



Nghiên cứu ứng dụng mô hình học máy khả diễn XAI trong phân tích rủi ro đầu tư ESG: Thông tin chi tiết từ các doanh nghiệp S&P 500

HOÀNG ANH ^{*},^a, TRƯƠNG CHÚC THANH ^a, LÊ THỊ TUYẾT MAI ^b

^a Đại học Kinh tế Thành phố Hồ Chí Minh

^b Trường Đại học Khoa học tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh

THÔNG TIN	TÓM TẮT
Ngày nhận: 02/02/2025 Ngày nhận lại: 11/03/2025 Duyệt đăng: 12/03/2025 Mã phân loại JEL: C52; F18; O3; M14; Q56; Y10. Từ khóa: Phân tích rủi ro ESG; XAI; SHAP; Đầu tư ESG; Đầu tư bền vững. Keywords: ESG risk analysis; XAI; SHAP;	Các yếu tố môi trường, xã hội, và quản trị trở nên quan trọng trong việc đánh giá hiệu suất và rủi ro của doanh nghiệp. Nghiên cứu này khám phá ứng dụng các mô hình học máy khả diễn XAI nhằm cải thiện khả năng diễn giải, tăng độ tin cậy trong phân tích rủi ro ESG. Thực nghiệm dữ liệu ESG của các doanh nghiệp S&P 500 cho thấy, mô hình LightGBM có độ chính xác cao nhất so với XGBoost và Random Forest. Phân tích giá trị SHAP chỉ ra rằng rủi ro ESG bị chi phối chủ yếu bởi ba yếu tố chính: rủi ro môi trường, rủi ro xã hội, và rủi ro quản trị. Ngoài ra, mức độ rủi ro ESG tổng thể của doanh nghiệp cũng đóng vai trò quan trọng. Kết quả nghiên cứu làm nổi bật tiềm năng của các mô hình XAI trong việc tăng cường báo cáo và tuân thủ ESG, đồng thời cung cấp thông tin chi tiết cho các nhà đầu tư và nhà hoạch định chính sách. Hơn nữa, nghiên cứu này minh chứng cho việc tích hợp ML/AI khả diễn vào quy trình quản lý rủi ro của tổ chức, thúc đẩy tính minh bạch, trách nhiệm giải trình, và sự tin tưởng vào các đánh giá ESG. Abstract Environmental, Social, and Governance (ESG) factors have become crucial in assessing corporate performance and risks. This study examines the application of explainable artificial intelligence (XAI) to enhance interpretability and improve trustworthiness in ESG risk analysis. By experimenting with ESG data from S&P 500 companies, this study shows that the LightGBM model achieves the highest accuracy,

^{*} Tác giả liên hệ.

Biên tập viên: Trần Lê Thùy Duyên.

Email: anhh@ueh.edu.vn (Hoàng Anh); thanhtruong.522202180068@st.ueh.edu.vn (Trương Chúc Thanh); lttmai@hcmus.edu.vn (Lê Thị Tuyết Mai).

Trích dẫn bài viết: Hoàng Anh, Trương Chúc Thanh, & Lê Thị Tuyết Mai. (2025). Nghiên cứu ứng dụng mô hình học máy khả diễn XAI trong phân tích rủi ro đầu tư ESG: Thông tin chi tiết từ các doanh nghiệp S&P 500. *Tạp chí Nghiên cứu Kinh tế và Kinh doanh Châu Á*, 36(1), 20-34.

ESG investment; Sustainable investment.	outperforming XGBoost and Random Forest. SHAP value analysis indicates that ESG risks are primarily driven by three key factors: environmental risk, social risk, and governance risk. Additionally, the overall ESG risk level of the company also plays an important role. The findings highlight the potential of XAI models to enhance ESG reporting and compliance while providing valuable insights for investors and policymakers. Furthermore, this study demonstrates the viability of integrating explainable ML/AI approaches into organizational risk management, thereby fostering transparency, accountability, and trust in ESG assessments.
--	---

1. Giới thiệu

Xu hướng áp dụng các yếu tố ESG (Environmental - Social - Governance – môi trường, xã hội và quản trị doanh nghiệp) ngày càng trở nên quan trọng trong phân tích doanh nghiệp và hỗ trợ ra quyết định đầu tư (Bùi Minh Tiến và cộng sự, 2024; Hồ Thị Hải Ly, 2024; Nguyễn Thị Liên Hương, 2023; Nguyễn Minh Phương và cộng sự, 2024; Zhang & Zhang, 2024). Vì vậy, hiểu được rủi ro và cơ hội ESG là cần thiết đối với các doanh nghiệp muốn tạo ra giá trị bền vững. Tuy nhiên, các nghiên cứu này đều sử dụng các phương pháp thống kê, kiểm định giả thuyết, và phân tích định lượng truyền thống. Nghiên cứu này xây dựng hệ thống kiến thức và công cụ cần thiết để hiểu, áp dụng, và giải thích các mô hình học máy khả diễn XAI trong việc đánh giá rủi ro ESG. Khai thác dữ liệu ESG từ các doanh nghiệp S&P 500 (Ernst & Woithe, 2024), nghiên cứu hướng đến việc đánh giá các yếu tố ESG có thể tác động đến hoạt động kinh doanh và hiệu suất tài chính. Ngoài ra, nghiên cứu tăng cường khả năng giải thích quá trình huấn luyện mô hình và diễn giải kết quả đạt được từ các mô hình thực nghiệm.

Cách tiếp cận báo cáo ESG toàn diện xem xét hiệu suất của một doanh nghiệp không chỉ dựa trên các số liệu tài chính truyền thống mà còn dựa trên tác động của doanh nghiệp đó đối với môi trường, mối quan hệ với các bên liên quan và chất lượng quản lý, giám sát của doanh nghiệp đó (Assael và cộng sự, 2023; Ernst & Woithe, 2024; Initiative, 2021). Trong đó, các yếu tố môi trường xem xét việc sử dụng năng lượng, quản lý chất thải, lượng khí thải carbon và các sáng kiến khác của doanh nghiệp với chiến lược phát triển bền vững. Các yếu tố xã hội xem xét cách đối xử của doanh nghiệp với nhân viên, mối quan hệ của doanh nghiệp với cộng đồng địa phương và các chính sách của doanh nghiệp về tính đa dạng, công bằng và hòa nhập. Các yếu tố quản trị đánh giá khả năng lãnh đạo, cơ cấu hội đồng quản trị, tính minh bạch và việc tuân thủ các tiêu chuẩn đạo đức của doanh nghiệp.

Bằng cách phân tích đánh giá các yếu tố ESG, các nhà đầu tư, cơ quan quản lý và công chúng có thể hiểu rõ hơn về tính bền vững và khả năng phát triển lâu dài của doanh nghiệp. Hiệu suất ESG mạnh mẽ không chỉ dẫn đến kết quả tài chính tốt hơn mà còn góp phần vào lợi ích lớn hơn của xã hội và cộng đồng. Khi các cân nhắc về ESG trở nên phổ biến, các doanh nghiệp chủ động giải quyết các vấn đề này sẽ góp phần gia tăng lợi thế cạnh tranh.

Tuy nhiên, các phương pháp đánh giá rủi ro ESG truyền thống tồn tại những hạn chế như đánh giá rủi ro ESG dựa trên sự kết hợp giữa các phương pháp định tính và định lượng (Nguyễn Dương Hoa, 2024; Jain và cộng sự, 2023). Doanh nghiệp thường sử dụng bảng câu hỏi, khảo sát và phỏng

vấn chuyên sâu để thu thập thông tin về hoạt động và hiệu suất ESG. Họ cũng đã phân tích dữ liệu công khai (chẳng hạn như báo cáo phát triển bền vững của công ty và hồ sơ nộp theo quy định) để đánh giá các rủi ro và cơ hội liên quan đến ESG. Mặc dù các phương pháp tiếp cận truyền thống này đã cung cấp những hiểu biết có giá trị, nhưng chúng thường bị hạn chế về khả năng nắm bắt toàn bộ phạm vi và tính phức tạp của việc đánh giá rủi ro ESG. Ví dụ, (1) dữ liệu ESG thiếu chuẩn hóa và các tiêu chuẩn báo cáo khác nhau giữa các ngành và khu vực, gây khó khăn cho việc so sánh và tổng hợp thông tin; (2) đánh giá định tính có thể bị ảnh hưởng bởi quan điểm chủ quan của người đánh giá, dẫn đến khả năng thiên vị và không nhất quán; (3) các phương pháp truyền thống thường tập trung vào một tập hợp hạn chế các yếu tố ESG, bỏ qua các mối quan hệ phụ thuộc quan trọng và rủi ro chưa được ghi nhận; (4) việc dựa vào dữ liệu lịch sử có thể khiến việc dự đoán và ứng phó với các rủi ro ESG đang phát triển nhanh chóng trở nên khó khăn; (5) điểm số hoặc xếp hạng ESG tổng hợp có thể không cung cấp mức độ chi tiết cần thiết để hiểu và giảm thiểu các rủi ro cụ thể.

Để giải quyết những hạn chế này, các tổ chức ngày càng chuyển sang học máy (Machine Learning – ML) và các kỹ thuật phân tích tiên tiến khác để nâng cao khả năng đánh giá rủi ro ESG (Dwivedi và cộng sự, 2023; Tiwari, 2023). Việc giới thiệu các phương pháp tiếp cận dựa trên ML/AI có thể giúp khắc phục những thiếu sót của các phương pháp truyền thống và mở ra những cơ hội mới cho quản lý rủi ro ESG toàn diện hơn dựa trên dữ liệu và thuật toán. Tuy nhiên, khả năng giải thích quá trình huấn luyện mô hình học máy và ý nghĩa của kết quả dự báo vẫn còn hạn chế. Bài báo này trình bày nghiên cứu ứng dụng phương pháp phân tích giá trị Shapley (SHAP) từ lý thuyết trò chơi giúp đánh giá và giải thích vai trò của các đặc trưng tham gia vào quá trình huấn luyện mô hình, đồng thời gia tăng tính thuyết phục của các kết quả dự đoán (Nguyễn Quốc Huy & Từ Lăng Phiêu, 2023).

2. Các nghiên cứu liên quan và cơ sở lý thuyết

2.1. Nghiên cứu ESG tác động đến doanh nghiệp

Zhang và Zhang (2024) phân tích các nghiên cứu đã được công bố, tập trung vào ba lĩnh vực chính: kết nối học máy với việc công bố ESG, tích hợp học máy với sự bất đồng trong xếp hạng ESG, và sử dụng học máy trong các bối cảnh liên quan đến ESG.

Tại Việt Nam, Nguyễn Thị Liên Hương (2023) nghiên cứu tác động của phát triển bền vững doanh nghiệp đến chi phí sử dụng vốn tại các doanh nghiệp niêm yết. Nghiên cứu sử dụng ba mô hình hồi quy chính (OLS, REM, FEM) để kết luận rằng phát triển bền vững doanh nghiệp giúp giảm chi phí sử dụng vốn chủ sở hữu và vốn bình quân, nhưng không ảnh hưởng đến chi phí sử dụng nợ. Bên cạnh đó, Trần Ngọc Hùng (2023) sử dụng dữ liệu thứ cấp thông qua phương pháp nghiên cứu định tính để phân tích thực trạng công bố chỉ tiêu ESG của các doanh nghiệp niêm yết tại Việt Nam trong bối cảnh đại dịch COVID-19. Nghiên cứu này một lần nữa nhấn mạnh tầm quan trọng của ESG trong việc thu hút đầu tư và đề xuất giải pháp nhằm nâng cao chất lượng công bố ESG.

2.2. Ứng dụng học máy trong nghiên cứu ESG

Veeramani và cộng sự (2024) ứng dụng học máy để xác định loại tác động ESG từ dữ liệu văn bản, phân loại thông tin thành các nhóm như "Cơ hội", "Rủi ro", và "Không thể phân biệt". Nghiên cứu này sử dụng các kỹ thuật học tập tổng hợp và kết hợp nhiều mô hình ngôn ngữ như mBERT,

FlauBERT, ALBERT cùng với các đặc trưng như LSA và TF-IDF nhằm hỗ trợ quyết định có trách nhiệm và bền vững trong quản trị doanh nghiệp và tài chính.

Tiwari (2023) đã khảo sát ứng dụng của Explainable AI (XAI) trong chăm sóc sức khỏe, tài chính, xe tự hành, và các lĩnh vực khác. Nghiên cứu này nhấn mạnh việc sử dụng các kỹ thuật XAI như phân tích tầm quan trọng của đặc trưng và khả năng diễn giải mô hình, từ đó tạo ra sự minh bạch và niềm tin từ phía người dùng đối với quyết định của hệ thống AI.

Ngoài ra, Raza và cộng sự (2022) áp dụng các thuật toán học máy như ANN, KNN, SVM và Random Forest để dự đoán điểm số ESG từ dữ liệu bảng cân đối kế toán và báo cáo thu nhập của các công ty phi tài chính tại Mỹ, Anh và Đức. Cùng với đó, Dwivedi và cộng sự (2023) đã sử dụng Gradient Boosting, Neural Network, Forest, Decision Tree và Linear Regression để xây dựng mô hình dự đoán điểm ESG cho các công ty dựa trên dữ liệu tài chính và phi tài chính từ S&P Global và Moody's Orbis.

Ở một nghiên cứu khác đáng chú ý từ Assael và cộng sự (2023), học máy khả diễn sử dụng để điều tra mối liên hệ giữa lợi nhuận giá và điểm ESG trên thị trường chứng khoán châu Âu. Nghiên cứu này sử dụng mô hình Gradient Boosting, cụ thể là LightGBM, và phương pháp SHAP để giải thích sự đóng góp của từng đặc điểm ESG vào dự đoán.

2.3. Ứng dụng SHAP để phân tích rủi ro ESG

Để áp dụng SHAP (Shapley, 1953) vào mô hình rủi ro ESG, bước đầu tiên là tính toán các giá trị SHAP cho từng đặc trưng trong tập dữ liệu. Các giá trị SHAP thể hiện sự đóng góp của từng đặc trưng vào đầu ra của mô hình, định lượng tác động của đặc trưng đó đối với dự đoán cuối cùng. Thông tin này có thể được sử dụng để xác định các động lực quan trọng nhất của rủi ro ESG, cho phép các bên liên quan tập trung sự chú ý của họ vào các yếu tố chính ảnh hưởng đến quyết định của mô hình.

Ngoài tầm quan trọng của đặc trưng, SHAP cũng có thể tiết lộ các tương tác giữa các đặc trưng đầu vào, làm sáng tỏ cách các đặc điểm khác nhau của một công ty hoặc tài sản phối hợp với nhau để ảnh hưởng đến rủi ro ESG. Bằng cách phân tích các biểu đồ tương tác SHAP, các nhà phân tích có thể khám phá ra các mối quan hệ phụ thuộc ẩn và hiểu được sự tương tác phức tạp của các yếu tố góp phần vào các kết quả liên quan đến ESG.

Cuối cùng, việc áp dụng SHAP vào các mô hình rủi ro ESG giúp tăng cường khả năng ra quyết định dựa trên dữ liệu, cho phép các tổ chức ưu tiên các lĩnh vực có tác động lớn nhất để giảm thiểu rủi ro và thực hành kinh doanh bền vững. Bằng cách tận dụng kỹ thuật XAI mạnh mẽ này, các công ty có thể đưa ra những lựa chọn sáng suốt, minh bạch và có trách nhiệm hơn, phù hợp với mục tiêu ESG của họ.

3. Phương pháp nghiên cứu

Nghiên cứu này được thực hiện theo quy trình đề xuất trong Hình 1 và chi tiết được trình bày trong các phần tiếp theo.

3.1. Xác định bài toán và thu thập dữ liệu ESG

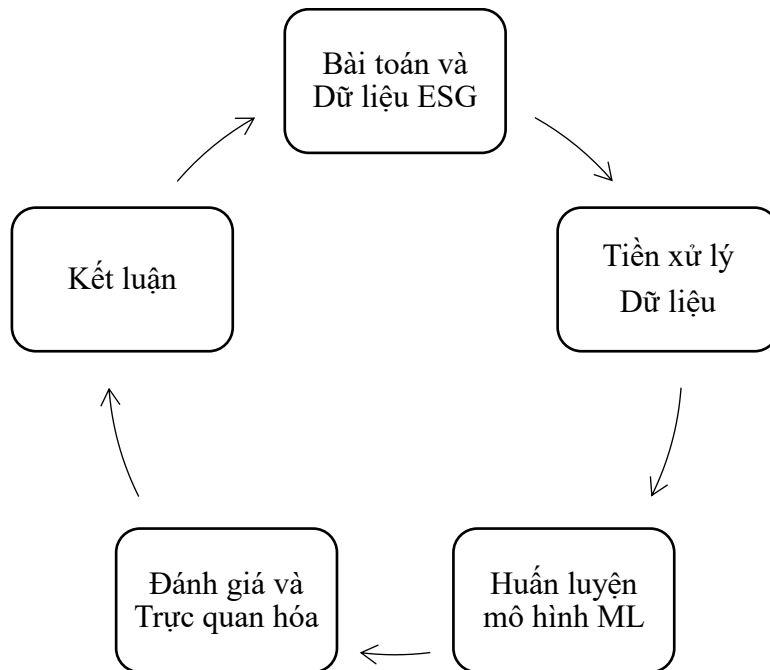
Đây là bước đầu tiên, nơi vấn đề cần giải quyết được xác định và dữ liệu liên quan có thể được thu thập.

- Nguồn dữ liệu ESG: MSCI, Sustainalytics và S&P 500 (Ernst & Woithe, 2024) - Cung cấp dữ liệu ESG toàn diện cho phép phân tích chuyên sâu về hiệu suất môi trường, xã hội và quản trị của các công ty. Bằng cách tận dụng các tập dữ liệu này, các tổ chức có thể đưa ra quyết định chính xác hơn và thúc đẩy thay đổi bền vững.

- Xếp hạng ESG của MSCI: MSCI cung cấp dữ liệu và xếp hạng ESG toàn diện cho hơn 8.500 công ty trên toàn cầu, bao gồm nhiều chỉ số về môi trường, xã hội và quản trị.

- Điểm ESG của Sustainalytics: Sustainalytics cung cấp nghiên cứu ESG chuyên sâu và xếp hạng rủi ro, phân tích mức độ tiếp xúc và quản lý các vấn đề ESG quan trọng của các công ty.

- Điểm ESG của S&P 500: S&P 500 cung cấp điểm và dữ liệu ESG cho hơn 7.000 công ty, cho phép các nhà đầu tư đánh giá hiệu suất và tiến độ phát triển bền vững của các công ty.



Hình 1. Sơ đồ phương pháp nghiên cứu

Các loại dữ liệu ESG có cấu trúc và không có cấu trúc:

- Dữ liệu ESG có cấu trúc: Dữ liệu ESG có cấu trúc bao gồm các số liệu và chỉ số định lượng có thể dễ dàng sắp xếp thành các hàng và cột, chẳng hạn như lượng khí thải nhà kính, mức sử dụng nước, số liệu thống kê về sự đa dạng của nhân viên và thành phần hội đồng quản trị. Loại dữ liệu này rất phù hợp để phân tích thống kê và lập mô hình vì nó có cấu trúc rõ ràng, được xác định trước.

- Dữ liệu ESG không có cấu trúc: Dữ liệu ESG không có cấu trúc gồm nhiều thông tin định tính, bao gồm: báo cáo về tính bền vững của công ty, bài báo, bài đăng trên mạng xã hội và các nguồn văn

bản khác. Dữ liệu này khó phân tích hơn vì không có định dạng chuẩn, đòi hỏi các kỹ thuật như xử lý ngôn ngữ tự nhiên để trích xuất thông tin chi tiết có liên quan.

- Phương pháp tiếp cận kết hợp: Nhiều tổ chức hiện đang áp dụng phương pháp tiếp cận kết hợp, kết hợp các số liệu ESG có cấu trúc với các nguồn dữ liệu không có cấu trúc để hiểu toàn diện hơn về hiệu suất phát triển bền vững của công ty. Điều này cho phép phân tích sâu hơn về bối cảnh và sắc thái đằng sau dữ liệu số, dẫn đến việc ra quyết định sáng suốt hơn.

Mặc dù các nguồn dữ liệu ESG như MSCI, Sustainalytics và S&P 500 cung cấp dữ liệu cho hàng nghìn công ty, nhưng không phải tất cả dữ liệu đó đều công khai hoặc có sẵn để sử dụng trong nghiên cứu học thuật. Việc thu thập dữ liệu ESG đầy đủ thường gặp hạn chế do quyền truy cập, giấy phép sử dụng dữ liệu, chi phí mua dữ liệu hoặc giới hạn của nền tảng cung cấp dữ liệu. Do đó, nghiên cứu này chỉ có thể tiếp cận được một tập hợp con đã được thu thập và chuẩn hóa từ nguồn công khai, tập trung vào thông tin ESG và điểm rủi ro của các công ty thuộc chỉ số S&P 500 chiếm tỷ trọng lớn trong nền kinh tế toàn cầu. Tập dữ liệu này bao gồm các chỉ số định lượng về hiệu suất môi trường, xã hội và quản trị, giúp cung cấp một nền tảng vững chắc để đánh giá mức độ phát triển bền vững của doanh nghiệp.

3.2. Tiền xử lý dữ liệu

Trong bước này, dữ liệu thu thập được làm sạch và chuẩn bị cho việc phân tích. Điều này có thể bao gồm việc xử lý các giá trị thiếu, chuẩn hóa dữ liệu, và chuyển đổi dữ liệu sang định dạng phù hợp.

3.2.1. Xác định và xử lý giá trị thiếu (*Missing Values*)

Loại bỏ tất cả dòng dữ liệu chỉ chứa thông tin định danh và thiếu giá trị ở tất cả các cột: Total ESG Risk score, Environment Risk Score, Governance Risk Score, Social Risk Score, Controversy Level, Controversy Score, ESG Risk Percentile, ESG Risk Level. Giá trị N/A trong cột Controversy Score chuyển thành giá trị 0.

3.2.2. Xác định và xử lý giá trị ngoại lệ (*Outliers*)

Các giá trị cực đại hoặc giá trị ngoại lệ trong dữ liệu ESG có thể ảnh hưởng đáng kể đến hiệu suất của mô hình. Các kỹ thuật như Winsorization, Capping và Robust Scaling có thể được sử dụng để giảm thiểu tác động của các giá trị ngoại lai mà không làm mất đi thông tin có giá trị.

3.2.3. Chuyển đổi dữ liệu

Tỷ lệ các biến số: Các số liệu ESG thường có các thang đo rất khác nhau, có thể ảnh hưởng đến hiệu suất của các mô hình học máy. Các kỹ thuật như chuẩn hóa (điểm z) và tỷ lệ min-max có thể được sử dụng để đảm bảo tất cả các biến đều có thang đo tương tự nhau, cải thiện sự hội tụ và khả năng diễn giải của mô hình.

Mã hóa các biến phân loại: Nhiều chỉ số ESG có bản chất phân loại, chẳng hạn như phân loại ngành hoặc xếp hạng tính bền vững. Mã hóa một lần và mã hóa mục tiêu là các phương pháp phổ biến để chuyển đổi các biến này thành định dạng phù hợp với các thuật toán học máy, bảo toàn các mối quan hệ vốn có trong dữ liệu.

3.3. Huấn luyện mô hình

Dữ liệu đã được tiền xử lý sẽ được sử dụng để huấn luyện các mô hình học máy. Đây là quá trình mà mô hình học từ dữ liệu để có thể đưa ra dự đoán hoặc phân loại. Trong nghiên cứu này, tác giả đã tiến hành huấn luyện mô hình dựa trên tập dữ liệu bao gồm 430 dòng dữ liệu và 146 trường dữ liệu. Trong Machine Learning, nhiều nghiên cứu vẫn có thể đưa ra những phát hiện quan trọng ngay cả khi số lượng dữ liệu không quá lớn, miễn là dữ liệu có chất lượng cao. Việc sử dụng các phương pháp tiên tiến như LightGBM (Ke và cộng sự, 2017), XGBoost (Chen & Guestrin, 2016), và Random Forest (Rigatti, 2017), kết hợp với SHAP Values, giúp tối ưu hóa quá trình khai thác thông tin từ tập dữ liệu nhỏ. Kết quả mô hình vẫn cung cấp những hiểu biết quan trọng về yếu tố ESG ảnh hưởng đến rủi ro ESG, ngay cả khi số lượng dữ liệu còn hạn chế.

Nghiên cứu sử dụng phương pháp K-Fold Cross-Validation để huấn luyện mô hình, dữ liệu được chia thành 4 Folds để đảm bảo mô hình được đánh giá trên toàn bộ tập dữ liệu. Tập dữ liệu huấn luyện có 430 dòng và 146 cột, được chia thành 4 phần bằng nhau, mỗi phần có khoảng 107–108 mẫu. Trong mỗi vòng lặp của quá trình huấn luyện, 3 Folds (~322 mẫu) được sử dụng làm tập huấn luyện, trong khi 1 Fold (~108 mẫu) đóng vai trò là tập kiểm tra. Quá trình này được lặp lại 4 lần, giúp mỗi điểm dữ liệu có cơ hội làm tập kiểm tra một lần và làm tập huấn luyện ba lần. Điều này giúp tận dụng tối đa dữ liệu, đồng thời cung cấp một đánh giá chính xác hơn về hiệu suất của mô hình.

3.4. Đánh giá mô hình, biểu diễn và diễn giải kết quả

Sau khi mô hình được huấn luyện, nó cần được đánh giá để xác định hiệu suất. Kết quả đánh giá sau đó được biểu diễn để phân tích và diễn giải. Bằng cách phân tích các số liệu chính, chúng ta có thể đánh giá khả năng diễn giải, độ chính xác và độ mạnh mẽ của từng mô hình, cho phép chúng ta lựa chọn phương pháp phù hợp nhất.

Các số liệu đánh giá chính:

- Khả năng diễn giải: Hiểu cách mô hình đưa ra quyết định và các yếu tố ảnh hưởng đến dự đoán của mô hình. Điều này rất quan trọng đối với tính minh bạch và khả năng giải thích trong đánh giá rủi ro ESG.

- Độ chính xác: Đo lường độ chính xác tổng thể của các dự đoán của mô hình. Các số liệu như R bình phương (đối với hồi quy) và độ chính xác (đối với phân loại) cung cấp thông tin chi tiết về sức mạnh dự đoán của mô hình.

Khi nói đến dự đoán rủi ro ESG, khả năng diễn giải là một cân nhắc quan trọng. Không giống như các mô hình tài chính thuần túy, các mô hình rủi ro ESG cần phải minh bạch và dễ hiểu đối với nhiều bên liên quan, bao gồm các nhà đầu tư, cơ quan quản lý và công chúng nói chung. Điều này đặc biệt quan trọng trong bối cảnh tài chính bền vững, nơi các quyết định được đưa ra dựa trên các mô hình này có thể có tác động đáng kể đến xã hội và môi trường. Để đảm bảo khả năng diễn giải của các mô hình rủi ro ESG, cần giải quyết các khía cạnh chính sau:

- Đặc trưng quan trọng: Xác định rõ ràng các đặc trưng hoặc biến số quan trọng nhất thúc đẩy các dự đoán của mô hình, cho phép hiểu rõ hơn về các động lực cơ bản của rủi ro ESG.

- Mối quan hệ logic: Đảm bảo rằng các đầu ra của mô hình phù hợp với các mối quan hệ logic và trực quan giữa các yếu tố ESG và rủi ro, giúp người dùng dễ dàng diễn giải và tin tưởng vào kết quả.

- AI khả diễn: Áp dụng các kỹ thuật như giá trị SHAP để cung cấp các giải thích chi tiết cho từng dự đoán, cho phép các bên liên quan hiểu được lý do đằng sau các đánh giá của mô hình.

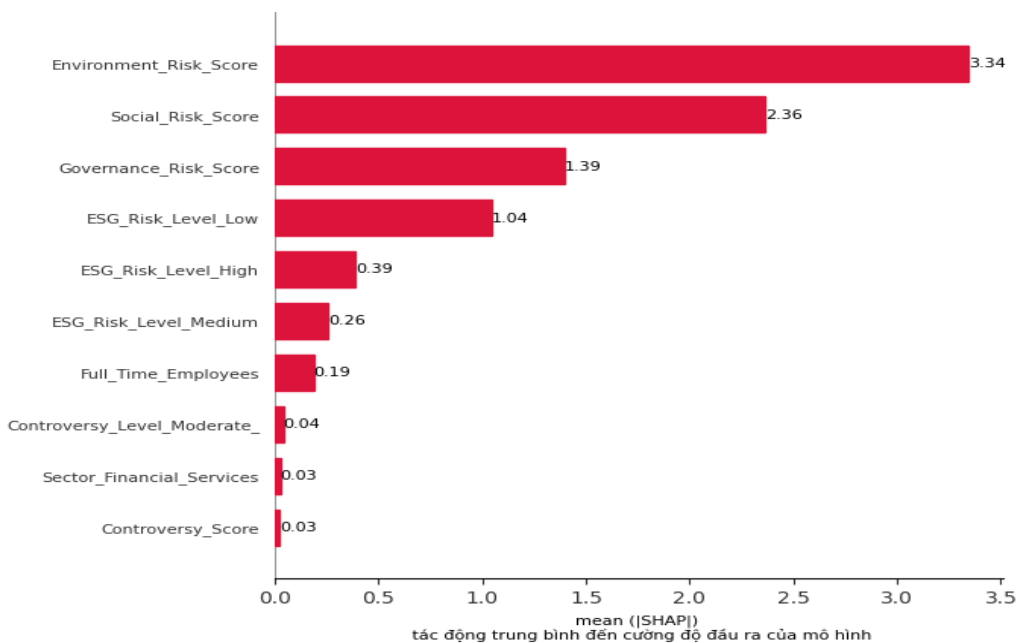
- Sự tham gia của bên liên quan: Chủ động thu hút các bên liên quan chính vào quá trình phát triển mô hình, kết hợp phản hồi của họ và đảm bảo đầu ra của mô hình được điều chỉnh theo nhu cầu và ưu tiên của họ.

Bằng cách ưu tiên khả năng diễn giải trong các mô hình rủi ro ESG, các tổ chức có thể xây dựng lòng tin, thúc đẩy quá trình ra quyết định sáng suốt hơn và cuối cùng, thúc đẩy tác động tích cực đến môi trường và xã hội thông qua các chiến lược đầu tư bền vững của họ.

4. Kết quả nghiên cứu và thảo luận

4.1. Kết quả thực nghiệm

Giải thích kết quả thực nghiệm mô hình dựa vào giá trị SHAP trung bình tuyệt đối cho mỗi đặc trưng trên toàn bộ dữ liệu. Giá trị này giúp định lượng, và đánh giá mức độ của sự đóng góp của từng đặc trưng vào kết quả dự đoán. Các đặc trưng có giá trị SHAP trung bình tuyệt đối cao hơn có ảnh hưởng lớn hơn (Hình 2). Giá trị SHAP tuyệt đối trung bình liên quan đến các đặc trưng tác động nhiều nhất đến dự đoán.



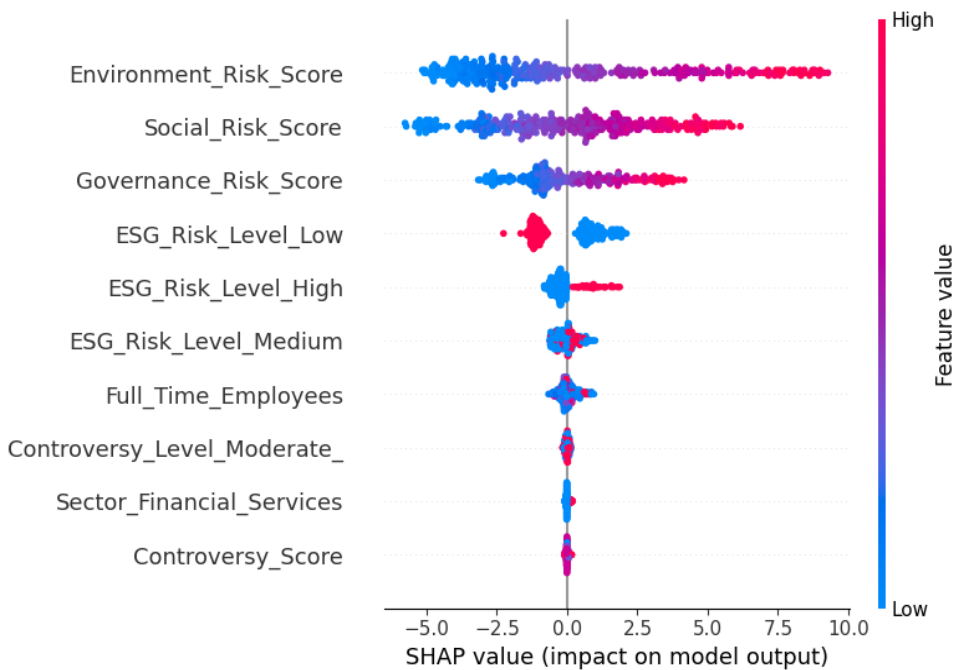
Hình 2. Giá trị SHAP trung bình tuyệt đối (mô hình LightGBM)

Hình 2 cho thấy yếu tố Environment_Risk_Score có tác động lớn nhất với giá trị 3,34, nhấn mạnh tầm quan trọng của rủi ro môi trường trong việc định hình kết quả phân tích. Tiếp theo, Social_Risk_Score đạt 2,36, cho thấy rằng các yếu tố liên quan đến trách nhiệm xã hội cũng có ảnh hưởng đáng kể, điều này cho thấy doanh nghiệp cần quản lý tốt các vấn đề xã hội để nâng cao hiệu

quả và uy tín. `Governance_Risk_Score` (1,39) cho thấy quản trị cũng đóng vai trò quan trọng, mặc dù tác động nhỏ hơn so với các yếu tố môi trường và xã hội, đòi hỏi doanh nghiệp cần cải thiện trong lĩnh vực này. Ngoài ra, `ESG_Risk_Level_Low` (1,04) thể hiện rằng mức độ rủi ro thấp trong ESG cũng là yếu tố quan trọng. Các chỉ số `ESG_Risk_Level_High` (0,39) và `ESG_Risk_Level_Medium` (0,26) có tác động thấp hơn, cho thấy rằng quản lý các rủi ro ở mức độ này là cần thiết nhưng không phải là ưu tiên hàng đầu. Các yếu tố khác như `Full_Time_Employees` (0,19), `Controversy_Level_Moderate` (0,04), `Controversy_Score` (0,03) và `Sector_Financial_Services` (0,03) không có tác động đáng kể, cho thấy rằng chúng không phải là yếu tố quyết định trong bối cảnh này. Tóm lại, các yếu tố môi trường và xã hội cần được ưu tiên trong xây dựng chiến lược phát triển bền vững, trong khi quản trị cũng cần cải thiện để đảm bảo sự phát triển lâu dài.

Tổng thể cho thấy sự ưu tiên rõ ràng cho yếu tố môi trường. Chỉ số Rủi ro môi trường luôn đạt điểm cao nhất trong cả ba mô hình, nhấn mạnh vai trò quan trọng của nó trong các đánh giá ESG. Sự khác biệt giữa các chỉ số mức độ rủi ro cho thấy cách tiếp cận khác nhau trong từng mô hình, từ đó có thể ảnh hưởng đến quyết định đầu tư và chiến lược quản lý rủi ro của các tổ chức. Nhìn chung, rủi ro môi trường được coi là yếu tố chính trong việc đánh giá tính bền vững và trách nhiệm xã hội của doanh nghiệp.

Biểu đồ đàn ong (Beewarm Plots) là một cách hiển thị phức tạp hơn và giàu thông tin hơn về các giá trị SHAP, không chỉ cho thấy tầm quan trọng tương đối của các đặc trưng mà còn cho thấy mối quan hệ thực tế của chúng với kết quả dự đoán (Hình 3). Bằng cách tính toán và tổng hợp các giá trị SHAP của tất cả các trường hợp trong tập dữ liệu, chúng ta có cái nhìn tổng thể về hành vi của mô hình.

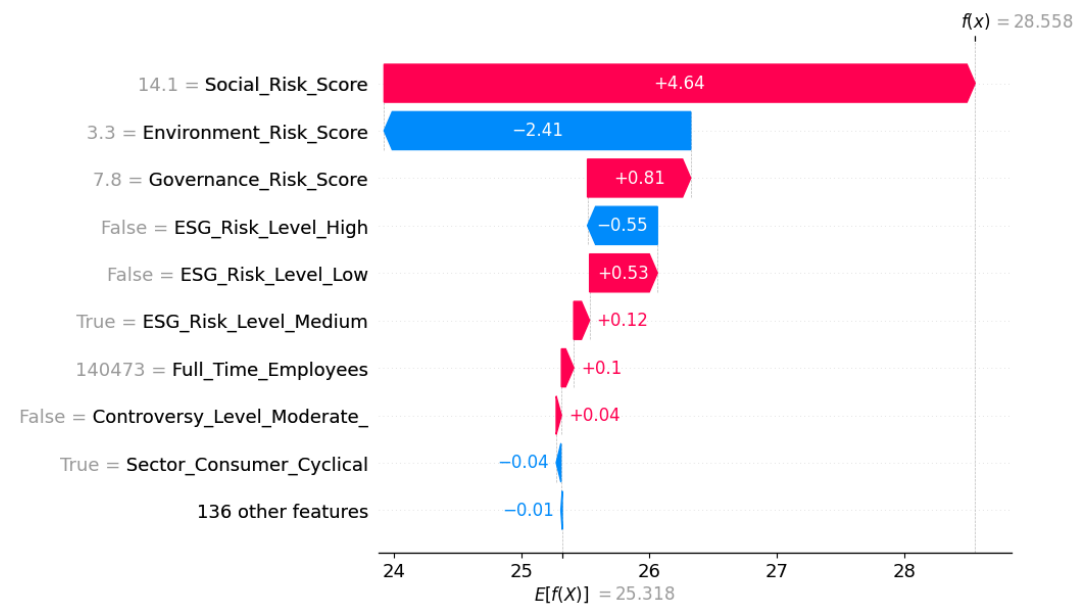


Hình 3. Kết quả thực nghiệm mô hình LightGBM (Biểu đồ đàn ong)

Phân tích biểu đồ (Hình 3) cho thấy sự phân bố và tác động của các yếu tố rủi ro ESG đến kết quả mô hình. Cụ thể, `Environment_Risk_Score` có phân phối rộng nhất từ -5 đến +10, với mật độ điểm dữ liệu cao ở cả hai đầu, cho thấy tầm ảnh hưởng mạnh mẽ theo cả hai chiều hướng. `Social_Risk_Score` thể hiện xu hướng tương tự nhưng có độ tập trung cao hơn ở phía âm (khoảng -2,0 đến -1,0), ngụ ý rằng yếu tố này thường mang tính rủi ro nhiều hơn. `Governance_Risk_Score` có phân phối hẹp hơn với nhiều điểm dữ liệu tập trung ở phía dương, đặc biệt trong khoảng 0,5 đến 1,0. Đối với các mức độ rủi ro ESG, `ESG_Risk_Level_Low` có xu hướng tích cực rõ rệt với đa số điểm dữ liệu nằm ở phía dương (0 đến 0,5), trong khi `ESG_Risk_Level_High` nghiêng về phía âm (-0,5 đến 0). Các yếu tố còn lại như `Full_Time_Employees`, `Controversy_Level_Moderate` và `Sector_Financial_Services` có phân bố tập trung quanh điểm 0 với độ phân tán thấp ($\pm 0,2$), cho thấy tác động hạn chế đến kết quả tổng thể của mô hình. Điều này phản ánh rõ nét thực tế rằng các yếu tố rủi ro cốt lõi về ESG có ảnh hưởng quyết định hơn so với các chỉ số về quy mô doanh nghiệp hay các yếu tố ngành.

Các giá trị SHAP dương cho thấy rằng khi giá trị của một đặc trưng tăng lên, dự đoán của mô hình cũng có xu hướng tăng lên. Ngược lại, các giá trị SHAP âm cho thấy rằng khi giá trị của đặc trưng tăng lên, dự đoán của mô hình có xu hướng giảm xuống. Sự phân bố này giúp hiểu rõ hơn về cách từng đặc trưng tác động đến kết quả, cho phép nhìn thấy cả những hình ảnh tích cực và tiêu cực.

Biểu đồ thác nước (Waterfall Plots) là biểu đồ hiển thị đầy đủ nhất của một dự đoán duy nhất, giải thích những đóng góp cơ bản của từng đặc trưng vào dự đoán. Cấu trúc thác nước nhấn mạnh bản chất cộng gộp của các yếu tố đóng góp tích cực và tiêu cực, và cách chúng xây dựng trên giá trị cơ sở để tạo ra dự đoán của mô hình (Hình 4). SHAP giải thích cách đưa ra các dự đoán riêng lẻ theo các đóng góp từ từng biến đầu vào của mô hình. Đây là một phương pháp tiếp cận rất trực quan tạo ra các đầu ra đơn giản nhưng nhiều thông tin.



Hình 4. Kết quả thực nghiệm mô hình LightGBM (Biểu đồ thác nước)

Hình 4 mô tả ảnh hưởng của các yếu tố đến chỉ số tổng quát $f(x)=28,558$. Trong đó, *Social_Risk_Score* có tác động tích cực lớn nhất (+4,64), cho thấy rủi ro xã hội đóng vai trò quan trọng trong việc tăng giá trị chỉ số. Ngược lại, *Environment_Risk_Score* lại có ảnh hưởng tiêu cực (-2,41), làm giảm giá trị của tổng điểm rủi ro ESG (*Total_ESG_Risk_Score*). *Governance_Risk_Score* cũng góp phần tích cực (+0,81). Ngoài ra, các yếu tố khác như mức độ rủi ro ESG và lực lượng lao động cũng ảnh hưởng, nhưng với mức tác động nhỏ hơn. Giá trị trung bình của $E(f(x))$ là 25,318, cho thấy sự chênh lệch so với giá trị hiện tại của tổng điểm rủi ro ESG (*Total_ESG_Risk_Score*).

Trong tổng thể các biểu đồ, *Social Risk Score* nổi bật là yếu tố quyết định nhất, luôn tạo ra giá trị dương cao, cho thấy tác động tích cực từ các yếu tố xã hội đối với hoạt động của doanh nghiệp. Ngược lại, *Environment Risk Score* lại liên tục thể hiện những giá trị âm, chỉ ra rằng các rủi ro liên quan đến môi trường cần được chú trọng và cải thiện ngay lập tức để hạn chế tác động xấu. Mặc dù *Governance Risk Score* có giá trị dương, nhưng ảnh hưởng của nó vẫn hạn chế hơn so với các yếu tố chính, phản ánh rằng quản trị tốt là cần thiết nhưng chưa đủ mạnh mẽ. Bên cạnh đó, các yếu tố khác như số lượng nhân viên (*Full_Time_Employees*) cũng có tác động tích cực, cho thấy rằng lực lượng lao động mạnh có thể thúc đẩy hiệu quả hoạt động. Mặc dù các yếu tố như mức độ rủi ro ESG và mức độ mâu thuẫn có giá trị thấp, chúng vẫn đóng vai trò quan trọng trong việc hình thành bức tranh tổng thể về rủi ro mà doanh nghiệp phải đối mặt. Do đó, việc xem xét đa chiều các đặc trưng này là cần thiết để xây dựng chiến lược quản lý rủi ro hiệu quả.

4.2. Độ chính xác của các mô hình

Bảng 1.

So sánh chỉ số RMSE, MAE, MSE, và MAPE của các mô hình

Model	RMSE	MAE	MSE	MAPE
Random Forest	1,6588	1,1612	2,7517	5,71%
XGB	1,6461	1,2641	2,7098	6,28%
LightGBM	1,3303	0,9233	1,7696	4,21%

MAE: LightGBM (0.9233) < Random Forest (1.1612) < XGB (1.2641)

MSE: LightGBM (1.7696) < XGB (2.7098) < Random Forest (2.7517)

RMSE: LightGBM (1.3303) < XGB (1.6461) < Random Forest (1.6588)

MAPE: LightGBM (4.21%) < Random Forest (5.71%) < XGB (6.28%)

Từ kết quả trên, LightGBM là mô hình tốt nhất trong cả 4 tiêu chí (MAE, MSE, RMSE, MAPE), nghĩa là nó có độ chính xác cao hơn XGB và Random Forest. XGB có hiệu suất kém nhất trong nhóm này vì nó có giá trị lỗi cao nhất trên tất cả các chỉ số. Random Forest có hiệu suất trung bình, tốt hơn XGB nhưng kém hơn LightGBM. Từ đó có thể kết luận rằng: LightGBM là lựa chọn tốt nhất theo kết quả này.

4.3. Thảo luận

Các giá trị Shapley giúp giải thích vai trò các đặc trưng đóng góp trong từng mô hình học máy cụ thể. Phương pháp này cung cấp các giải thích tương phản có nghĩa thay vì so sánh một dự đoán với

dự đoán trung bình của toàn bộ tập dữ liệu, chúng ta có thể so sánh nó với một tập hợp con hoặc thậm chí với một điểm dữ liệu duy nhất. Nguyên lý giải thích một dự đoán được thực hiện như một trò chơi được chơi bởi các giá trị đặc trưng.

Hơn nữa, phương pháp SHAP là một trong số ít phương pháp giải thích có lý thuyết vững chắc, vì các tiên đề toán học (hiệu quả, tính đối xứng, giả định, tính cộng) cung cấp nền tảng hợp lý cho lời giải thích.

Cuối cùng, cần đề cập rằng chiến lược này chỉ là một trong nhiều giải pháp khả thi. Tuy nhiên, giá trị Shapley thường được ưa chuộng vì nó dựa trên một lý thuyết vững chắc, phân chia công bằng các hiệu ứng và cung cấp lời giải thích đầy đủ. Tính toán giá trị Shapley dựa vào mô hình như một công cụ dự đoán với đầu vào và đầu ra cụ thể, giá trị Shapley có thể được tính toán mà không cần bất kỳ kiến thức nào về cơ chế bên trong của mô hình. Điều này làm cho nó trở thành một kỹ thuật không phụ thuộc vào mô hình và cũng tạo điều kiện so sánh trực tiếp các giá trị Shapley cho các đặc trưng đầu vào trên các loại mô hình khác nhau.

5. Kết luận và hướng phát triển

5.1. Kết luận

Nghiên cứu ứng dụng các mô hình học máy khả diễn trong đánh giá rủi ro ESG. Khi các tổ chức nỗ lực quản lý hiệu quả các rủi ro về môi trường, xã hội và quản trị (ESG) của mình, việc áp dụng các mô hình học máy khả diễn ngày càng trở nên quan trọng. Nghiên cứu đã phân tích so sánh các mô hình XAI khác nhau, nêu bật điểm mạnh và hạn chế của từng phương pháp trong việc dự đoán và phân tích các rủi ro liên quan đến ESG trên nhiều ngành khác nhau. Các khía cạnh chính của phân tích bao gồm khả năng diễn giải, độ chính xác của mô hình, yêu cầu dữ liệu và khả năng xử lý dữ liệu ESG phức tạp, đa chiều.

Ngoài ra, nghiên cứu đi sâu vào việc ứng dụng phương pháp SHAP giải thích quá trình huấn luyện mô hình học máy và kết hợp kiến thức chuyên ngành để tăng cường tính mạnh mẽ và độ tin cậy của các đánh giá rủi ro ESG. Những phát hiện từ nghiên cứu này cung cấp bằng chứng và những hiểu biết có giá trị cho các tổ chức đang tìm cách tối ưu hóa các chiến lược quản lý rủi ro ESG của mình thông qua việc triển khai chiến lược sử dụng khả năng học máy có giải thích.

5.2. Hạn chế của nghiên cứu

Phân tích rủi ro ESG bằng XAI đã cung cấp một số thông tin chi tiết có giá trị, nhưng cũng nêu bật một số hạn chế quan trọng cần cân nhắc:

- *Thứ nhất*, việc thu thập dữ liệu toàn diện và đáng tin cậy về các chỉ số ESG đầy thách thức, vì các tiêu chuẩn báo cáo và tính khả dụng của dữ liệu khác nhau giữa các ngành và khu vực. Việc tích hợp nhiều nguồn dữ liệu thông qua các lược đồ XAI đã giúp giảm thiểu điều này, nhưng vẫn còn những khoảng trống trong dữ liệu.

- *Thứ hai*, diễn giải các chỉ số phức tạp: Việc chuyển đổi dữ liệu ESG thô thành các đánh giá rủi ro có ý nghĩa đòi hỏi phải phân tích và ngữ cảnh hóa cẩn thận. Khung XAI tạo điều kiện thuận lợi cho quá trình này, nhưng việc hiểu được các sắc thái và sự phụ thuộc lẫn nhau của các yếu tố ESG khác nhau vẫn đặt ra một thách thức phân tích đáng kể.

- *Thứ ba*, các yếu tố bên ngoài: Hồ sơ rủi ro ESG của một công ty hoặc ngành có thể bị ảnh hưởng nặng nề bởi các yếu tố bên ngoài như thay đổi về quy định, tiến bộ công nghệ và xu hướng của người tiêu dùng. Việc kết hợp các yếu tố động này vào khuôn khổ phân tích rủi ro ESG dựa trên XAI rất khó khăn và đòi hỏi phải theo dõi và điều chỉnh liên tục.

Phương pháp tiếp cận dựa trên XAI để phân tích rủi ro ESG đã chứng minh là một công cụ mạnh mẽ, cung cấp một cách có cấu trúc và có thể mở rộng để đánh giá và diễn giải dữ liệu ESG phức tạp. Bằng cách liên tục tinh chỉnh các nguồn dữ liệu, mô hình phân tích và khuôn khổ diễn giải, các tổ chức có thể tận dụng phương pháp luận này để đưa ra quyết định sáng suốt hơn và thực hành kinh doanh bền vững.

5.3. Hướng phát triển

Nghiên cứu cho thấy việc tích hợp học máy có giải thích và các nguyên tắc ESG có tiềm năng to lớn trong việc chuyển đổi phân tích rủi ro và ra quyết định trong nhiều ngành. Bằng cách tận dụng sức mạnh của phân tích nâng cao, các tổ chức có thể có được những hiểu biết sâu sắc hơn, sắc thái hơn về các yếu tố môi trường, xã hội và quản trị định hình nên tính bền vững và thành công lâu dài của họ.

Tương lai sẽ chứng kiến sự nhấn mạnh hơn vào việc phát triển các mô hình học máy không chỉ chính xác mà còn minh bạch và có thể diễn giải được. Phương pháp "học máy có thể giải thích" (XAI) này sẽ rất quan trọng để xây dựng lòng tin, trách nhiệm giải trình và ra quyết định có đạo đức trong phân tích rủi ro ESG.

Trong tương lai, chúng ta có thể mong đợi các cân nhắc về ESG sẽ được nhúng sâu hơn vào cấu trúc của chiến lược kinh doanh, hoạt động và báo cáo. Học máy sẽ đóng vai trò quan trọng trong việc tổng hợp và phân tích các nguồn dữ liệu ESG đa dạng để cung cấp góc nhìn toàn diện, dựa trên dữ liệu về hiệu suất phát triển bền vững của công ty.

Các kỹ thuật học máy tiên tiến, chẳng hạn như học sâu và phân tích chuỗi thời gian, sẽ cho phép các tổ chức dự báo và dự đoán các rủi ro và cơ hội liên quan đến ESG với độ chính xác cao hơn. Điều này sẽ cho phép giảm thiểu rủi ro chủ động và định vị chiến lược trước bối cảnh môi trường, xã hội và quy định đang thay đổi nhanh chóng.

Khi lĩnh vực ESG và học máy tiếp tục phát triển, chúng ta có thể mong đợi một tương lai mà các tổ chức được trao quyền để đưa ra các quyết định sáng suốt hơn, bền vững hơn và có đạo đức hơn, thúc đẩy việc tạo ra giá trị lâu dài cho tất cả các bên liên quan.

Tài liệu tham khảo

- Assael, J., Carlier, L., & Challet, D. (2023). Dissecting the explanatory power of ESG features on equity returns by sector, capitalization, and year with interpretable machine learning. *Journal of Risk and Financial Management*, 16(3), 159.
- Bùi Minh Tiến, Nguyễn Thị Thùy Tiên, & Đặng Thanh Tùng. (2024). Xu hướng áp dụng ESG và đề xuất cho Tập đoàn Dầu khí Việt Nam. *Đặc san Dầu khí - Khoa học, Công nghệ và Đổi mới sáng tạo*, 4, 6-16.

- 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* (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>
- Dwivedi, D., Batra, S., & Pathak, Y. K. (2023). A machine learning based approach to identify key drivers for improving corporate's esg ratings. *Journal of Law and Sustainable Development*, 11(1), e0242-e0242.
- Ernst, D., & Woithe, F. (2024). Impact of the environmental, social, and governance rating on the cost of capital: Evidence from the S&P 500. *Journal of Risk and Financial Management*, 17(3), 91.
- Hồ Thị Hải Ly. (2024). Mối quan hệ giữa thực hiện trách nhiệm xã hội, mức độ nắm giữ tiền mặt, và khả năng tài chính của các doanh nghiệp ở khu vực Châu Á. *Tạp chí Kinh tế và Phát triển*, 319, 2-12.
- Nguyễn Dương Hoa. (2024). Trờ ngại với nhà đầu tư khi đánh giá các yếu tố ESG trong doanh nghiệp tại việt nam. *Tạp chí Khoa học và Công nghệ Trường Đại học Hùng Vương*, 36(3), 26-38.
- Nguyễn Minh Phương, Trần Thúy Anh, Bùi Thị Dạ Lý, Trần Bình Minh, & Đinh Phương Hà. (2024). Tác động của công bố thông tin ESG tới khả năng sinh lời của ngân hàng thương mại Việt Nam. *Tạp chí Kinh tế và Phát triển*, 330, 23-33.
- Nguyễn Quốc Huy, & Từ Lăng Phiêu. (2023). Giải thích đặc trưng thẻ tín dụng theo phương pháp lime và shap sau giai đoạn học sâu. *Tạp chí Khoa học*, 20(10), 1746-1760.
- Nguyễn Thị Liên Hương. (2023). *Nghiên cứu tác động của phát triển bền vững doanh nghiệp đến chi phí sử dụng vốn tại các doanh nghiệp niêm yết Việt Nam*. Luận án Tiến sĩ kinh tế, Trường Đại học Thương mại.
- Initiative, G. R. (2021). *How to use the GRI Standards*. Retrieved from <https://www.globalreporting.org/how-to-use-the-gri-standards/gri-standards-english-language>
- Jain, Y., Gupta, S., Yalciner, S., Joglekar, Y. N., Khetan, P., & Zhang, T. (2023). *Overcoming complexity in ESG investing: The role of generative AI integration in identifying contextual ESG factors*. Available at SSRN: <https://ssrn.com/abstract=4495647> or <http://dx.doi.org/10.2139/ssrn.4495647>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf
- Raza, H., Khan, M. A., Mazliham, M., Alam, M. M., Aman, N., & Abbas, K. (2022). Applying artificial intelligence techniques for predicting the environment, social, and governance (ESG) pillar score based on balance sheet and income statement data: A case of non-financial companies of USA, UK, and Germany. *Frontiers in Environmental Science*, 10, 975487. <https://doi.org/10.3389/fenvs.2022.975487>
- Rigatti, S. J. (2017). Random forest. *Journal of Insurance Medicine*, 47(1), 31-39.

- Shapley, L. (1953). A Value for n-Person Games. In: Kuhn, H., & Tucker, A. (Eds.), *Contributions to the Theory of Games II*. Princeton: Princeton University Press (pp.307-317). <https://doi.org/10.1515/9781400881970-018>
- Tiwari, R. (2023). Explainable ai (xai) and its applications in building trust and understanding in AI decision making. *International Journal of Scientific Research in Engineering and Management*, 7, 1-13.
- Trần Ngọc Hùng. (2023). ESG Trong môi trường bất định Covid-19: Nghiên cứu thực nghiệm tại các doanh nghiệp Việt Nam. *Tạp chí Kinh tế và Phát triển*, 311(2), 44-53.
- Veeramani, H., Thapa, S., & Naseem, U. (2024). Enhancing esg impact type identification through early fusion and multilingual models. *arXiv preprint arXiv:2402.10772*. <https://doi.org/10.48550/arXiv.2402.10772>
- Zhang, A. Y., & Zhang, J. H. (2024). Renovation in environmental, social and governance (ESG) research: the application of machine learning. *Asian Review of Accounting*, 32(4), 554-572.