

NHẬN DIỆN SAI SÓT BÁO CÁO TÀI CHÍNH BẰNG HỌC MÁY: NGHIÊN CỨU THỰC NGHIỆM TẠI THỊ TRƯỜNG VIỆT NAM

**Detecting financial reporting misstatements using machine learning:
an empirical study in Vietnam**

TIÊU THỊ THANH HOA
TRẦN NGỌC HÙNG

*N*ghiên cứu này tập trung nhận diện sai sót trọng yếu trong báo cáo tài chính của các doanh nghiệp niêm yết tại Việt Nam thông qua ứng dụng học máy. Bài viết khai thác dữ liệu tài chính và phi tài chính từ nền tảng FiinPro giai đoạn 2016 - 2023, sử dụng hồi quy LASSO để chọn lọc 33 biến đầu vào. Ba mô hình dự đoán được áp dụng gồm hồi quy logistic, rừng ngẫu nhiên và XGBoost, với kỹ thuật xác thực chéo k -fold ($k=10$). Kết quả cho thấy, các mô hình học máy vượt trội hơn thống kê truyền thống, đặc biệt mô hình rừng ngẫu nhiên đạt độ chính xác 81,55% và F1-score 73,38%. Nghiên cứu khẳng định hiệu quả của việc kết hợp dữ liệu sẵn có với thuật toán học máy trong phát hiện sai sót báo cáo tài chính, đồng thời đề xuất hướng mở rộng tích hợp thêm biến phi tài chính và chuyên môn kế toán nhằm nâng cao khả năng phát hiện trong tương lai.

Từ khóa: sai sót báo cáo tài chính, học máy, rừng ngẫu nhiên.

This study aims to identify material misstatements in the financial statements of listed companies in Vietnam using machine learning techniques. Financial and non-financial data were collected from the FiinPro platform, covering the period from 2016 to 2023. LASSO regression was applied to select 33 input variables. Three predictive models - Logistic Regression, Random Forest, and XGBoost - were employed, with 10-fold cross-validation used for model evaluation. The results show that machine learning models outperform traditional statistical methods, with the Random Forest model achieving an accuracy of 81.55% and an F1-score of 73.38%. The study demonstrates the effectiveness of combining available financial data with machine learning algorithms in detecting misstatements in financial reporting. It also suggests that future research should integrate additional non-financial indicators and accounting expertise to further enhance prediction performance.

Keywords: financial statement misstatements, machine learning, random forest.

1. Đặt vấn đề

Báo cáo tài chính là kênh thông tin cốt lõi phản ánh sức khỏe tài chính, hiệu quả hoạt động và dòng tiền của doanh nghiệp, từ đó định hướng các quyết định kinh tế của nhà đầu tư, cơ quan quản lý và các bên liên

quan. Một sai sót trọng yếu - dù do vô ý hay gian lận - có thể bóp méo bức tranh tài chính,

Tiêu Thị Thanh Hoa, ThS., NCS; Trần Ngọc Hùng, PGS.TS. (tác giả liên hệ), Trường đại học Công nghiệp Thành phố Hồ Chí Minh.