**LỰA CHỌN TÍNH NĂNG BẰNG CÁCH**

**SỬ DỤNG MÔ HÌNH TỐI ƯU HOÁ**

**ĐA CẤP BẬC ĐỂ XẾP HẠNG ĐIỂM TÍN DỤNG**

**Tóm lược**: Nhận dạng và phòng ngừa gian lận tài chính đang trở thành thách thức quyết định đối với bất kỳ tổ chức tài chính và doanh nghiệp nào. Phần lớn các ngân hàng sử dụng điểm tín dụng hoặc một giá trị bằng số hoặc thống kê để ước tính mức độ tin cậy của khách hàng dựa trên lịch sử tín dụng của họ. Mô hình điểm tín dụng sẽ được xây dựng bằng cách sử dụng đào tạo dữ liệu và hơn nữa, một quy trình phân tích sẽ được áp dụng để ước tính điểm tín dụng của từng khách hàng.

Quá trình này không chỉ giúp ngân hàng định rõ hơn về khả năng thanh toán nợ của khách hàng dựa trên điểm tín dụng của họ mà còn giúp họ phát hiện và phòng ngừa các hoạt động gian lận tiềm ẩn. Bằng cách tự động hóa quy trình này, ngân hàng có thể tiết kiệm thời gian và nguồn lực, đồng thời cải thiện độ chính xác trong việc đánh giá rủi ro tài chính. Tuy nhiên, cần lưu ý rằng không có mô hình nào là hoàn hảo và luôn cần phải cập nhật và tinh chỉnh mô hình để phù hợp với thực tế tài chính đang thay đổi liên tục.

Đối với việc cải tiến và cập nhật mô hình, việc thu thập và phân tích dữ liệu là yếu tố then chốt. Ngân hàng cần phải tổ chức và lưu trữ dữ liệu một cách hợp lý để dễ dàng truy cập và sử dụng nó cho việc phân tích và dự đoán. Đồng thời, việc sử dụng công nghệ tiên tiến như trí tuệ nhân tạo và học máy có thể giúp cải thiện hiệu suất của mô hình và cho phép ngân hàng xử lý một lượng lớn dữ liệu một cách nhanh chóng và chính xác.

Không chỉ vậy, việc cung cấp đào tạo cho nhân viên về cách sử dụng và hiểu các mô hình điểm tín dụng cũng vô cùng quan trọng. Điều này không chỉ giúp họ nắm bắt được cách thức hoạt động của mô hình mà còn giúp họ nắm bắt được các tín hiệu cảnh báo về gian lận tài chính một cách kịp thời.

Như vậy, đối với bất kỳ mô hình điểm tín dụng, người cho vay thu thập nhiều dữ liệu khác nhau về khách hàng từ nhiều cơ quan bên ngoài khác nhau. Việc thu thập được dữ liệu có thể bao gồm các tham số không liên quan sẽ không giúp đưa ra bất kỳ quyết định nào và cũng làm giảm tốc độ toàn cầu hiệu suất của mô hình. Theo đó, việc lựa chọn đặc tính là một quá trình bắt buộc để loại bỏ các thuộc tính ít liên quan hơn. cho bất kỳ tập dữ liệu nào, đặc biệt là trong việc chấm điểm tín dụng.

Bài viết này nhấn mạnh vào việc làm sáng tỏ hiệu suất của thuộc tính quá trình lựa chọn bằng cách sử dụng mô hình điểm số nhiều cấp bậc. Phương pháp đề xuất tích lũy kết quả bằng cách sử dụng phương pháp tối ưu hóa thuật toán ngưỡng và hoạt động tốt hơn trong việc lựa chọn thuộc tính chất lượng cho mô hình điểm tín dụng cơ bản. Các đánh giá thực nghiệm đã được thực hiện và đã chứng minh được tính chính xác và hiệu quả của các phương pháp đề xuất phương pháp này tương đối tốt hơn so với các kỹ thuật xếp hạng đơn hiện có.

1. **Giới thiệu**

Các tổ chức tài chính hoặc ngân hàng trở thành trung tâm tập trung vào 90% công chúng ở Ấn Độ một cách rõ ràng sau khi hủy kiếm tiền. Rõ ràng, khi ngân hàng sở hữu giao dịch tiền, đó là một trong những nơi hấp dẫn để kẻ lừa đảo thực hiện các hoạt động liên quan đến tiền giao dịch. Gian lận trong lĩnh vực tài chính đã tồn tại từ giai đoạn đầu trong hơn một thập kỷ. Tuy nhiên, trong những năm qua, các vụ lừa đảo trong lĩnh vực tài chính đã trở nên phức tạp hơn do sự mở rộng tất cả các dịch vụ dựa trên công nghệ được cung cấp cho khách hàng trên toàn thế giới. Khu vực tài chính Ấn Độ cũng chịu đựng nỗi đau do gian lận leo thang sự cố. Một cuộc khảo sát cho thấy rằng 93% gian lận có tăng trưởng trong giai đoạn 2018-2020 [1]. Theo Ngân hàng Dự trữ Ấn Độ (RBI), trong 5 năm qua, hơn 23.000 trường hợp gian lận bao gồm một số tiền khổng lồ 1 vạn rupee đã được báo cáo từ một số ngân hàng. Theo chi tiết được cung cấp bởi ngân hàng trung ương, số tiền cao nhất là 28.459 rupee có liên quan đến tất cả các trường hợp gian lận được thông báo từ tháng 4 năm 2017 đến tháng 3 năm 2018 [2]. Dù gian lận các hoạt động phổ biến trong các ứng dụng web tiếp xúc với nhiều lỗ hổng khác nhau [3] cho một giao dịch tài chính diễn ra, trọng tâm chính của việc này bài viết chỉ nói về việc quản lý rủi ro tín dụng trong lĩnh vực tài chính.

Chính vì vậy, việc áp dụng các phương pháp, công cụ và kỹ thuật mới nhất để phát hiện và ngăn chặn gian lận tài chính là điều cần thiết. Điều này không chỉ giúp bảo vệ ngân hàng và khách hàng khỏi rủi ro tài chính mà còn góp phần vào việc xây dựng niềm tin và uy tín cho tổ chức tài chính.

Ví dụ, việc sử dụng trí tuệ nhân tạo và học máy trong việc xử lý và phân tích dữ liệu có thể giúp ngân hàng dự đoán và phát hiện sớm các hành vi gian lận. Những công nghệ này cũng giúp các ngân hàng tự động hóa quy trình đánh giá rủi ro, giảm thiểu khả năng sai sót và tăng cường hiệu quả của các hoạt động tài chính.

Ngoài ra, việc đào tạo và nâng cao năng lực cho nhân viên cũng cần được coi trọng. Khi nhân viên có đủ kiến thức và kỹ năng, họ sẽ có khả năng phát hiện và xử lý kịp thời các tình huống gian lận, đồng thời đưa ra những quyết định chính xác hơn trong việc quản lý rủi ro tài chính.

Trên hết, việc xây dựng một nền văn hóa minh bạch và tuân thủ nguyên tắc trong ngân hàng cũng đóng vai trò quan trọng. Một nền văn hóa mạnh mẽ về tuân thủ có thể ngăn ngừa gian lận bằng cách khuyến khích các hành vi đúng đắn và trừng phạt những hành vi sai trái.

Tóm lại, việc ngăn chặn gian lận tài chính đòi hỏi một cách tiếp cận toàn diện, từ việc áp dụng công nghệ mới, đào tạo nhân viên, tới việc xây dựng văn hóa tuân thủ. Chỉ khi thực hiện đồng bộ các biện pháp này, ngân hàng và các tổ chức tài chính mới có thể bảo vệ mình và khách hàng khỏi những rủi ro do gian lận tài chính gây ra. Mỗi tổ chức tài chính đều thúc đẩy hoạt động riêng của mình chính sách rủi ro tín dụng kiểm soát các cam kết cấp tín dụng của ngân hàng và áp dụng các biện pháp thiết yếu kỹ thuật điều khiển các hoạt động. Cách tiếp cận này cung cấp mức độ thỏa đáng của sự đánh đổi giữa rủi ro và lợi nhuận cho những thành tựu của nó [4].

Về cơ bản, lĩnh vực tài chính thực hiện việc chấm điểm tín dụng Mô hình bao gồm thông tin về khách hàng và các giao dịch của họ sẽ được hữu ích trong việc đưa ra quyết định thông minh và hiệu quả [5]. Dựa trên thông tin tín dụng của khách hàng, tín dụng mô hình điểm kiểm tra và tính toán điểm tín dụng. Mô hình điểm tín dụng là bài kiểm tra có thể định lượng được được các chuyên gia tư vấn tín dụng khai thác để ước tính giá trị của khách hàng trong việc tiếp cận tín dụng. Điểm số là được hình thành dựa trên phân tích thống kê của hành vi của khách hàng và đặc điểm của quá khứ khoản vay để dự đoán các khoản nợ phải trả cho các khoản vay trong tương lai [6]. Các dự đoán rủi ro tín dụng là vấn đề chiếm ưu thế nhất mà các tổ chức tài chính phải đối mặt. Do tăng cường cạnh tranh, nâng cao sự đáp ứng của khách hàng với rủi ro khả năng kiểm soát và cải thiện giao dịch trình độ đã trở thành ý nghĩa quan trọng nhất trong lĩnh vực tài chính [7]

Chấm điểm tín dụng là cần thiết để thế chấp các khoản vay khác nhau và các dịch vụ cần thiết như bảo hiểm và mua thẻ tín dụng. Tuy nhiên, các công cụ chấm điểm tín dụng tự động sẽ nâng cao mối quan ngại lâu dài về tính chính xác và không công bằng [8].

Ở Ấn Độ, bốn công ty thông tin tín dụng đang được RBI cấp phép. Mỗi công ty phục vụ với điểm tín dụng cá nhân bằng cách sử dụng các mô hình khác nhau. Phương pháp được sử dụng rộng rãi nhất là điểm tín dụng CIBIL. Phương pháp này có số điểm dao động từ 300 đến 900, trong đó 750 là điểm trung bình và 900 là điểm cao nhất [9]. Hai thước đo quan trọng như độ chính xác và sự rõ ràng phải được quan tâm trong khi triển khai mô hình quyết định chấm điểm tín dụng [10]. Chỉ có điểm do cơ quan bên ngoài đưa ra không đủ điều kiện để đưa ra quyết định. Kết quả là, các tổ chức tài chính phát triển rủi ro tín dụng của họ chính sách tùy thuộc vào các giao dịch được thực hiện và các chính sách khác chi tiết liên quan đến khách hàng. Như vậy, ngân hàng thu thập thông tin từ nhiều nơi khác nhau và thực hiện mô hình của họ bên cạnh điểm số do các cơ quan bên ngoài cung cấp. Việc thu thập được thông tin là nền tảng cho việc đưa ra bất kỳ quyết định nào.

Tuy nhiên, tập dữ liệu cơ bản bao gồm thuộc tính ngoại lệ không liên quan đến tín dụng mô hình tính điểm. Các thuộc tính không liên quan đến nghiên cứu và dư thừa trong khi đồng hóa dữ liệu mua từ hai nơi trở lên được đặt tên là thuộc tính ngoại lệ hoặc thuộc tính ngoại lệ. Những thuộc tính này các ngoại lệ không chỉ làm tăng thời gian tính toán mà còn cũng làm giảm hiệu quả và độ chính xác của mô hình [11]. Vì vậy, việc lựa chọn đặc tính đóng vai trò thiết yếu trong bất kỳ mô hình ra quyết định nào.

Bài viết này tập trung vào việc lựa chọn các tính năng có liên quan đến việc phân loại rủi ro tín dụng bằng cách sử dụng tối ưu hóa nhiều điểm xếp hạng. Phương pháp đề xuất mang lại độ chính xác tốt hơn vì nó sử dụng nhiều cấp bậc chấm điểm cho các tính năng Cuối cùng, điểm xếp hạng là tích lũy dựa vào đó sẽ đưa ra quyết định được thực hiện trong việc lựa chọn các tính năng hữu ích cho việc nâng cao mô hình quyết định cơ bản lên cấp độ tiếp theo. Hơn nữa, phương pháp này sử dụng thuật toán tối ưu hóa trong việc tích lũy điểm số.

1. **Dự đoán và quản lý rủi ro tín dụng bằng cách triển khai thẻ điểm bằng chiến lược kết hợp với xếp hạng độ tin cậy**

Các tổ chức tài chính trên toàn cầu đều phải đối mặt với rủi ro tín dụng lớn khi tiến hành đánh giá hồ sơ xin cấp tín dụng của khách hàng. Trong thời gian gần đây, để có thể đánh giá, quản lý và đưa ra quyết định một cách chính xác về rủi ro tín dụng từ phía khách hàng, các tổ chức tài chính đã quyết định áp dụng hệ thống thẻ điểm nội bộ. Tuy nhiên, các ngân hàng có quy mô lớn lại thường xuyên sử dụng một số mô hình chấm điểm tín dụng một chiều hiện có trên thị trường, điều này có thể dẫn đến kết quả đánh giá không chính xác và khách quan về khả năng trả nợ của khách hàng. Trong bài báo này, chúng tôi đã đề xuất một kỹ thuật chấm điểm tín dụng kết hợp ba chiều mới, bao gồm việc chấm điểm ứng dụng tuần tự cùng với mô hình ma trận chấm điểm tín dụng kép. Mô hình chấm điểm tín dụng kép này sử dụng cả điểm tín dụng hành vi lẫn điểm tín dụng của cơ quan tín dụng để tính toán và xếp hạng độ tin cậy của khách hàng, giúp tăng độ chính xác và minh bạch trong việc đánh giá rủi ro tín dụng.

Mô hình chấm điểm tín dụng kép này được thiết kế dựa trên hai yếu tố chính: điểm tín dụng hành vi và điểm tín dụng của cơ quan tín dụng. Điểm tín dụng hành vi được tính toán dựa trên hành vi thanh toán nợ của khách hàng trong quá khứ, còn điểm tín dụng của cơ quan tín dụng được xác định dựa trên đánh giá của các cơ quan tín dụng đối với khách hàng.

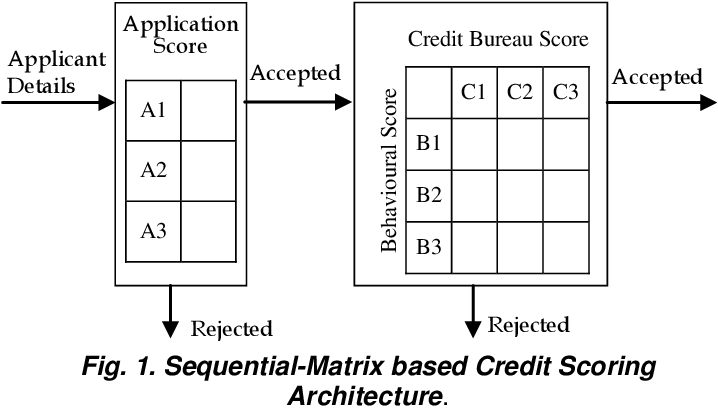
Tính năng độc đáo của mô hình này là việc kết hợp hai yếu tố trên để tạo ra một ma trận chấm điểm tín dụng kép. Ma trận này không chỉ giúp đánh giá được khả năng thanh toán nợ của khách hàng một cách toàn diện hơn mà còn giúp ngân hàng đưa ra quyết định về việc cấp tín dụng một cách chính xác hơn.

Ngoài ra, kỹ thuật chấm điểm tín dụng kết hợp ba chiều còn bao gồm việc chấm điểm ứng dụng tuần tự. Điều này có nghĩa là, thay vì chỉ xem xét thông tin tín dụng hiện tại của khách hàng, ngân hàng còn xem xét các hồ sơ ứng dụng tín dụng trước đó của khách hàng để đánh giá khả năng thanh toán nợ của họ. Qua đó, ngân hàng có thể đưa ra những quyết định chính xác hơn về việc cấp tín dụng cho khách hàng.

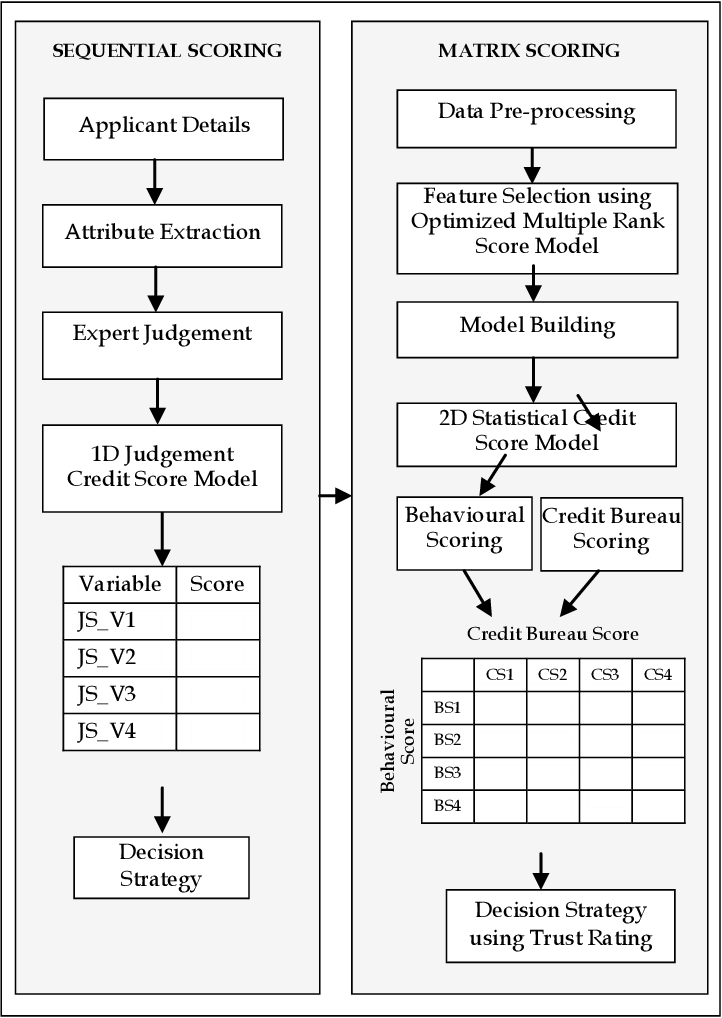
Với sự kết hợp của các phương pháp chấm điểm tín dụng truyền thống và kỹ thuật chấm điểm tín dụng kết hợp ba chiều mới này, ngân hàng có thể cải thiện đáng kể độ chính xác trong việc đánh giá rủi ro tín dụng, đồng thời giảm thiểu rủi ro tài chính và tăng cường sự tin tưởng của khách hàng đối với tổ chức tài chính.

Ngoài ra, mô hình chấm điểm hành vi sử dụng lựa chọn tính năng dựa trên nhiều điểm xếp hạng được tối ưu hóa để ghi điểm chính xác. Khi sử dụng phương pháp xếp hạng độ tin cậy dựa trên phương pháp đã ký, khách hàng được phân loại thành ba nhóm rủi ro để đánh giá và quản lý rủi ro tín dụng của khách hàng. Các chiến lược tín dụng cần tuân thủ khi đưa ra quyết định cũng được trình bày cùng với các phân tích thực nghiệm.

Kết quả phân tích cho thấy phương pháp đề xuất cho độ chính xác 88% với giá trị thống kê 43,17 K-S.[12]



Kiến trúc chấm điểm tín dụng dựa trên ‘ma trận tuần tự’.



Khung chi tiết của ‘mô hình đề xuất’.

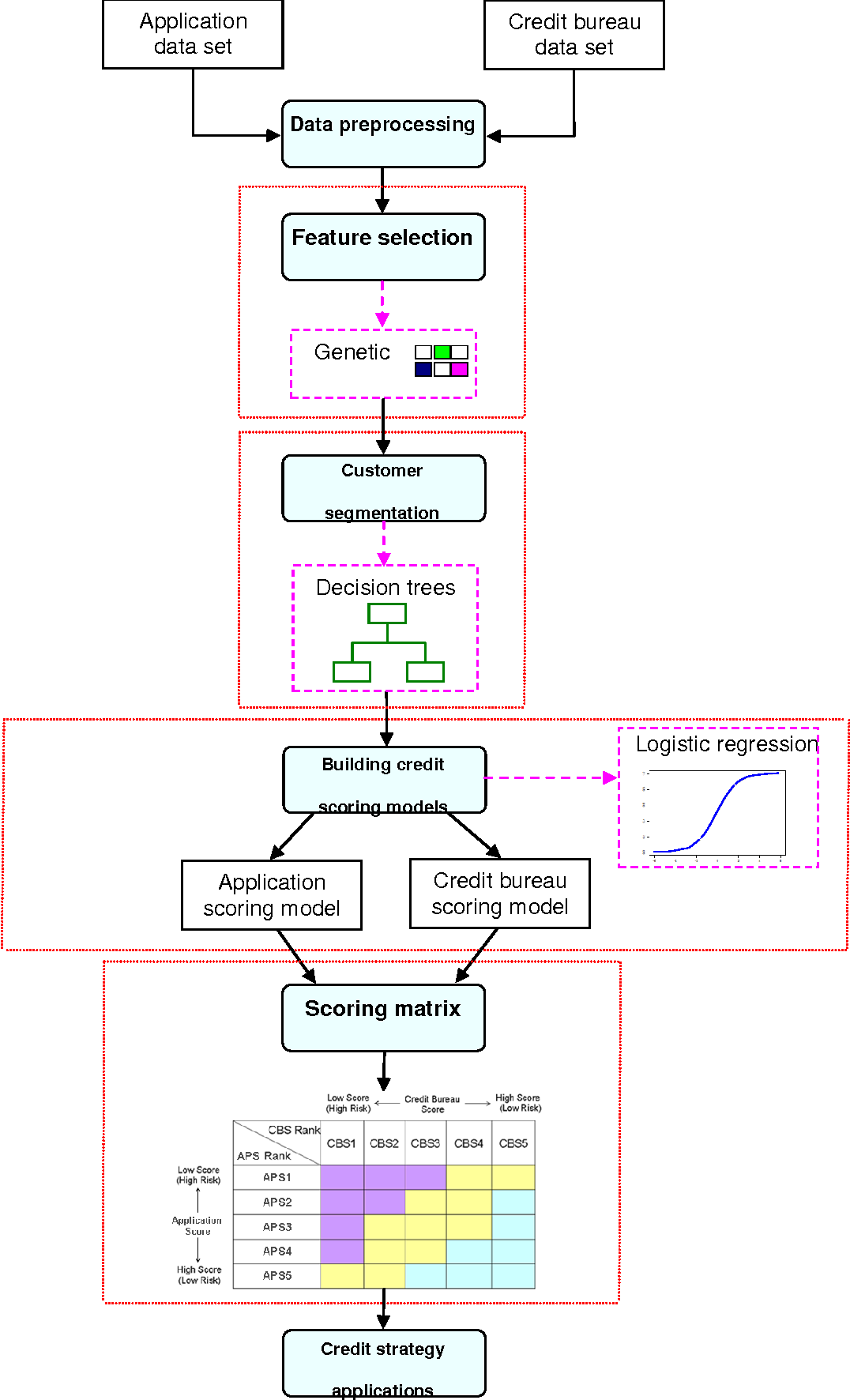
1. **Nâng cao hiệu quả hoạt động của mô hình chấm điểm tín dụng bằng ma trận chấm điểm kết hợp**

Trong thời gian gần đây, cuộc cạnh tranh trên thị trường tín dụng tiêu dùng ở Đài Loan đã trở nên khốc liệt hơn bao giờ hết. Đối mặt với tình hình này, hầu hết các tổ chức tài chính đang tích cực triển khai các biện pháp nhằm xây dựng mô hình chấm điểm tín dụng. Mô hình này dựa trên việc đánh giá quá trình phê duyệt tín dụng của khách hàng mới cũng như việc quản lý rủi ro tín dụng của khách hàng hiện tại.

Nghiên cứu này tiếp tục đi sâu vào việc áp dụng thuật toán di truyền, một phương pháp tiên tiến, để lựa chọn các đặc điểm quan trọng. Ngoài ra, nghiên cứu cũng sử dụng cây quyết định, một công cụ phân loại mạnh mẽ, để phân khúc khách hàng.

Bên cạnh đó, hồi quy logistic, một phương pháp thống kê phổ biến, được sử dụng để xây dựng mô hình tính điểm cho phòng ứng dụng và tín dụng. Trong quá trình này, hai mô hình tính điểm khác nhau được kết hợp để tạo ra ma trận tính điểm.

Ma trận tính điểm này sau đó được sử dụng trong quá trình đánh giá và phân đoạn rủi ro. Quá trình này giúp xác định thêm các phần cần được tăng cường quản lý hoặc kiểm soát. Kết quả cuối cùng là việc tạo ra một danh mục cho vay cá nhân hiệu quả và bền vững hơn.Kết quả phân tích chứng minh rằng khả năng dự đoán của ma trận chấm điểm vượt trội hơn cả mô hình chấm điểm ứng dụng và phòng tín dụng. Về giá trị K-S, ma trận chấm điểm tăng độ chính xác dự đoán so với mô hình chấm điểm ứng dụng và phòng tín dụng lần lượt là 18,40 và 5,70%. Về giá trị AUC, ma trận tính điểm tăng độ chính xác dự đoán so với mô hình tính điểm của phòng ứng dụng và tín dụng lần lượt là 10,90 và 6,40%. Hơn nữa, nghiên cứu này áp dụng ma trận tính điểm cho các quyết định phê duyệt tín dụng cho các nhóm rủi ro tương ứng nhằm tăng cường hoạt động quản lý rủi ro của ngân hàng. [13]

Dưới đây là quá trình phát triển ma trận tính điểm.

1. **Một mô hình tập hợp nhiều giai đoạn mới với nhiều mẫu lấy mẫu chọn lọc dựa trên K-means: Một ứng dụng trong chấm điểm tín dụng**

Với sự phát triển không ngừng nghỉ của học máy, nó đã mở ra cánh cửa cho việc cải tiến quy trình chấm điểm tín dụng, làm cho nó trở nên hiệu quả hơn. Học tập tổng thể, một trong những phương pháp học máy được công nhận rộng rãi, đã chứng minh những cải tiến đáng kể về độ chính xác dự đoán so với các mô hình học máy riêng lẻ trước đây được sử dụng để chấm điểm tín dụng. Điều này đã mở ra khả năng cải tiến đáng kể về hiệu quả trong việc đánh giá rủi ro tín dụng.

Nghiên cứu này tiếp tục tiến xa hơn bằng cách đề xuất một mô hình tập hợp nhiều giai đoạn mới, mô hình này sử dụng nhiều mẫu lấy mẫu chọn lọc. Sự đổi mới đến từ việc áp dụng thuật toán K-means, một phương pháp phân cụm dữ liệu mạnh mẽ, để chọn lọc mẫu. Sự kết hợp này tạo ra một quy trình chấm điểm tín dụng tinh vi hơn, có khả năng đánh giá rủi ro tín dụng một cách chính xác hơn, giúp các tổ chức tài chính đưa ra quyết định dựa trên dữ liệu chính xác hơn.

Thông qua việc kết hợp các phương pháp học máy hiện đại với các thuật toán phân tích dữ liệu truyền thống, mô hình này mở ra một cách tiếp cận mới trong việc chấm điểm tín dụng. Điều này không chỉ nâng cao khả năng dự đoán của hệ thống mà còn tạo ra một quy trình dễ dàng điều chỉnh và tùy chỉnh, giúp nó có thể được tối ưu hóa cho một loạt các ứng dụng khác nhau.

Đầu tiên, một phương pháp lấy mẫu dưới nhiều phương tiện K mới được đề xuất. Phương pháp này tiếp cận vấn đề dữ liệu mất cân bằng thông qua việc áp dụng thuật toán K-means, một phương pháp phân cụm dữ liệu mạnh mẽ, để chọn lọc mẫu. Điều này không chỉ giúp xử lý hiệu quả dữ liệu mất cân bằng, mà còn tạo nên sự cải tiến đáng kể trong việc chấm điểm tín dụng.

Sau đó, một cơ chế lấy mẫu chọn lọc mới cũng được đề xuất. Điểm nổi bật của cơ chế này là khả năng chọn lựa các bộ phân loại cơ sở hoạt động tốt hơn một cách thích ứng, giúp nâng cao khả năng xử lý và phân loại dữ liệu. Đây là một bước tiến rất quan trọng trong việc tối ưu hóa quy trình học máy, giúp cải thiện chất lượng và độ chính xác của mô hình.

Cuối cùng, một phương pháp xếp chồng nâng cao tính năng mới cũng được đề xuất. Đây chính là bước tiếp theo quan trọng trong việc xây dựng một mô hình tập hợp hiệu quả. Bằng cách tổng hợp các bộ phân loại cơ sở được đưa vào danh sách rút gọn, mô hình này không chỉ nâng cao hiệu suất mà còn cải thiện đáng kể độ chính xác trong việc chấm điểm tín dụng. Kết quả là một mô hình tập hợp mạnh mẽ và hiệu quả, có thể đáp ứng nhu cầu của nhiều tổ chức tài chính khác nhau.

Trong các thử nghiệm, bốn bộ dữ liệu với bốn chỉ số đánh giá được sử dụng để đánh giá hiệu quả của mô hình đề xuất và kết quả thử nghiệm chứng minh tính ưu việt của mô hình đề xuất so với các mô hình chuẩn khác.[14]

1. **Một mô hình kết hợp với phương pháp lựa chọn tính năng mới và phương thức bỏ phiếu nâng cao để chấm điểm tín dụng**

Trong thời đại công nghệ thứ tư với sự thống trị của trí tuệ nhân tạo, việc ứng dụng chuyên sâu công nghệ này vào lĩnh vực tài chính trở nên không thể thiếu. Cụ thể, các mô hình chấm điểm tín dụng được xây dựng bằng thuật toán machine learning đã trở thành xu hướng chủ đạo, mang lại sự minh bạch và chính xác hơn trong việc đánh giá khả năng trả nợ của khách hàng. Tuy nhiên, các đặc điểm thuộc tính phức tạp và có chiều cao của người đi vay đặt ra thách thức đối với khả năng dự đoán của mô hình. Bài viết này đề xuất một mô hình kết hợp với phương pháp lựa chọn tính năng mới và phương pháp bỏ phiếu nâng cao để chấm điểm tín dụng. Đầu tiên, một phương pháp kết hợp lựa chọn tính năng mới dựa trên thuật toán di truyền (FSCM-GA) được đề xuất, trong đó các bộ phân loại khác nhau được sử dụng để chọn các tính năng kết hợp với thuật toán di truyền và kết hợp chúng để tạo ra một tập hợp con tính năng tối ưu. Điều này giúp cải thiện tính linh hoạt và độ chính xác của mô hình đánh giá tín dụng. Hơn nữa, một phương pháp bỏ phiếu nâng cao (EVM) được đề xuất để tích hợp các bộ phân loại, nhằm mục đích cải thiện kết quả phân loại trong đó các giá trị xác suất dự đoán gần với ngưỡng. Cuối cùng, khả năng dự đoán của mô hình đề xuất đã được xác thực trên ba bộ dữ liệu công khai và năm chỉ số đánh giá (độ chính xác, AUC, điểm F, mất nhật ký và điểm Brier). Các kết quả thử nghiệm so sánh và kiểm tra ý nghĩa đã xác nhận tính hiệu quả và độ tin cậy tốt của mô hình đề xuất, đánh dấu một bước tiến quan trọng trong lĩnh vực này.

Lĩnh vực tài chính đã trải qua nhiều giai đoạn biến đổi và những thay đổi này đã đẩy mạnh việc ứng dụng trí tuệ nhân tạo. Thực tế, việc chấm điểm tín dụng, một khía cạnh quan trọng của hoạt động ngân hàng, đã được cải tiến đáng kể nhờ vào sự phát triển của thuật toán học máy. Một mô hình chấm điểm tín dụng không chỉ yêu cầu độ chính xác cao mà còn đòi hỏi khả năng xử lý dữ liệu phức tạp và đa chiều.

Để đáp ứng yêu cầu này, nghiên cứu này đã đề xuất một mô hình mới kết hợp với phương pháp lựa chọn tính năng mới và phương pháp bỏ phiếu nâng cao. Đặc biệt, một phương pháp kết hợp lựa chọn tính năng mới dựa trên thuật toán di truyền (FSCM-GA) đã được giới thiệu. Trong đó, các bộ phân loại khác nhau được sử dụng để chọn các tính năng kết hợp với thuật toán di truyền và kết hợp chúng lại với nhau để tạo ra một tập hợp con tính năng tối ưu. Điều này không chỉ giúp cải thiện khả năng xử lý dữ liệu phức tạp mà còn nâng cao độ chính xác của mô hình.

Ngoài ra, một phương pháp bỏ phiếu nâng cao (EVM) cũng đã được đề xuất. Mục đích của phương pháp này là để tích hợp các bộ phân loại, cải thiện kết quả phân loại mà trong đó các giá trị xác suất dự đoán gần với ngưỡng. Điều này giúp tăng cường khả năng dự đoán và độ tin cậy của mô hình.

Để đánh giá hiệu suất của mô hình, các thử nghiệm đã được tiến hành trên ba bộ dữ liệu công khai. Kết quả cho thấy rằng mô hình đã đạt được hiệu suất xuất sắc trên năm chỉ số đánh giá quan trọng: độ chính xác, AUC, điểm F, mất nhật ký và điểm Brier. Điều này đã xác nhận rằng mô hình có hiệu quả và đáng tin cậy, đánh dấu một bước tiến quan trọng trong lĩnh vực chấm điểm tín dụng.

Tóm lại, thông qua việc kết hợp các thuật toán tiên tiến, nghiên cứu này đã đóng góp vào việc cải tiến quy trình chấm điểm tín dụng, giúp các tổ chức tài chính đưa ra quyết định dựa trên dữ liệu chính xác và đáng tin cậy hơn..[15]

1. **Lựa chọn tính năng bao bọc để dự đoán mức độ nghèo dựa trên bộ dữ liệu thương mại điện tử**

Nghèo đói là tình trạng một cá nhân không thể đáp ứng được nhu cầu của mình, cả vật chất và phi vật chất. Cơ quan Thống kê Indonesia (BPS) sử dụng một phương pháp về nhu cầu cơ bản hoặc có thể gọi là nhu cầu về nhu cầu cơ bản để đo lường mức độ phúc lợi trong một khu vực. Khái niệm này sử dụng nhu cầu cơ bản của con người về lương thực và phi lương thực để đo lường mức độ nghèo đói trong khu vực thông qua cuộc khảo sát kinh tế xã hội quốc gia (SUSENAS). SUSENAS dự kiến sẽ tiêu tốn một lượng lớn thời gian và nguồn lực tài chính vì trong các hoạt động này, các cuộc phỏng vấn sẽ được thực hiện cho từng hộ gia đình bằng bảng câu hỏi. Ngày nay, việc thực hiện bảng câu hỏi, phỏng vấn và khảo sát không còn là cách duy nhất để thu thập dữ liệu. Người ta cũng có thể lấy dữ liệu từ bộ dữ liệu thương mại điện tử. Năm 2018, hiệp hội công đoàn gồm các nhà tư vấn quản lý McKinsey and Company đã công bố kết quả nghiên cứu của họ dưới dạng hiện trạng của thị trường thương mại điện tử Indonesia trong tương lai. Kết quả nghiên cứu cho thấy sự tăng trưởng về giá trị và tác động của tăng trưởng đến khía cạnh kinh tế xã hội của thị trường thương mại điện tử Indonesia hiện nay cho đến năm 2022. Nghiên cứu này đề xuất lựa chọn tính năng bao bọc kết hợp với thuật toán cây quyết định để dự đoán mức nghèo nhằm bổ sung cho khảo sát và điều tra dân số do BPS thực hiện.[16]

1. **Quan điểm thống kê về việc khám phá sự phụ thuộc chức năng trong dữ liệu nhiễu loạn**

Chúng tôi đã tiến hành một nghiên cứu sâu rộng về vấn đề phát hiện các phụ thuộc hàm (FD) từ một tập dữ liệu chứa nhiều nhiễu. Trong quá trình nghiên cứu, chúng tôi đã áp dụng các quan điểm thống kê và từ đó, chúng tôi đã rút ra được một mối liên hệ rõ ràng giữa việc khám phá FD và việc học cấu trúc trong các mô hình đồ họa xác suất.

Chúng tôi đã chỉ ra rằng việc khám phá các FD từ một tập dữ liệu ồn ào tương đương với việc tìm hiểu cấu trúc của mô hình qua các biến ngẫu nhiên nhị phân, trong đó mỗi biến ngẫu nhiên tương ứng với một hàm của các thuộc tính của tập dữ liệu. Điều này cung cấp một cái nhìn mới mẻ và mạnh mẽ về cách chúng tôi tiếp cận vấn đề.

Dựa trên những quan sát này, chúng tôi đã giới thiệu FDX, một khung làm việc đơn giản về mặt khái niệm, trong đó việc học các phụ thuộc hàm tương ứng với việc giải quyết vấn đề hồi quy thưa thớt. Khung làm việc này đã cho thấy sự hiệu quả đáng kể trong việc khám phá FD.

Chúng tôi đã kiểm chứng rằng FDX có thể khôi phục các phụ thuộc chức năng thực sự trên một loạt các tập dữ liệu tổng hợp và thế giới thực, ngay cả khi có dữ liệu nhiễu hoặc bị thiếu. Điều này cho thấy sự mạnh mẽ và linh hoạt của FDX.

Chúng tôi cũng nhận thấy rằng FDX có thể mở rộng quy mô thành các phiên bản dữ liệu lớn với hàng triệu bộ dữ liệu và hàng trăm thuộc tính trong khi nó mang lại mức cải thiện F1 trung bình gấp 2 lần so với các phương pháp khám phá FD hiện đại. Điều này cho thấy sự tiềm năng cao của FDX trong việc giải quyết các vấn đề liên quan đến dữ liệu lớn.

Chúng tôi đã tiến hành một nghiên cứu sâu rộng về vấn đề phát hiện các phụ thuộc hàm (FD) từ một tập dữ liệu chứa nhiều nhiễu. Trong quá trình nghiên cứu, chúng tôi đã áp dụng các quan điểm thống kê và từ đó, chúng tôi đã rút ra được một mối liên hệ rõ ràng giữa việc khám phá FD và việc học cấu trúc trong các mô hình đồ họa xác suất.

Chúng tôi đã chỉ ra rằng việc khám phá các FD từ một tập dữ liệu ồn ào tương đương với việc tìm hiểu cấu trúc của mô hình qua các biến ngẫu nhiên nhị phân, trong đó mỗi biến ngẫu nhiên tương ứng với một hàm của các thuộc tính của tập dữ liệu. Điều này cung cấp một cái nhìn mới mẻ và mạnh mẽ về cách chúng tôi tiếp cận vấn đề.

Dựa trên những quan sát này, chúng tôi đã giới thiệu FDX, một khung làm việc đơn giản về mặt khái niệm, trong đó việc học các phụ thuộc hàm tương ứng với việc giải quyết vấn đề hồi quy thưa thớt. Khung làm việc này đã cho thấy sự hiệu quả đáng kể trong việc khám phá FD.

Chúng tôi đã kiểm chứng rằng FDX có thể khôi phục các phụ thuộc chức năng thực sự trên một loạt các tập dữ liệu tổng hợp và thế giới thực, ngay cả khi có dữ liệu nhiễu hoặc bị thiếu. Điều này cho thấy sự mạnh mẽ và linh hoạt của FDX.

Chúng tôi cũng nhận thấy rằng FDX có thể mở rộng quy mô thành các phiên bản dữ liệu lớn với hàng triệu bộ dữ liệu và hàng trăm thuộc tính trong khi nó mang lại mức cải thiện F1 trung bình gấp 2 lần so với các phương pháp khám phá FD hiện đại. Điều này cho thấy sự tiềm năng cao của FDX trong việc giải quyết các vấn đề liên quan đến dữ liệu lớn.

Bên cạnh đó, chúng tôi cũng giới thiệu các kỹ thuật mới để tận dụng tối đa sức mạnh của FDX. Chẳng hạn, chúng tôi đã phát triển các phương pháp để giảm thiểu ảnh hưởng của nhiễu và dữ liệu bị thiếu, từ đó tăng cường khả năng khám phá FD của FDX.

Chúng tôi đang tiếp tục nghiên cứu để tìm hiểu thêm về cách tối ưu hóa FDX cho các tập dữ liệu cụ thể và cách đánh giá chất lượng của những FD mà nó khám phá. Chúng tôi tin rằng với những tiến bộ này, FDX sẽ trở thành một công cụ mạnh mẽ hơn nữa trong việc khám phá FD từ dữ liệu lớn.

Chúng tôi đã tiến hành một nghiên cứu sâu rộng về vấn đề phát hiện các phụ thuộc hàm (FD) từ một tập dữ liệu chứa nhiều nhiễu. Trong quá trình nghiên cứu, chúng tôi đã áp dụng các quan điểm thống kê và từ đó, chúng tôi đã rút ra được một mối liên hệ rõ ràng giữa việc khám phá FD và việc học cấu trúc trong các mô hình đồ họa xác suất.

Chúng tôi đã chỉ ra rằng việc khám phá các FD từ một tập dữ liệu ồn ào tương đương với việc tìm hiểu cấu trúc của mô hình qua các biến ngẫu nhiên nhị phân, trong đó mỗi biến ngẫu nhiên tương ứng với một hàm của các thuộc tính của tập dữ liệu. Điều này cung cấp một cái nhìn mới mẻ và mạnh mẽ về cách chúng tôi tiếp cận vấn đề.

Dựa trên những quan sát này, chúng tôi đã giới thiệu FDX, một khung làm việc đơn giản về mặt khái niệm, trong đó việc học các phụ thuộc hàm tương ứng với việc giải quyết vấn đề hồi quy thưa thớt. Khung làm việc này đã cho thấy sự hiệu quả đáng kể trong việc khám phá FD.

Chúng tôi đã kiểm chứng rằng FDX có thể khôi phục các phụ thuộc chức năng thực sự trên một loạt các tập dữ liệu tổng hợp và thế giới thực, ngay cả khi có dữ liệu nhiễu hoặc bị thiếu. Điều này cho thấy sự mạnh mẽ và linh hoạt của FDX.

Chúng tôi cũng nhận thấy rằng FDX có thể mở rộng quy mô thành các phiên bản dữ liệu lớn với hàng triệu bộ dữ liệu và hàng trăm thuộc tính trong khi nó mang lại mức cải thiện F1 trung bình gấp 2 lần so với các phương pháp khám phá FD hiện đại. Điều này cho thấy sự tiềm năng cao của FDX trong việc giải quyết các vấn đề liên quan đến dữ liệu lớn.

Chúng tôi cũng đã thảo luận về một số kỹ thuật mới nhằm tận dụng tối đa sức mạnh của FDX. Chẳng hạn, chúng tôi đã phát triển các phương pháp để giảm thiểu ảnh hưởng của nhiễu và dữ liệu bị thiếu, từ đó tăng cường khả năng khám phá FD của FDX.

Chúng tôi cũng nhấn mạnh việc tiếp tục nghiên cứu để tìm hiểu thêm về cách tối ưu hóa FDX cho các tập dữ liệu cụ thể và cách đánh giá chất lượng của những FD mà nó khám phá. Chúng tôi tin rằng với những tiến bộ này, FDX sẽ trở thành một công cụ mạnh mẽ hơn nữa trong việc khám phá FD từ dữ liệu lớn.

Cuối cùng, chúng tôi đã đề xuất một số hướng tiếp theo cho nghiên cứu. Điều này bao gồm việc khám phá các cách tiếp cận mới để tăng cường hiệu suất của FDX, cũng như việc thực hiện các nghiên cứu sâu hơn về các thuộc tính và cấu trúc dữ liệu mà FDX có thể khám phá. Chúng tôi tin rằng những nghiên cứu tiếp theo này sẽ giúp chúng tôi đưa FDX lên một tầm cao mới, không chỉ trong việc khám phá FD từ dữ liệu lớn, mà còn trong việc giúp hiểu rõ hơn về cấu trúc và mối liên hệ trong dữ liệu. [17]

1. **Ứng dụng xử lý dữ liệu dự đoán nghèo đói dựa trên thương mại điện tử bằng cách sử dụng cây quyết định và lựa chọn tính năng bao bọc**

Nghèo đói là một tình trạng khốn khổ mà trong đó một cá nhân không thể đáp ứng được các nhu cầu cơ bản của mình, không chỉ về vật chất mà còn về phi vật chất. Cơ quan Thống kê Trung ương đã áp dụng một phương pháp phức tạp nhưng cực kỳ hữu ích, dựa trên khái niệm về cách tiếp cận nhu cầu cơ bản - một khái niệm hay còn được gọi là khả năng đáp ứng các nhu cầu cơ bản, để đo lường mức độ nghèo đói trong một khu vực nhất định.

Phương pháp này tận dụng cách tiếp cận các nhu cầu cơ bản của con người, bao gồm nhu cầu lương thực và nhu cầu phi lương thực, để đo lường mức độ nghèo đói trong một khu vực cụ thể thông qua khảo sát kinh tế xã hội quốc gia (SUSENAS). Tuy nhiên, việc tiến hành SUSENAS đều gặp phải những khó khăn và trở ngại bởi quá trình này mất rất nhiều thời gian do yêu cầu phải phỏng vấn từng chủ hộ bằng bảng câu hỏi cụ thể, điều đó cần một khoản chi phí khá lớn và chủ hộ thường khó phát hiện, đôi khi thậm chí tránh xa vì sợ xảy ra lừa đảo.

Một giải pháp thay thế để hoàn thiện kết quả khảo sát và điều tra dân số của BPS là sử dụng cây quyết định học máy - một công cụ mạnh mẽ của trí tuệ nhân tạo, cùng với trình bao bọc lựa chọn tính năng được hỗ trợ bởi dữ liệu thương mại điện tử thu được từ một trong các công ty thương mại điện tử hàng đầu ở Indonesia. Cách tiếp cận này cho phép chúng ta có được dữ liệu dự đoán nghèo đói chính xác hơn.

Kết quả thực nghiệm đã thực hiện cho thấy rằng phương pháp này phù hợp với kết quả BPS nếu có nhiều đặc điểm. Tuy nhiên, nếu chỉ có 10 - 40 và 80-90 đặc điểm, kết quả không phù hợp.

Tóm tắt, Nghèo đói là tình trạng mà một người không thể đáp ứng được nhu cầu cơ bản của mình, bao gồm cả nhu cầu vật chất và phi vật chất. Cơ quan Thống kê Trung ương đã sử dụng các phương pháp tiếp cận dựa trên nhu cầu cơ bản để đo lường mức độ nghèo đói trong một khu vực nhất định. Điều này đòi hỏi phải tiến hành các cuộc khảo sát kinh tế xã hội quốc gia (SUSENAS), một quá trình tốn thời gian và tiền bạc. Tuy nhiên, việc sử dụng công nghệ học máy và dữ liệu thương mại điện tử có thể giúp cải tiến quá trình này và cung cấp dữ liệu dự đoán nghèo đói chính xác hơn. Từ khóa: Nghèo đói, BPS, máy học, cây quyết định, trình bao bọc. [18]

TÀI LIỆU THAM KHẢO

[1] S.V. Ramana and S.G. Krishna, “A study on the impact of fraud in Indian banking sector (with special reference on retail banking products)”, International Journal of Academic Research and Development, Vol.2, No. 6, pp.544-547, 2017.

[2] The Times of India Business, PTI, May 2, 2018. https://timesofindia.indiatimes.com/business/in dia-business/over-23000-bank-frauds-worth-rs1-lakh-crore-reported-in-5-yearsrbi/articleshow/63998429.cms

[3] A. Saravanan, M. S. Irfan Ahmed, and S. Sathya Bama, “A Survey on Exposed Vulnerabilities in Web Applications”, Asia Pacific Journal of Research, Vol.1, No. 35, pp.84-89, 2016.

[4] RBI Notification on Credit Risks. https://www.rbi.org.in/scripts/NotificationUser. aspx?Mode=0&Id=906

[5] W. Bouaguel, G. B. Mufti, and M. Limam, “Similarity Aggregation a New Version of Rank Aggregation Applied to Credit Scoring Case”, Mining Intelligence and Knowledge Exploration, Lecture Notes in Computer Science, Springer, Cham, Vol.8284, pp.618-628, 2013.

[6] S. Arya, C. Eckel, and C. Wichman, "Anatomy of the Credit Score", Journal of Economic Behavior & Organization, Vol.95, pp.175-185, 2013. 85 86 86 89 92 89 71 72 77 75 73 76 89 86 86 83 89 90 88 87 86 87 90 90 Existing Proposed Existing Proposed Existing Proposed Test Option 1 Test Option 2 Test Option 3 Accuracy in % Australian Credit German Credit Japanese Credit HMEQ Received: November 20, 2018 83 International Journal of Intelligent Engineering and Systems, Vol.12, No.2, 2019 DOI: 10.22266/ijies2019.0430.08

[7] X. Chen, C. Zhou, X. Wang, and Y. Li, “The Credit Scoring Model Based on Logistic-BPAdaBoost Algorithm and its Application in P2P Credit Platform”, In: Proc. of International Forum on Decision Sciences, pp. 119-130, 2017.

[8] M. Hurley and J. Adebayo, “Credit Scoring in the Era of Big Data”, Yale Journal of Law and Technology, Vol.18, 148, 2016.

[9] RBI Notification. RBI/2009-10/354 RPCD.CO RRB.BC No.62 /03.05.33/2009-10

[10] B. Baesens, C. Mues, M. De Backer, J. Vanthienen, and R. Setiono, “Building Intelligent Credit Scoring Systems using Decision Tables”, Enterprise Information Systems V, pp.131-137, 2004.

[11] T. Howley, M.G. Madden, ML. O’Connell, and A.G. Ryder, “The Effect of Principal Component Analysis on Machine Learning Accuracy with High-Dimensional Spectral Data”, Applications and Innovations in Intelligent Systems XIII, pp.209-222, 2006.

[12] M.S.Irfan Ahmed, P.Ramila Rajaleximi Published 25 October 2019 ‘Business, Computer Science, Economics International Journal of Scientific & Technology Research’

[13] Department of Information Management, National Taiwan University of Science and Technology, No. 43, Sec. 4, Keelung Road, Taipei 10607, Taiwan.

[14] Authors: Jin, Yilun | Liu, Yanan; \* | Zhang, Wenyu | Zhang, Shuai | Lou, Yu from School of Information Management and Artificial Intelligence, Zhejiang University of Finance and Economics, Hangzhou, China

[15] Yao, J., Wang, Z., Wang, L., Zhang, Z., Jiang, H., & Yan, S. (2021). A hybrid model with novel feature selection method and enhanced voting method for credit scoring. J. Intell. Fuzzy Syst., 42, 2565-2579.

[16] Pangestu, A.D., Wijaya, D.R., Hernawati, E., & Hidayat, W.Y. (2020). Wrapper Feature Selection for Poverty Level Prediction Based on E-Commerce Dataset. 2020 International Conference on Data Science and Its Applications (ICoDSA), 1-7.

[17] Zhang, Y., Guo, Z., & Rekatsinas, T. (2020). A Statistical Perspective on Discovering Functional Dependencies in Noisy Data. Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data.

[18] Pangestu, A.G., Wijaya, D.R., & Hernawati, E. (2020). Aplikasi Pengolahan Data Prediksi Kemiskinan Berbasis E-commerce Menggunakan Decision Tree Dan Wrapper Feature Selection.