1. **TÓM TẮT CÁC BÀI NGHIÊN CỨU**
2. **The dilemma of accuracy in bankruptcy prediction: a new approach using explainable AI techniques to predict corporate crises**

[**https://www.emerald.com/insight/content/doi/10.1108/ejim-06-2024-0633/full/pdf?title=the-dilemma-of-accuracy-in-bankruptcy-prediction-a-new-approach-using-explainable-ai-techniques-to-predict-corporate-crises**](https://www.emerald.com/insight/content/doi/10.1108/ejim-06-2024-0633/full/pdf?title=the-dilemma-of-accuracy-in-bankruptcy-prediction-a-new-approach-using-explainable-ai-techniques-to-predict-corporate-crises)

1. **Tóm tắt của bài báo**

* Mục tiêu của bài báo: Phát triển 1 mô hình dự đoán sự khủng hoảng của doanh nghiệp có tính chính xác cao hơn so với các phiên bản trước đó.
* Phương pháp được sử dụng: sử dụng AI cho dự đoán vỡ nợ của doanh nghiệp với một cách tiếp cận mới lạ dựa trên việc xen lẫn các kĩ thuật với nhau, cho phép mô hình đạt được độ chính xác cao hơn. Nhóm nghiên cứu đã xây dựng 1 mô hình với các các độ cho dài chuỗi khác nhau, có thể là độ dài cố định hoặc thay đổi. Sau đó chọn ra mô hình có biến độ dài chuỗi là tốt nhất.
* Kết quả: Kỹ thuật AI ra kết quả dự đoán có độ chính xác rất cao so với các nghiên cứu trước đó, ngay cả khi so sánh với những nghiên cứu sử dụng chuỗi thời gian dài hoặc là số lượng quan sát lớn.

1. **Các phương pháp được sử dụng trong bài báo:**

* Các kỹ thuật được sử dụng trong bài:

+ Synthetic Minority Over-sampling Technique (SMOTE) -> dùng để xử lý dữ liệu unbalance

+ Model: RNN, LSTM

Train with fixed sequence lengths (FSLs; from lengths 2 to 9)

Train with variable sequence lengths (VSL)

+ Loss function: Binary Cross-Entropy -> mục tiêu tối ưu hóa của mô hình

+ AUC, Accuracy, ROC, … -> Đánh giá hiệu suất mô hình

+ K-fold cross-validation -> đánh giá hiệu suất mô hình

1. **Sử dụng mô hình RNN (Recurrent Neural Network) để xây dựng mô hình**

* Các bước:

+ B1: Sử dụng 4 kiến trúc RNN khác nhau về số lượng và dạng của recurrent layer

+ B2: Đánh giá và lựa chọn ra 1 mô hình tốt nhất dựa trên các chỉ số đo độ chính xác (accuracy, ROC curve).

+ B3: Tiến hành cross-validation để đảm bảo mô hình là khái quát và mạnh cho tất cả các phần của bộ dữ liệu: Train mô hình nhiều lần với các tập dữ liệu đầu vào khác nhau và với các độ dài chuỗi cố định hoặc thay đổi (Sử dụng K-fold cross-validation technique)

1. Sử dụng phương phá AI có khả năng giải thích, gọi là SHAP (SHapley Additive explanation) để giải thích sự đóng góp của mỗi chỉ số tài chính trong việc dự đoán.
2. **Khoảng trống nghiên cứu:**

Bài nghiên cứu mới chỉ sử dụng dữ liệu kế toán, có thể sử dụng dụng các biến phi tài chính khác để cải thiện độ chính xác:

+ Tích hợp thông tin về quản trị, ESG (Environmental (Môi trường), Social (Xã hội) và Governance (Quản trị)) và mạng xã hội vào mô hình RNN -> cung cấp thông tin chi tiết theo thời gian thực, khắc phục những hạn chế của báo cáo tài chính chậm trễ và cải thiện độ chính xác dự đoán.

+ Sử dụng LSTM để xử lý văn bản không có cấu trúc từ các chỉ số quản trị, ESG mạng xã hội và các bài báo tin tức có tiềm năng nắm bắt xu hướng kinh tế, do đó nâng cao khả năng dự đoán tổng thể của mạng nơ-ron trong việc xác định các tình huống kinh doanh rủi ro và báo hiệu tình trạng khó khăn tài chính tiềm ẩn.

1. **AI FOR CREDIT RISK MODELING: A DEEP LEARNINGAPPROACH**

[**http://www.upubscience.com/upload/20250314112655.pdf**](http://www.upubscience.com/upload/20250314112655.pdf)

1. **Tóm tắt:**

Bài báo sử dụng các loại mô hình DL khác nhau và sau đó so sánh performance giữa các mô hình

1. **Các kỹ thuật sử dụng:**

+ GNN

+ GAN

+ Kỹ thuật AI có thể giải thích (explainable AI - XAI)

1. **Khoảng trống của nghiên cứu:**

* Tính toán phức tạp với GNNs và mạng lưới đối nghịch (adversarial networks) trong bộ dữ liệu tài chính với quy mô lớn.
* Đánh đổi giữa fairness constraints and classification accuracy, các nghiên cứu sau có thể khám phá ra cách thức để cân bằng giữa tính công bằng và độ mạnh của dự đoán để đảm bảo rằng model giữ được cả tính ensuring that credit risk models remain both tính đúng đắn và tính hiệu quả.
* Các nghiên cứu tiếp theo cũng có thể tập trung vào việc tích hợp các nguồn dữ liệu tài chính đa phương thức, bao gồm hành vi giao dịch, mô hình chi tiêu và các chỉ số chấm điểm tín dụng thay thế, để cải thiện hồ sơ rủi ro của người vay.
* Ngoài ra, bài báo gợi ý việc mở rộng framework của mô hình rủi ro tín dụng xuyên biên giới sẽ tăng cường khả năng áp dụng của nó trong các thị trường tài chính quốc tế, đảm bảo rằng các đánh giá rủi ro do AI điều khiển vẫn hiệu quả trong các môi trường cho vay đa dạng.

1. **AI-Based Credit Scoring Models in Microfinance: Improving Loan Accessibility, Risk Assessment, and Financial Inclusion**

[**http://thecrsss.com/index.php/Journal/article/view/370/424**](http://thecrsss.com/index.php/Journal/article/view/370/424)

1. **Tóm tắt:**

Bài biết mô tả rất chi tiết về sự phát triển của các mô hình chấm điểm tín dụng, từ các mô hình truyền thống tới các mô hình AI. Và nhấn mạnh vai trò cũng như những lợi ích mà AI mang lại nhưng cũng lưu ý phải đảm bảo tính công bằng, minh bạch và sử dụng công nghệ có trách nhiệm trong tài chính vi mô. Bài báo cũng cung cấp các ví dụ trong việc áp dụng AI và khám phá tiềm năng của AI trong việc tăng cường hòa nhập tài chính,

Ngoài ra tác giả đưa ra các nguồn dữ liệu mới, như việc sử dụng điện thoại di động, giao dịch, tài khoản mạng xã hội và dữ liệu sinh trắc học để tạo điểm tín dụng mới cho người vay

1. **Hạn chế/khoảng trống nghiên cứu:**

Bài viết mới chỉ dừng lại ở việc đưa ra những đánh giá và khung lý thuyết tổng quan mà chưa có áp dụng vào xây dựng mô hình trên dữ liệu thực tế.

1. **A Deep Learning Approach for Credit Scoring of Peer-to-Peer Lending Using Attention Mechanism LSTM**

[**https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8579130**](https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8579130)

1. **Tóm tắt**

Tác giả sử dụng dữ liệu về hành vi hoạt động trực tuyến của người vay và đề xuất phương pháp chấm điểm tín dụng tiêu dùng dựa trên cơ chế chú ý LSTM, đây là ứng dụng mới của thuật toán học sâu. Tác giả đã lấy ý tưởng dựa trên Word2vec

1. **Hạn chế/Khoảng trống của mô hình**

Nghiên cứu chỉ xem xét tới trình tự các sự kiện trong dữ liệu hành vi hoạt động của người dùng trực tuyến mà không xem xét khoảng thời gian mà sự kiện xảy ra. Tuy nhiên, khoảng thời gian của sự kiện cũng chứa thông tin quan trọng và cần được xem xét để đưa vào mô hình.

1. **Neural network credit scoring models**

[**https://www.researchgate.net/profile/David-West/publication/223425357\_Neural\_Network\_Credit\_Scoring\_Models/links/5ae9c71c45851588dd826629/Neural-Network-Credit-Scoring-Models.pdf**](https://www.researchgate.net/profile/David-West/publication/223425357_Neural_Network_Credit_Scoring_Models/links/5ae9c71c45851588dd826629/Neural-Network-Credit-Scoring-Models.pdf)

1. **Explainable** **Deep Learning Approaches to Credit Risk Evaluation**

[**https://dl.acm.org/doi/pdf/10.1145/3675888.3676052**](https://dl.acm.org/doi/pdf/10.1145/3675888.3676052)

1. **DỮ LIỆU NGHIÊN CỨU**

– Dữ liệu khách hàng cá nhân:

+ Bộ cá nhân cho dự đoán multi class: <https://www.kaggle.com/datasets/parisrohan/credit-score-classification/data>

-> Bộ dữ liệu này không có thông tin tường minh về dữ liệu được collect từ đâu.

+ Bộ KH cá nhân của Home Credit cho dự đoán 2 lớp:

<https://www.kaggle.com/competitions/home-credit-credit-risk-model-stability/data>

-> cần check lại quyền access (có thể add thêm các thông tin về mức thu nhập theo ngành nghề, tình hình kinh tế ở khu vực đang sinh sống,.v..v.. )

* Dữ liệu doanh nghiệp

+ Bộ dữ liệu bankruptcy của doanh nghiệp tại HCM:

<https://opendata.hochiminhcity.gov.vn/dataset/danh-s%C3%A1ch-c%C3%A1c-doanh-nghi%E1%BB%87p-hi%E1%BB%87n-nay> ,

và collect thêm các thông tin khác trên các trang truyền thông như báo cáo tài chính của các doanh nghiệp trên web, thông tin báo chí trên mạng truyền thông..

Dữ liệu doanh nghiệp đang hoạt động: là thông tin các doanh nghiệp mới thành lập từ 2016 -> 2020 và còn hoạt động

Dữ liệu doanh nghiệp đã dừng hoạt động: doanh nghiệp có thể được thành lập trước đó nhưng dừng hoạt động trong khoảng thời gian 2026 -> 2020

* Cần collect thêm cả những doanh nghiệp được thành lập trước đó, cùng thời gian với những doanh nghiệp đã phá sản nữa.
* Hoặc là ở những công ty đã phá sản, chỉ lấy những công ty nào được thành lập trong thời gian 2016 -> 2020
* Hoặc là collect thông tin các doanh nghiệp trong danh sách thành lập từ 2016-2020 mà đã phá sản

**Dữ liệu**

Dữ liệu ngày thành lập của các doanh nghiệp:

**Thư viện pháp luật**: [Tại đây](https://thuvienphapluat.vn/ma-so-thue/tra-cuu-ma-so-thue-doanh-nghiep?timtheo=ma-so-thue&tukhoa=&ngaycaptu=&ngaycapden=&ngaydongmsttu=&ngaydongmstden=&vondieuletu=&vondieuleden=&loaihinh=0&nganhnghe=0&tinhthanhpho=79&quanhuyen=0&phuongxa=0) -> kiểm tra xem có lấy được full danh sách không

* Có ngày đóng mã số thuế, trường hợp đã giải thể nhưng chưa đóng thì sao?

**Tổng cục thuế**: [Tại đây](https://tracuunnt.gdt.gov.vn/tcnnt/mstdn.jsp) -> chỉ có cách search từng mst -> kiểm tra xem có lấy full danh sách được k

* Phải điền nhiều thông tin như căn cước, mã xác thực -> k lấy được

**Masothue:** [Tại đây](https://masothue.com/tra-cuu-ma-so-thue-theo-tinh/ho-chi-minh-23) -> kiểm tra xem có lấy được full danh sách k

* K có thông tin ngày giải thể của doanh nghiệp

**Cổng thông tin quốc gia về đăng kí doanh nghiệp:** [Tại đây](https://dichvuthongtin.dkkd.gov.vn/inf/Forms/Products/ProductCatalog.aspx?h=2511) -> kiểm tra xem có lấy được full danh sách k

* Phải điền nhiều thông tin + mã xác thực -> k lấy được

**Doanh nghiệp mới:** [Tại đây](https://doanhnghiepmoi.vn/thong-tin/Cong-Ty-TNHH-Thuong-Mai-Dich-Vu-Dau-Tu-Xay-Dung-BT-1ec32.html)

* Phải đi qua 2 layer, có thông tin ngày giải thể tuy nhiên phải cắt chuỗi, 1 vài thông tin chưa chính xác (ví dụ mã 0310227642), có thể cân nhắc lấy
* Với những doanh nghiệp giải thể, lấy thông tin trong Masothue do chỉ cần thông tin ngày bắt đầu thành lập
* Với những doanh nghiệp hoạt động/đang nộp đơn xin giải thể thì ưu tiên lấy thông tin ở Thư viện pháp luật -> sau đó tới doanh nghiệp mới