Một số kĩ thuật học máy cho chấm điểm tín dụng

Giang Thị Thu Huyền

Học viện Ngân hàng

Ngày nhận: 23/10/2020 Ngày nhận bản sửa: 29/01/2021 Ngày duyệt đăng: 23/03/2021

Tóm tắt: Chấm điểm tín dụng là một giai đoạn quan trọng trong quy trình quản lý rủi ro của các tổ chức tài chính và ngân hàng. Chấm điểm tín dụng tốt sẽ góp phần làm cho chất lượng cho vay tốt hơn. Chất lượng cho vay là yếu tố quyết định hàng đầu đến sự cạnh tranh, tồn tại và lợi nhuận của các tổ chức tài chính ngân hàng. Các kỹ thuật học máy đã góp phần cải tiến đáng kể về hiệu suất phân loại trong chấm điểm tín dụng, đồng thời giảm thiểu rủi ro cho đánh giá tín dụng. Bài báo đưa ra một số kỹ thuật học máy cho chấm điểm tín dụng đã và đang được các tổ chức tài chính và ngân hàng sử dụng; đưa ra kết quả thử nghiệm các kỹ thuật học máy trên bộ dữ liệu Kaggle- bộ dữ liệu miễn phí cho nghiên cứu về khoa học dữ liệu, từ đó đánh giá hiệu suất phân loại của các kỹ thuật học máy này. Qua đánh giá thì các thuật toán về học sâu cho kết quả khá tốt và cần tiếp tục nghiên cứu trong tương lai. Từ khóa: Kỹ thuật học máy, Chấm điểm tín dụng, Học sâu.

1. Giới thiệu

Xếp hạng tín dụng của khách hàng là nội

dung quan trọng trong quản lý rủi ro đối với hệ thống các ngân hàng thương mại. Kết quả xếp hạng tín dụng là một yếu tố quan

Some machine learning techniques for credit scoring

Abstract: Credit scoring is an important stage in the risk management process of financial institutions and banks. A good credit score will contribute to a better loan quality. Loan quality is the decisive aspect of the competition, existence and profitability of banks and financial institutions. Machine learning techniques have contributed to a significant improvement in the grading performance in credit scoring while also minimizing the risk of credit rating. The paper outlines some machine learning techniques for credit scoring that have been used by financial terms. The test results of machine learning techniques are obtained based on Kaggle datasets, the free datasets for data science research, which evaluate the classification efficiency of these machine learning techniques and suggest some recommendations. The deep learning algorithms give good results to study in the future. **Keywords**: Machine learning techniques, Credit scoring, Deep learning.

Huyen Thi Thu Giang

Email: huyengtt@hvnh.edu.vn Banking Academy of Vietnam trong cần xem xét trong việc ra quyết định cho vay. Do đó, hệ thống xếp hang tín dung là một công cụ quan trọng để tăng cường tính khách quan, nâng cao chất lượng và hiệu quả hoạt động tín dụng. Hiện nay, các ngân hàng thương mại đã và đang áp dụng nhiều kĩ thuật khác nhau để xếp hang tín dung khách hàng. Mô hình tính điểm tín dụng là phương pháp lượng hoá mức độ rủi ro, thông qua đánh giá thang điểm, các chỉ tiêu đánh giá trong những mô hình chấm điểm được áp dụng khác nhau đối với từng loại khách hàng. Cùng với sự phát triển của khoa học máy tính và các kĩ thuật khai phá dữ liêu, nhiều phương pháp khai phá dữ liêu đã được đưa vào sử dung trong thực tế như cây quyết đinh và các kĩ thuật phân lớp, phân cum, học sâu... Theo Amitha Mathew (2021), kỹ thuật học sâu đang thực hiện tốt hơn các kỹ thuật học máy hiện tại đặc biệt là với dữ liệu phi cấu trúc.

Hand và Jacka (1998) đã nói rằng "Quá trình (bởi các tổ chức tài chính) mô hình hóa mức độ tín nhiệm được gọi là chấm điểm tín dụng". Quá trình này bao gồm thu thập, phân tích và phân loại các yếu tố và biến số tín dung khác nhau để đánh giá các quyết định tín dụng. Trong các mô hình chấm điểm tín dụng truyền thống được sử dụng, người vay tiềm năng phải có đủ thông tin tín dụng. Cách làm này phụ thuộc rất nhiều vào quan điểm chủ quan của nhân viên tín dung, do đó các ngân hàng phải đối mặt với nhiều rủi ro bởi trình đô thẩm đinh của nhân viên còn han chế, nhân viên có thể thông đồng với khách hàng để nâng hạng tín dụng... Trong trường hợp không có thông tin này, điểm tín dụng không thể tính toán được, và một người đi vay có khả năng đáng tin cậy thường khó có khả năng tiếp cận tín dụng và xây dựng một lịch sử tín dụng với ngân hàng. Với việc sử dụng các nguồn dữ liêu thay thế và ứng dung các thuật toán học máy để giúp phát triển khả năng đánh giá và sư sẵn lòng trả nơ, người cho vay có thể đưa ra các quyết đinh tín dụng trước đây không thể.

Nhu cầu chấm điểm tín dụng mà thực tế đặt ra đã khiến các nhà nghiên cứu phát triển các mô hình chấm điểm tín dụng, từ đó giúp người cho vay quyết đinh cấp hay từ chối tín dung cho người nôp đơn. Công cu chấm điểm tín dung sử dung học máy được thiết kế để tăng tốc các quyết định cho vay, đồng thời có khả năng hạn chế rủi ro. Áp dụng thuật toán học máy đã cho phép phân khúc chất lượng của người đi vay lớn hơn, nhanh hơn và rẻ hơn, giúp tạo điều kiện tiếp cân tín dung tốt hơn. Muc tiêu của mô hình chấm điểm tín dung là phân loại người nôp đơn tín dụng thành hai hạng: hạng "tín dụng tốt" có trách nhiệm bồi hoàn nghĩa vu tài chính và hạng "tín dụng xấu" sẽ bị từ chối tín dụng do khả năng võ nợ cao. Việc phân loại này phụ thuộc vào đặc điểm xã hội học của người đi vay (như tuổi, trình độ học vấn, nghề nghiệp và thu nhập), tình hình trả nơ các khoản vay trước đó và loai khoản vay... Môt số kỹ thuật học máy cho chấm điểm tín dung như: mang nơ ron nhân tao, K-Nearest neighbour, support vector machine - SVM, cây quyết định... Theo Amitha Mathew (2021), kỹ thuật học sâu dựa trên mạng nơ ron sâu và ứng dụng của nó đang là hướng nghiên cứu được nhiều người quan tâm và thu được những kết quả khả quan. Kỹ thuật học sâu đã giải quyết được một số vướng mắc mà các thuật toán hiện nay gặp phải đó là xử lý với dũ liệu phi cấu trúc.

Bài báo giới thiệu một số thuật toán học máy và thử nghiệm, so sánh, đánh giá các thuật toán trên bộ dữ liệu Kaggle và từ đó đưa ra kết luân.

2. Các kỹ thuật học máy cho chấm điểm tín dụng

Một vấn đề lớn đối với các tổ chức tài

chính và ngân hàng là làm thể nào để xác đinh được tín dung xấu, vì tín dung xấu có thể gây ra các vấn đề nghiêm trọng trong tương lai. Điều này dẫn đến thất thoát vốn, giảm doanh thu và làm tăng tổn thất của ngân hàng, có thể dẫn đến mất khả năng thanh toán hoặc phá sản. Hiện nay, có rất nhiều kĩ thuật học máy để phân loại khách hàng. Bài báo sẽ đưa ra một số kĩ thuật hay được sử dung:

2.1. K láng giềng gần nhất (K-Nearest Neighbors - KNN)

Theo Ram Babu, Mr.A.Rama Satish (2013) thì thuật toán K láng giềng gần nhất là một kĩ thuật học có giám sát (supervised learning) dùng để phân loai đổi tương mới bằng cách tìm điểm tương đồng giữa quan sát mới với dữ liệu sẵn có. Một đối tượng được phân lớp dựa vào K láng giềng của nó. K là số nguyên dương được xác định trước khi thực hiện thuật toán. Người ta thường dùng khoảng cách Euclidean để tính khoảng cách giữa các đối tương.

Thuật toán KNN được mô tả như sau:

- Xác đinh giá tri tham số K (số láng giềng gần nhất).
- Tính khoảng cách giữa đối tượng cần phân lớp với tất cả các đối tượng trong tập huấn luyên (thường sử dung khoảng các Euclidean).
- Sắp xếp khoảng cách theo thứ tư tăng dần và xác đinh K láng giềng gần nhất với đối tương cần phân lớp.
- Lấy tất cả các lớp của K láng giềng gần nhất đã xác đinh.
- Dựa vào phần lớn lớp của láng giềng gần nhất để xác định lớp cho đối tượng cần phân lớp.

KNN là một mô hình đơn giản và trực quan nhưng vẫn có hiệu quả cao vì nó không tham số; mô hình không đưa ra giả đinh nào về việc phân phối dữ liệu. Hơn nữa, nó có thể được sử dụng trực tiếp để phân loại đa lớp.

2.2. Cây quyết định

Theo Joao A. Bastos (2008), Cây quyết định là kỹ thuật phân loại khác được sử dụng để phát triển các mô hình tính điểm tín dung. Cây quyết đinh là một trong những thuật toán máy học phổ biến nhất hiện nay. Với kỹ thuật phân lớp dựa trên cây quyết định, kết quả của quá trình xây dựng mô hình sẽ cho ra một cây quyết định. Thuật toán bắt đầu với một nút gốc chứa một mẫu người nôp đơn tín dung tốt và xấu. Sau đó, thuật toán lặp lai tất cả các phép phân tách có thể có để tìm thuộc tính và giá tri giới han tương ứng mang lai sư phân tách tốt nhất thành một bên có phần lớn là tín dung tốt và bên kia chủ yếu là tín dụng xấu. Cây quyết định sẽ sinh ra các luật để dự đoán lớp của các đối tượng chưa biết. Cây quyết định là cây mà mỗi nút biểu diễn một đặc trưng (tính chất), mỗi nhánh (branch) biểu diễn một quy luật (rule) và mỗi lá biểu biển một kết quả (giá tri cu thể hay một nhánh tiếp tuc). Có rất nhiều thuật toán cây quyết định như ID3, J48, C4.5, CART (Classification and Regression Tree)... CART là một kĩ thuật học máy được áp dụng để dự đoán. CART dùng Gini Index để kiểm tra. Chỉ số Gini là thước đo sư phân tán hoặc tính đa dang trong môt nút.

2.3. Naive Bayes

Theo Olatunji J. Okesola (2017), Thuật toán Naive Bayes (NB) là một thuật toán dựa trên đinh lý Bayes về lý thuyết xác suất để đưa ra các phán đoán cũng như phân loại dữ liệu dựa trên các dữ liệu được quan sát và thống kê, được ứng dụng để đưa các dự đoán có đô chính xác cao, dưa trên một tập dữ liệu đã được thu thập. Thuật toán Naive Bayes thuộc vào nhóm học máy có giám sát.

2.4. Máy hỗ trợ vector- Support Vector Machine (SVM)

Theo R.H. Davis, D.B. Edelman, và A.J. Gammerman (1992), Máy hỗ trợ vector-Support Vector Machine (SVM) là một thuật toán học máy có giám sát được sử dụng phô biến cho các bài toán phân lớp. SVM được Vapnik giới thiệu lần đầu tiên và sau đó trở nên phổ biến. Ý tưởng của SVM là tìm một siêu phẳng (hyper lane) đóng vai trò là ranh giới quyết định, để tách hai lớp khác nhau. Vân đề là có rất nhiều siêu phẳng, chúng ta phải chọn cái nào để tối ưu nhất? Siêu phẳng tối ưu mà chúng ta cần chon là siêu phẳng phân tách có lề lớn nhất. Lý thuyết học máy đã chỉ ra rằng một siêu phẳng như vậy sẽ cực tiểu hóa giới hạn lỗi mắc phải.

2.5. Kỹ thuật học sâu (Deep learning)

Theo Cuicui Luo, Desheng Wu, Dexiang Wu (2017) khái niệm học sâu được phát triển lần đầu vào những năm 1960. Đến năm 2011 và 2012, thuật toán học sâu được thúc đây bởi sự gia tăng mạnh mẽ trong sức mạnh tính toán của máy tính hiện đại và đạt được nhiều thành tựu trong nhiều lĩnh vực trong đó có tài chính ngân hàng. Học sâu cho phép một lượng lớn dữ liệu được phân tích rất nhanh chóng, nó có thể xử lý phạm vi đầu vào tín dung rông hơn, giảm chi phí đánh giá rủi ro tín dung.

Kỹ thuật học sâu là một thuật toán học máy được xây dựng dựa trên một số ý tưởng mô phỏng hệ thông não bộ của con người. Nó biểu diễn dữ liệu thông qua nhiều tầng từ cụ thể đến trừu tượng qua đó trích rút được các đặc trưng có ý nghĩa. Theo Amitha Mathew (2021) Học sâu (deep learning) còn được gọi là học có cấu trúc sâu hoặc học phân cấp, là một phần của học máy dựa trên các mạng nơ ron nhân tạo. Học sâu là một tập hợp các thuật toán để mô hình dữ liệu trừu tượng hóa ở mức cao bằng cách sử dụng nhiều lớp xử lý với cấu trúc phức tạp, hoặc bằng cách khác bao gồm nhiều biến đổi phi tuyến. Theo Cuicui Luo (2017) thì có nhiều kỹ thuật học sâu Deep learning cho chấm điểm tín dụng, bài báo sẽ đưa ra bốn kỹ thuật học sâu cho chấm điểm tín dung:

2.5.1. Mạng nơ ron nhiều lớp (Deep Sequential Neural Network)

Mạng nơ-ron nhiều lớp là một kiểu đơn giản của mô hình Deep learning với các mạng nơ-ron chứa lớp liên tiếp xếp chồng nhau. Trong thực nghiệm, để nâng cao hiệu quả của mô hình, ta thiết kế các tầng mang với mật đô nơ-ron khác nhau cho mỗi tầng. Sử dụng hàm kích hoạt (activation function) sau rectified linear (ReLU): $\theta(x)$ = max(0, x); hyperbolic tangent (tanh): $\theta(x) = \sinh(x) \div \cosh(x) = (e^x - e^{-x}) \div (e^x$ $+ e^{-x}$; sigmoid: $\theta(x) = (1 + e^{-x})^{-1}$. Để đánh giá mô hình một cách khách quan, sử dụng Hàm loss cross-entropy.

$$L(\theta) = -\frac{1}{n} \sum_{i=1}^{n} [y_i \log(p_i) + (1 - p_i) \log(1 - p_i)]$$

$$= -\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij} \log(p_{ij})$$

Trong đó i là chỉ số của các quan sát, j là các lốp, y là nhãn và $p_{ij} \in (0,1)$: $\sum_{j} p_{ij} = 1$, $\forall i, j$ là giá trị dự đoán.

Kiến trúc của mô hình được mô tả chi tiết dưới đây (áp dụng cho dữ liệu Kaggle Credit):

Thuật toán được mô tả như sau:

- 1. Input (length = 60)
- 2. Dense layer (60, input $\dim = 10$, kernel initializer = 'uniform', activation function ='relu')
- 3. Dense layer (5, kernel initializer = 'uniform', activation function ='relu')
- 4. Dense layer (1, kernel initializer = 'uniform', activation function = 'sigmoid')

2.5.2. Mang no ron tích chập (Convolutional Neural Network)

Mạng nơ ron tích chập là một mạng nơ ron nhân tạo với các toán tử tích chập. Nó có khả năng học một lượng lớn các dữ liệu trong khoảng thời gian ngắn hơn nhiều so với mang nơ ron thông thường. Lý do là nó sử dung ít trong số hơn trong khi đô chính xác chỉ kém hơn một phần nhỏ so với kiến trúc truyền thống.

Thuật toán được mô tả như sau:

- 1. Input (length = 14, shape(14,1))
- 2. 1D Convolution layer (filters = 150, kernel size=1)
- 3. Max pooling layer (pool size = 1)
- 4. Fully connected (N = 14, activation)function = 'relu')
- 5. Fully connect (N = 1, activation function)= 'sigmoid')

2.5.3. Mang no ron hồi quy (Recurrent Neural Network)

Ý tưởng chính của mang nơ ron hồi quy là sử dung chuỗi các thông tin. Trong các mang nơ-ron truyền thống tất cả các đầu vào và đầu ra là độc lập với nhau, không liên kết thành chuỗi. Mang nơ ron hồi quy được gọi là hồi quy bởi lẽ chúng thực hiện cùng một tác vụ cho tất cả các phần tử của một chuỗi với đầu ra phụ thuộc vào cả các phép tính trước đó. Nói cách khác mạng nơ ron hồi quy có khả năng nhớ các thông tin được tính toán trước đó.

Thuật toán được mô tả như sau:

- 1. Input (length = 24)
- 2. RNN = SimpleRNN(N = 100, return)sequences = True, dropout = self.dropout, activation function = 'sigmoid')
- 3. Dropout(0.5)
- 4. Full connected (N = 1, activation function)= 'sigmoid')

2.5.4. Mạng bộ nhớ ngắn-dài hạn (Long Short-Term Memory)

Mô hình mạng bộ nhớ ngắn-dài hạn không khác mô hình truyền thống của mạng nơ ron hồi quy, chúng chỉ khác nhau ở cách tính toán của các nút ẩn, mạng bộ nhớ ngắn-dài hạn sử dụng hàm tính toán khác ở các trạng thái ẩn. Bộ nhớ của mạng bộ nhớ ngắn-dài han được gọi là tế bào (Cell) và ban có thể tưởng tương rằng chúng là các hộp đen nhân đầu vào là trang thái phía trước và đầu vào hiện tại. Bên trong hộp đen này sẽ tự quyết định cái gì cần phải nhớ và cái gì sẽ xoá đi. Sau đó, chúng sẽ kết hợp với trạng thái phía trước, nhớ hiện tại và đầu vào hiện tại. Vì vậy mà ta có thể truy xuất được quan hệ phu thuộc xa nhau rất hiệu quả.

Thuật toán được mô tả như sau:

- 1. Input (length = 24)
- 2. LSTM(N = 24, activation function = 'sigmoid', loss = 'binary cross entropy', optimizer = 'rmsprop')
- 3. Dropout(0.5)
- 4. Full connected(N = 1,activation code = 'sigmoid')

3. Kết quả thử nghiệm các kỹ thuật học máy

Bài báo sử dụng bộ dữ liệu Kaggle đánh giá hiệu quả các kỹ thuật học máy. Kaggle có nhiều bộ dữ liệu khác nhau cho các lĩnh vưc nhằm hỗ trơ cho nghiên cứu về học máy và khoa học dữ liêu. Kaggle đã được các nhà nghiên cứu trên thế giới sử dung rông rãi. Kaggle credit là một bộ dữ liệu tốt được sử dụng cho các cuộc thi và các nghiên cứu về chấm điểm tín dụng. Bộ dữ liệu này bao gồm thông tin của 250.000 người vay vốn với 150.000 bản ghi có nhãn và 100.000 trường hợp không có nhãn. Các bản ghi được gán nhãn tương ứng với "tín dụng xấu" hoặc "tín dụng tốt".

Thực nghiệm trên bô dữ liêu Kaggle và sử dụng phương pháp đánh giá chéo (10-fold cross validation), lấy kết quả trung bình để

so sánh hiệu quả các thuật toán K-Nearest neighbor (kNN), cây hồi quy (CART), Naive Bayes (NB) và máy hỗ trợ vector-Support Vector Machine (SVM), các thuật toán học sâu (mang nơ ron nhiều lớp (Deep Sequential Neural Network- DSNN), mang no ron tích châp (Convolutional Neural Network- CNN), mang no ron hồi quy (Recurrent Neural Network- RNN), mạng bộ nhớ ngắn-dài hạn (Long Short-Term Memory-LSTM). Đánh giá hiệu quả của các thuật toán dựa trên 4 độ đo sau đây: Accuracy, Precision, Recall và F1-score.

 $TP = s\hat{o}$ lương đúng tích cực

 $FP = s\hat{o}$ lương sai tích cực

 $TN = s\delta$ lương đúng tiêu cực

 $FN = s\delta$ lương sai tiêu cực

 $P = s\delta$ lương bản ghi lớp tích cực thực tế $N = s\delta$ lượng bản ghi lớp tiêu cực thực tế Kết quả về độ chính xác của các thuật toán được thể hiện trong Bảng 2.

Kết quả tổng hợp đối với bộ dữ liệu Kaggle

credit khi so sánh trên đô đo accuracy, các thuật toán đều có được kết quả rất tốt trên 90% (trừ CART là 89,2%). Khi so sánh trên độ đo Precision, Recall và F1 thì độ chính xác của các các thuật toán học sâu tốt hơn so với các thuật toán truyền thống. Kết quả tốt nhất đat được với đô đo F1-socre là 51,12%, đô đo Precision là 53,20%, đô đo Recall là 49,20% do mang RNN đem lai. Nguyên nhân là do khả năng học của các thuật toán học sâu tốt khi được điều chỉnh số lớp và số nơ ron ở mỗi lớp; mặt khác, bộ dữ liệu Kaggle credit có tính không cân bằng và các thuật toán khá nhay cảm với dữ liêu không cân bằng. Với bô dữ liêu Kaggle credit thì các thuật toán sử dung kỹ thuật học sâu cho kết quả phân loại khá tốt so với các thuật toán còn lai.

4. Kết luận

Các thuật toán chấm điểm tín dụng là công

Bảng 1. Mô tả các biến trong bộ dữ liệu Kaggle Credit

Thuộc tính	Mô tả
SeriousDlqin2yrs	Người đã trải qua 90 ngày quá hạn
RevolvingUtilizationOf_ UnsecuredLines	Tổng số dư trên thẻ tín dụng và hạn mức tín dụng cá nhân ngoại trừ bất động sản và không có nợ trả góp như khoản vay mua ô tô chia cho tổng hạn mức tín dụng
Age	Tuổi của người vay (năm)
NumberOfTime30- 59DaysPastDueNotWorse	Số lần người vay quá hạn 30-59 ngày trong 2 năm gần đây
DebtRatio	Tổng mức chi tiêu hàng tháng chia cho tổng mức thu nhập hàng tháng
MonthlyIncome	Thu nhập hàng tháng
NumberOfOpenCredit_ LinesAndLoans	Số lượng khoản vay mở (trả góp như vay mua ô tô hoặc thế chấp) và Dòng tín dụng (ví dụ: thẻ tín dụng)
NumberOfTimes90DaysLate	Số lần người vay quá hạn 90 ngày hoặc hơn.
NumberRealEstateLoans_ OrLines	Số lượng các khoản vay thế chấp và bất động sản bao gồm cả dòng vốn chủ sở hữu của tín dụng.
NumberOfTime60- 89DaysPastDueNotWorse	Số lần người vay quá hạn 60-89 ngày trong 2 năm gần đây.
NumberOfDependents	Số lượng người phụ thuộc (trẻ em, vợ/chồng).

Nguồn https://www.kaggle.com/

Bang 2. Độ chính xác của các thuật toàn										
	kNN	CART	NB	SVM	DSNN	CNN	RNN	LSTM		
Accuracy	0,9287	0,8920	0,9288	0,9301	0,9342	0,9277	0,9350	0,9360		
Precision	0,3031	0,2515	0,3585	0,3670	0,5200	0,3682	0,5320	0,5156		
Recall	0,0172	0,2786	0,0234	0,2840	0,4870	0,3745	0,4920	0,503		
F1	0,0325	0,2644	0,0438	0,3205	0,5029	0,3713	0,5112	0,509		

Rảng 2 Độ chính vác của các thuật toán

Nguồn: Tác giả chạy thử nghiệm và tổng hợp kết quả

cụ quan trọng cho các vấn đề dự đoán và phân loại. Bài báo đã giới thiệu một số kĩ thuật chấm điểm tín dụng với bộ dữ liệu Kaggle credit. Trên cơ sở những kết quả thu được, cho thấy kĩ thuật học sâu có thể là giải pháp cho các ứng dung chấm điểm tín dung. Các kỹ thuật hỗ trơ người ra quyết đinh tín dung dư đoán khách hàng sẽ có lịch sử tín dụng tốt hay tín dụng xấu dựa trên các thuộc tính và thông tin "tín dụng" của họ. Tuy nhiên, cần nhấn mạnh rằng không có mô hình tính điểm tín dụng lý tưởng cũng như không có kỹ thuật tốt nhất được sử dung để xây dưng mô hình chấm điểm tín dung; và kỹ thuật tốt cho tập dữ liêu này chưa chắc tốt trên tập dữ liêu khác. Kỹ thuật này hơn kỹ thuật khác khả năng dự đoán trong các trường hợp cụ thể khác nhau. "Kĩ thuật tốt nhất" phụ thuộc

vào từng yêu cầu cụ thể, cấu trúc của dữ liệu, các chức năng của ứng dụng, mức độ tách biệt các lớp bằng mục tiêu phân loại. Ngoài ra, việc sử dụng kết hợp các kỹ thuật hứa hẹn cho khả năng phân loại và dự đoán tốt hơn cũng như cần phải xem xét đến ảnh hưởng quan trọng của các chính sách và quy đinh của các tổ chức tài chính ngân hàng. Bài nghiên cứu mới chỉ thử nghiệm trên bộ dữ liệu dành cho nghiên cứu của Kaggle. Trên cơ sở những kết quả thu được, có thể kết luận rằng kĩ thuật học sâu có thể là giải pháp khá quan trọng cho các ứng dung chấm điểm tín dung hiện nay và trong tương lai. ■

Tài liệu tham khảo

Amitha Mathew, P.Amudha, S.Sivakumari. (2021), Deep Learning Techniques: An Overview, Advanced Machine Learning Technologies and Applications.

Cuicui Luo, Desheng Wu, Dexiang Wu. (2017), A deep learning approach for credit scoring using credit default swaps, Engineering Applications of Artificial Intelligence.

Denoyer, Gallinari. (2011), Deep sequential neural network, University Pierre et Marie Curie - Paris, France.

Hand, Jacka. (1998) Statistics in Finance, Arnold Applications of Statistics: London.

Hussein A. Abdou, John Pointon. (2011), Credit scoring, statistical techniques and evaluation criteria: A review of the literature, Intelligent Systems in Accounting, Finance & Management.

Joao A. Bastos. (2008), Credit scoring with boosted decision trees, School of Economics and Management (ISEG) Technical University of Lisbon, Portugal.

Olatunji J. Okesola et al. (2017), An improved Bank Credit Scoring Model A Naïve Bayesian Approach, International Conference on Computational Science and Computational Intelligence.

Ram Babu, Mr.A.Rama Satish. (2013), Improved of K-Nearest Neighbor Techniques in Credit Scoring, International Journal For Development of Computer Science & Technology.

R.H. Davis, D.B. Edelman, A.J. Gammerman. (1992), Machine learning algorithms for credit-card applications. IMA Journal of Management Mathematics.

https://www.hindawi.com/journals/aor/2019/1974794/; https://www.kaggle.com/