat. Manag. Syst.,* vol. 25, no. 5, pp. 1011–1020 (2022), doi: 10.1080/09720510.2022.2083828.

[11] R. Gibson Brandon, P. Krueger, and P. S. Schmidt, "ESG rating disagreement and stock returns," *Financ. Anal. J.,* vol. 77, no. 4, pp. 104–127 (2021).

[12] D. M. Christensen, G. Serafeim, and A. Sikochi, "Why is corporate virtue in the eye of the beholder? The case of ESG ratings," *Account. Rev.,* vol. 97, no. 1, pp. 147–175 (2022).

[13] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining (KDD)*, pp. 785–794 (2016).

[14] S. Y. In, D. Rook, and A. Monk, "Integrating alternative data (also known as ESG data) in investment decision making," *Glob. Econ. Rev.,* vol. 48, no. 3, pp. 237–260 (2019).

[15] P. Dugar, "S&P 500 ESG risk ratings" [Data set], Kaggle (2023), doi: 10.34740/KAGGLE/DS/3660201.

[16] Johnbrick123, "sp500-data: S&P 500 daily price history," *GitHub repository,* (2026).

[17] Board of Governors of the Federal Reserve System, "3-month Treasury bill secondary market rate, discount basis (TB3MS)," *FRED, Federal Reserve Bank of St. Louis,* (2026).

[18] L. Breiman, "Random forests," *Mach. Learn.,* vol. 45, no. 1, pp. 5–32 (2001), doi: 10.1023/A:1010933404324.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.,* vol. 9, no. 8, pp. 1735–1780 (1997), doi: 10.1162/neco.1997.9.8.1735.

[20] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Stat.,* vol. 13, no. 3, pp. 253–263 (1995), doi: 10.1080/07350015.1995.10524599.

[21] W. K. Newey and K. D. West, "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica,* vol. 55, no. 3, pp. 703–708 (1987), doi: 10.2307/1913610.

[22] C. Memmel, "Performance hypothesis testing with the Sharpe ratio," *Finance Lett.,* vol. 1, no. 1, pp. 21–23 (2003).

*Received December, 2025*