# **Credit score classification - Phân loại điểm tín dụng**

1. Giới thiệu
2. Điểm tín dụng là gì

Là con số thể hiện lịch sử tín dụng của một cá nhân theo tiêu chuẩn và quy tắc xếp hạng quốc gia dựa trên việc phân tích hồ sơ tín dụng của cá nhân nào đó và số điểm này sẽ được trung tâm thông tin tín dụng – CIC quản lý, đây là trung tâm thuộc ngân hàng nhà nước.

Hiểu đơn giản, điểm tín dụng là điểm các ngân hàng, tổ chức tài chính sử dụng để đánh giá sự uy tín của khách hàng khi họ sử dụng các dịch vụ tài chính hay các hình thức cho vay mà tổ chức đó cung cấp.

Điểm số càng cao thì khách hàng sẽ thuận lợi hơn trong quá trình vay, điểm càng thấp thì các tổ chức sẽ hạn chế tiếp nhận hồ sơ của người đi vay. Mức điểm 740 là mức điểm tuyệt vời để khách hàng có thể sử dụng dịch vụ vay của ngân hàng – tài chính với lãi suất tốt.

Rủi ro của người cho vay khi quay trở lại từ phân tích rủi ro tín dụng liên quan đến đầu tư của họ, chủ yếu là do người vay không trả lại các khoản vay, mà quá trình tính toán là đánh giá rủi ro tín dụng, cái gọi là *chấm điểm tín dụng*  [Moscato2021ABO].

1. Mục đích của điểm tín dụng

Hỗ trợ chính xác các tổ chức tài chính trong việc xác định chính sách ngân hàng và chiến lược thương mại của họ. Một mặt, để giảm rủi ro tài chính và mặt khác, để tăng lợi nhuận liên quan [Moscato2021ABO].

Điểm tín dụng sẽ rất có lợi để bạn có thể đi vay ở bất cứ tổ chức tín dụng – ngân hàng nào. Lãi suất mà bạn nhận được, ngân hàng sẽ ấn định nó dựa trên khung điểm tín dụng mà khách hàng có. Đó là lý do những cá nhân thường giao dịch với ngân hàng sẽ có được mức lãi suất khi vay cực kỳ cao. Hiện nay, điểm tín dụng không chỉ được sử dụng tại ngân hàng mà các công ty bảo hiểm, bất động sản, công ty cho vay cũng tận dụng điểm tín dụng để đánh giá sự uy tín của khách hàng. Nếu khách hàng chưa từng vay mượn nợ ngân hàng đồng nghĩa với điểm tín dụng bằng không và những người như vậy sẽ bị xếp ngang hàng với nhóm khách hàng nợ xấu.

Với những chiếc thẻ Mastercard/Visa cho phép khách hàng sử dụng xuyên quốc gia, biết cách sử dụng thẻ tín dụng sẽ giúp khách hàng nâng cao uy tín nhanh chóng.

1. Cách chấm điểm tín dụng của ngân hàng

Đánh giá rủi ro tín dụng là một vấn đề đa chiều và mất cân đối, chủ yếu dựa trên một khối lượng lớn dữ liệu lịch sử như: tình trạng công việc, lịch sử tín dụng, tình trạng tài khoản cá nhân, v.v [Moscato2021ABO].

Có 5 yếu tố mà trung tâm CIC trực thuộc ngân hàng nhà nước đã xét duyệt để các tổ chức/ngân hàng sử dụng để đánh giá điểm tín dụng của khách hàng:

* Lịch sử thanh toán (35%): Đây là thông số có tính quyết định quá trình duyệt hồ sơ nhanh hay chậm. Nếu khách hàng luôn thanh toán các khoản vay đúng hạn sẽ giúp điểm tín dụng nâng cao và quá trình duyệt hồ sơ của ngân hàng đối với khách sẽ đẩy nhanh tiến trình
* Khoản nợ tín dụng (30%): Nắm toàn bộ khoản nợ của khách bao gồm tổng số nợ và tỷ lệ nợ tín dụng từ các khoản vay mà ngân hàng cung cấp. Khách hàng nên duy trì tỷ lệ nợ tín dụng ở mức trung bình, nên hồ sơ của khách được duyệt hay không đều phụ thuộc vào khoản nợ tín dụng.
* Thời gian quan hệ tín dụng (15%): Thời gian quan hệ tín dụng được tính từ khi khách hàng mở tài khoản tín dụng. Thời gian quan hệ tín dụng càng lâu thì tỷ lệ khách hàng được nâng điểm tín dụng và đánh giá tốt càng cao.
* Khoản vay tín dụng mới (10%): là mức độ khách hàng mở tài khoản tín dụng các những khoản vay mới, càng mở nhiều tài khoản vay thì lịch sử tín dụng càng xấu.
* Các loại tín dụng (10%): Các loại tín dụng mà khách hàng sử dụng sẽ được phản ánh qua yếu tố này.

Điểm tín dụng tốt sẽ hội tụ tất cả 5 yếu tố trên và con số sẽ dao động từ 600 – 750 điểm. Lúc này, ngân hàng sẽ dựa trên những thang điểm sau để đánh giá khoản vay của khách hàng có được duyệt hay không và tỉ lệ rủi ro là bao nhiêu:

* Từ 150 – 321 điểm: Không đủ điều kiện vay vốn, mức độ rủi ro, nợ xấu cực cao
* Từ 322 – 430 điểm: Thuộc nhóm khách hàng không có khả năng trả nợ, rủi ro cao
* Từ 431 – 569 điểm: Cần xem xét kỹ các hạn mức và lãi suất cho vay, rủi ro trung bình
* Từ 570 – 679: Có khả năng trả nợ đúng hạn, rủi ro thấp, khách hàng đủ điều kiện vay, lãi suất thấp
* Từ 680 – 750 điểm: Điểm tín dụng lý tưởng, đủ điều kiện vay, hưởng mức lãi suất thấp, nhận được nhiều ưu đãi, hạn mức cao.

1. Nghiên cứu liên quan

Các nghiên cứu trước đây về phân loại điểm tín dụng đã tập trung vào việc áp dụng các phương pháp học máy tiên tiến để cải thiện hiệu suất dự đoán và tăng cường khả năng phát hiện rủi ro. Một số ví dụ về các nghiên cứu có thể kể đến như: Nghiên cứu “Chuẩn mực của các phương pháp học máy để dự đoán điểm tín dụng” của nhóm tác giả [V. Moscato](https://www.semanticscholar.org/author/V.-Moscato/1707400), [A. Picariello](https://www.semanticscholar.org/author/A.-Picariello/1695194), [Giancarlo Sperlí](https://www.semanticscholar.org/author/Giancarlo-Sperl%C3%AD/2682488) đã đề xuất phương pháp điểm chuẩn, cung cấp kết quả dễ hiểu của các chuyên gia với sự minh bạch đầy đủ đặc biệt là cho vấn đề đánh giá rủi ro tín dụng [Moscato2021ABO]. Nghiên cứu “Siêu hình hai cấp lai để chấm điểm tín dụng” của nhóm tác giả [Doruk şen](https://www.semanticscholar.org/author/Doruk-%C5%9Een/101618471), [cem çağrı dönmez](https://www.semanticscholar.org/author/Cem-%C3%87a%C4%9Fr%C4%B1-D%C3%B6nmez/38742740), [U. M. Yildirim](https://www.semanticscholar.org/author/U.-M.-Yildirim/33851788) sử dụng phương pháp lai giữa Thuật toán di truyền (GA) và Máy vector hỗ trợ (SVM) trong cơ chế cung cấp hai cấp để tăng độ chính xác, kết quả cho thấy phương pháp đề xuất sử dụng cơ chế phản hồi theo cấu trúc GA-SVM hai cấp vượt trội hơn các thuật toán phân loại khác trong tài liệu như Cây quyết định, SVM,... cải thiện hiệu quả độ chính xác của phân loại điểm tín dụng [en2020AHB].

Nghiên cứu “Dự đoán điểm tín dụng bằng thuật toán di truyền-Kỹ thuật LSTM” của nhóm tác giả [J. Adisa](https://www.semanticscholar.org/author/J.-Adisa/70264210), [S. Ojo](https://www.semanticscholar.org/author/S.-Ojo/1515810032), [P. Owolawi](https://www.semanticscholar.org/author/P.-Owolawi/21504196), [Agnieta Pretorius](https://www.semanticscholar.org/author/Agnieta-Pretorius/31760513), [S. Ojo](https://www.semanticscholar.org/author/S.-Ojo/9089552). Bài viết này áp dụng Bộ nhớ ngắn hạn dài (LSTM), một trong những thuật toán deep learning tiên tiến nhất vốn phù hợp với lĩnh vực tài chính nhưng hiếm khi được áp dụng để dự đoán chấm điểm tín dụng. Kết quả cho thấy mô hình LSTM được tối ưu hóa vượt trội hơn cả mô hình phân loại đơn và mô hình tập hợp [Adisa2022CreditSP].

Bài nghiên cứu “Phương pháp mới để chấm điểm tín dụng dựa trên chuyển đổi tính năng và mô hình tổng hợp” của nhóm tác giả [Hongxiang Li](https://www.semanticscholar.org/author/Hongxiang-Li/2108578676), [Ao Feng](https://www.semanticscholar.org/author/Ao-Feng/2038472092), [Bin Lin](https://www.semanticscholar.org/author/Bin-Lin/2068284058), [Houcheng Su](https://www.semanticscholar.org/author/Houcheng-Su/2038438668), [Zixi Liu](https://www.semanticscholar.org/author/Zixi-Liu/2127499640), [Xuliang Duan](https://www.semanticscholar.org/author/Xuliang-Duan/2067778409), [Haibo Pu](https://www.semanticscholar.org/author/Haibo-Pu/1890570), [Yifei Wang](https://www.semanticscholar.org/author/Yifei-Wang/2115569035) đã thiết kế một mô hình tập hợp không đồng nhất bằng cách tính trọng số của máy phân tích nhân tố (FM) và mạng lưới thần kinh sâu (DNN), có thể trích xuất một cách hiệu quả các giao lộ bậc thấp và giao lộ bậc cao. Các thử nghiệm toàn diện đã được tiến hành trên hai bộ dữ liệu tiêu chuẩn và kết quả chứng minh rằng phương pháp đề xuất vượt trội hơn các mô hình chấm điểm tín dụng hiện có về độ chính xác [Li2021ANM].Ngoài ra còn một số nghiên cứu liên quan khác được tổng hợp trong bảng dưới đây:

| id | cite | Abstract |
| --- | --- | --- |
| 1 | @article{Chicco2020TheAO,  title={The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation},  author={Davide Chicco and Giuseppe Jurman},  journal={BMC Genomics},  year={2020},  volume={21},  url={https://api.semanticscholar.org/CorpusID:209528322}  } | *This article shows how MCC produces a more informative and truthful score in evaluating binary classifications than accuracy and F1 score, by first explaining the mathematical properties, and then the asset of MCC in six synthetic use cases and in a real genomics scenario.*  Abstract  To evaluate binary classifications and their confusion matrices, scientific researchers can employ several statistical rates, according to the goal of the experiment they are investigating. Despite being a crucial issue in machine learning, no widespread consensus has been reached on a unified elective chosen measure yet. Accuracy and F1 score computed on confusion matrices have been (and still are) among the most popular adopted metrics in binary classification tasks. However, these statistical measures can dangerously show overoptimistic inflated results, especially on imbalanced datasets. The Matthews correlation coefficient (MCC), instead, is a more reliable statistical rate which produces a high score only if the prediction obtained good results in all of the four confusion matrix categories (true positives, false negatives, true negatives, and false positives), proportionally both to the size of positive elements and the size of negative elements in the dataset. In this article, we show how MCC produces a more informative and truthful score in evaluating binary classifications than accuracy and F1 score, by first explaining the mathematical properties, and then the asset of MCC in six synthetic use cases and in a real genomics scenario. We believe that the Matthews correlation coefficient should be preferred to accuracy and F1 score in evaluating binary classification tasks by all scientific communities. |
| 2 | @article{Markoulidakis2021MultiClassCM,  title={Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem},  author={Ioannis Markoulidakis and George Kopsiaftis and Ioannis Rallis and Ioannis Georgoulas},  journal={The 14th PErvasive Technologies Related to Assistive Environments Conference},  year={2021},  url={https://api.semanticscholar.org/CorpusID:235676827}  } | *Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem*  Abstract:  The current paper presents a novel method for reducing a multiclass confusion matrix into a 2 × 2 version enabling the exploitation of the relevant performance metrics and methods such as the receiver operating characteristic and area under the curve for the assessment of different classification algorithms. The reduction method is based on class grouping and leads to a special type of matrix called the reduced confusion matrix. The developed method is then exploited for the assessment of state of the art machine learning algorithms applied on the net promoter score classification problem in the field of customer experience analytics indicating the value of the proposed method in real world classification problems. |
| 3 | @article{Moscato2021ABO,  title={A benchmark of machine learning approaches for credit score prediction},  author={Vincenzo Moscato and Antonio Picariello and Giancarlo Sperl{\'i}},  journal={Expert Syst. Appl.},  year={2021},  volume={165},  pages={113986},  url={https://api.semanticscholar.org/CorpusID:224909498}  } | *A benchmark of machine learning approaches for credit score prediction.*Abstract Credit risk assessment plays a key role for correctly supporting financial institutes in defining their bank policies and commercial strategies. Over the last decade, the emerging of social lending platforms has disrupted traditional services for credit risk assessment. Through these platforms, lenders and borrowers can easily interact among them without any involvement of financial institutes. In particular, they support borrowers in the fundraising process, enabling the participation of any number and size of lenders. However, the lack of lenders’ experience and missing or uncertain information about borrower’s credit history can increase risks in social lending platforms, requiring an accurate credit risk scoring. To overcome such issues, the credit risk assessment problem of financial operations is usually modeled as a binary problem on the basis of debt’s repayment and proper [machine learning techniques](https://www.sciencedirect.com/topics/computer-science/machine-learning-technique) can be consequently exploited. In this paper, we propose a benchmarking study of some of the most used credit risk scoring models to predict if a loan will be repaid in a P2P platform. We deal with a [class imbalance problem](https://www.sciencedirect.com/topics/computer-science/class-imbalance-problem) and leverage several classifiers among the most used in the literature, which are based on different sampling techniques. A real social lending platform (*Lending Club*) data-set, composed by 877,956 samples, has been used to perform the experimental analysis considering different [evaluation metrics](https://www.sciencedirect.com/topics/computer-science/evaluation-metric) (i.e. AUC, Sensitivity, Specificity), also comparing the obtained outcomes with respect to the state-of-the-art approaches. Finally, the three best approaches have also been evaluated in terms of their explainability by means of different [eXplainable Artificial Intelligence](https://www.sciencedirect.com/topics/computer-science/explainable-artificial-intelligence) (XAI) tools. |
| 4 | @article{DeDiego2022GeneralPS,  title={General Performance Score for classification problems},  author={Isaac Mart{\'i}n De Diego and Ana R. Redondo and Rub{\'e}n R. Fern{\'a}ndez and Jorge Navarro and Javier M. Moguerza},  journal={Applied Intelligence},  year={2022},  volume={52},  pages={12049 - 12063},  url={https://api.semanticscholar.org/CorpusID:246462458}  } | *General Performance Score for classification problems* Abstract:  Several performance metrics are currently available to evaluate the performance of Machine Learning (ML) models in classification problems. ML models are usually assessed using a single measure because it facilitates the comparison between several models. However, there is no silver bullet since each performance metric emphasizes a different aspect of the classification. Thus, the choice depends on the particular requirements and characteristics of the problem. An additional problem arises in multi-class classification problems, since most of the well-known metrics are only directly applicable to binary classification problems. In this paper, we propose the General Performance Score (GPS), a methodological approach to build performance metrics for binary and multi-class classification problems. The basic idea behind GPS is to combine a set of individual metrics, penalising low values in any of them. Thus, users can combine several performance metrics that are relevant in the particular problem based on their preferences obtaining a conservative combination. Different GPS-based performance metrics are compared with alternatives in classification problems using real and simulated datasets. The metrics built using the proposed method improve the stability and explainability of the usual performance metrics. Finally, the GPS brings benefits in both new research lines and practical usage, where performance metrics tailored for each particular problem are considered. |
| 5 | @article{Xiao2021ImpactOR,  title={Impact of resampling methods and classification models on the imbalanced credit scoring problems},  author={Jin Xiao and Yadong Wang and Jingfang Chen and Ling Xie and Jing Huang},  journal={Inf. Sci.},  year={2021},  volume={569},  pages={508-526},  url={https://api.semanticscholar.org/CorpusID:235427287}  } | *Impact of resampling methods and classification models on the imbalanced credit scoring problems*Abstract For imbalanced credit scoring, the most common solution is to balance the class distribution of the training set with a [resampling method](https://www.sciencedirect.com/topics/mathematics/resampling-method), and then train a [classification model](https://www.sciencedirect.com/topics/computer-science/classification-models) and classify the customer samples in the test set. However, it is still difficult to select the most appropriate resampling methods and classification models, and the optimal combinations of them have not been identified. Therefore, this study proposes a new [benchmark models](https://www.sciencedirect.com/topics/computer-science/model-benchmark) comparison framework for imbalanced credit scoring. In the framework, we introduce the index of balanced accuracy and four other evaluation measures, experimentally compare the performance of 10 benchmark resampling methods and nine benchmark classification models respectively on six credit scoring data sets, and analyze the optimal combinations of them. The experimental result shows: (1) as for benchmark resampling methods, random under-sampling (a traditional resampling method) and synthetic minority over-sampling technique combined with Wilson’s edited nearest neighbor (an intelligent resampling method) present the best performance; (2) as for benchmark classification models, logistic regression (a single classification model) and [adaptive boosting](https://www.sciencedirect.com/topics/computer-science/adaptive-boosting) (an ensemble classification model) present the best performance; (3) as for optimal combinations, random under-sampling combined with random subspace (an ensemble classification model) can obtain the most satisfactory credit scoring performance. |
| 6 | @article{Wang2021MulticlassificationAO,  title={Multi-classification assessment of bank personal credit risk based on multi-source information fusion},  author={Tianhui Wang and Renjing Liu and Guohua Qi},  journal={Expert Syst. Appl.},  year={2021},  volume={191},  pages={116236},  url={https://api.semanticscholar.org/CorpusID:244755272}  } | *Multi-classification assessment of bank personal credit risk based on multi-source information fusion*Abstract There have been many studies on machine learning and data mining algorithms to improve the effect of credit risk assessment. However, there are few methods that can meet its universal and efficient characteristics. This paper proposes a new multi-classification assessment model of personal credit risk based on the theory of information fusion (MIFCA) by using six machine learning algorithms. The MIFCA model can simultaneously integrate the advantages of multiple classifiers and reduce the interference of uncertain information. In order to verify the MIFCA model, dataset collected from a real data set of commercial bank in China. Experimental results show that MIFCA model has two outstanding points in various assessment criteria. One is that it has higher accuracy for multi-classification assessment, and the other is that it is suitable for various risk assessments and has universal applicability. In addition, the results of this research can also provide references for banks and other financial institutions to strengthen their risk prevention and control capabilities, improve their credit risk identification capabilities, and avoid financial losses. |
| 7 | @article{Mahbobi2021CreditRC,  title={Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks},  author={Mohammad Mahbobi and Salman Kimiagari and Marriappan Vasudevan},  journal={Annals of Operations Research},  year={2021},  volume={330},  pages={609 - 637},  url={https://api.semanticscholar.org/CorpusID:237760427}  } | *Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks* Abstract  This study utilizes classification models to provide a robust algorithm for imbalanced data where the minority class is of the interest, that is, in the context of default payments. In developing an integrated predictive accuracy algorithm, this study proposes machine learning classifiers and applies DNN, SVM, KNN, and ANN. The proposed algorithm utilizes a 30,000 imbalanced dataset to improve the accuracy of the prediction of default payments by implementing oversampling and undersampling strategies, such as synthetic minority oversampling technique (SMOTE), SVM SMOTE, random undersampling, and ALL-KNN. The results indicate that the SVM under the ALL-KNN sampling technique is able to achieve an accuracy of 98.6%, with the lowest cross entropy loss measurement of 0.028. Through the accurate implementation of the neural networks and neurons used in the proposed algorithm, this paper presents better insights into the functioning of the neural networks when used in conjunction with the resampling techniques. Using the methodology and algorithm presented in this study, credit risk assessments can be more accurately predicted in practical applications where most of the clients are categorized as non-default payments. |
| 8 | @article{Imron2020ImprovingAA,  title={Improving Algorithm Accuracy K-Nearest Neighbor Using Z-Score Normalization and Particle Swarm Optimization to Predict Customer Churn},  author={Muhammad Ali Imron and Budi Prasetyo},  journal={Journal of Soft Computing Exploration},  year={2020},  url={https://api.semanticscholar.org/CorpusID:229529952}  } | *Improving Algorithm Accuracy K-Nearest Neighbor Using Z-Score Normalization and Particle Swarm Optimization to Predict Customer Churn* Abstract  Due to increased competition in the business world, many companies use data mining techniques to determine the loyalty level of customers. In this business, data mining can be used to determine the loyalty level of customers. Data mining consists of several research models, one of which is classification. One of the most commonly used methods in classification is the K-Nearest Neighbor algorithm. In this study, the data which used are from German Credit Datasets obtained from UCI machine learning repository. The purpose of this study is to find out how Z-Score works to normalize the data and Particle Swarm Optimization to find the most optimal K value parameters, so the performance of the K-Nearest Neighbor algorithm is more optimal during the classification. The methods which were used to normalize the data are Z-score and Particle Swarm Optimization to determine the most optimal K value. The classification was tested using confusion matrix to determine the generated accuracy. From the finding of this study, the application of Z-score normalization and Particle Swarm Optimization with the K Nearest Neighbor algorithm succeed in increasing the accuracy up to 14%. The initial accuracy was 68.5%, and after applying the normalization of Z-Score and Particle Swarm Optimization, the accuracy became 82.5%. |
| 9 | @article{Salekshahrezaee2023TheEO,  title={The effect of feature extraction and data sampling on credit card fraud detection},  author={Zahra Salekshahrezaee and Joffrey L. Leevy and Taghi M. Khoshgoftaar},  journal={Journal of Big Data},  year={2023},  volume={10},  pages={1-17},  url={https://api.semanticscholar.org/CorpusID:256134459}  } | *The effect of feature extraction and data sampling on credit card fraud detection* Abstract  Training a machine learning algorithm on a class-imbalanced dataset can be a difficult task, a process that could prove even more challenging under conditions of high dimensionality. Feature extraction and data sampling are among the most popular preprocessing techniques. Feature extraction is used to derive a richer set of reduced dataset features, while data sampling is used to mitigate class imbalance. In this paper, we investigate these two preprocessing techniques, using a credit card fraud dataset and four ensemble classifiers (Random Forest, CatBoost, LightGBM, and XGBoost). Within the context of feature extraction, the Principal Component Analysis (PCA) and Convolutional Autoencoder (CAE) methods are evaluated. With regard to data sampling, the Random Undersampling (RUS), Synthetic Minority Oversampling Technique (SMOTE), and SMOTE Tomek methods are evaluated. The F1 score and Area Under the Receiver Operating Characteristic Curve (AUC) metrics serve as measures of classification performance. Our results show that the implementation of the RUS method followed by the CAE method leads to the best performance for credit card fraud detection. |
| 10 | @article{Tripathi2020BinaryBA,  title={Binary BAT algorithm and RBFN based hybrid credit scoring model},  author={Diwakar Tripathi and Damodar Reddy Edla and Venkatanareshbabu Kuppili and Ramesh Dharavath},  journal={Multimedia Tools and Applications},  year={2020},  volume={79},  pages={31889 - 31912},  url={https://api.semanticscholar.org/CorpusID:221283850}  } | *Binary BAT algorithm and RBFN based hybrid credit scoring model* Abstract  Credit scoring is a process of calculating the risk associated with an applicant on the basis of applicant’s credentials such as social status, financial status, etc. and it plays a vital role to improve cash flow for financial industry. However, the credit scoring dataset may have a large number of irrelevant or redundant features which leads to poorer classification performances and higher complexity. So, by removing redundant and irrelevant features may overcome the problem with huge number of features. This work emphasized on the role of feature selection and proposed a hybrid model by combining feature selection by utilizing Binary BAT optimization technique with a novel fitness function and aggregated with for Radial Basis Function Neural Network (RBFN) for credit score classification. Further, proposed feature selection approach is aggregated with Support Vector Machine (SVM) & Random Forest (RF), and other optimization approaches namely: Hybrid Particle Swarm Optimization and Gravitational Search Algorithm (PSOGSA), Hybrid Particle Swarm Optimization and Genetic Algorithm (PSOGA), Improved Krill Herd (IKH), Improved Cuckoo Search (ICS), Firefly Algorithm (FF) and Differential Evolution (DE) are also applied for comparative analysis. |
| 11 | @article{Tian2021ANN,  title={A new non-kernel quadratic surface approach for imbalanced data classification in online credit scoring},  author={Ye Tian and Bo Bian and Xiaofei Tang and Jing Zhou},  journal={Inf. Sci.},  year={2021},  volume={563},  pages={150-165},  url={https://api.semanticscholar.org/CorpusID:233933973}  } | *A new non-kernel quadratic surface approach for imbalanced data classification in online credit scoring*AbstractImbalanced data sets are very common in online credit scoring. Their imbalanced structures may cause statistical bias and poor performance of those traditional models. Hence, how to properly deal with them and dig useful information from them are very crucial for the risk management. This paper proposes a new approach to address the imbalanced data classification in this area by directly generating two quadratic surfaces in the original space. In this way, it avoids the time-consuming task for searching a proper kernel function and its corresponding parameters in the traditional support vector machine (SVM) models, hence significantly improve the total efficiency of the approach. Moreover, the homocentric structure and maximization margin principle are applied to obtain a good performance of our model on the issue. Besides, fuzzy weight is also incorporated to further increase the classification accuracy and robustness. It is worth noting that the linear programming structure of our model not only guarantees the global optimality of its solution, but also leads to a much higher efficiency than those benchmark models. In addition, the fewer parameters in our model further saves more time in the tuning process. Hence it is quite suitable for handling those huge-sized problems in this big data era. Finally, we conduct a comprehensive experiment to compare the performances and efficiencies of different methods. The numerical results on various data sets strongly verify the superiority of our method to some benchmark methods in handling the imbalanced data classification in online credit scoring. |
| 12 | @article{en2020AHB,  title={A Hybrid Bi-level Metaheuristic for Credit Scoring},  author={Doruk Şen and Cem Çağrı D{\"o}nmez and Umman Mahir Yildirim},  journal={Information Systems Frontiers},  year={2020},  volume={22},  pages={1009 - 1019},  url={https://api.semanticscholar.org/CorpusID:220508755}  } | *A Hybrid Bi-level Metaheuristic for Credit ScoringA Hybrid Bi-level Metaheuristic for Credit Scoring*AbstractThis research aims to propose a framework for evaluating credit applications by assigning a binary score to the applicant. The score is targeted to determine whether the credit application is ‘good’ or ‘bad’ in small business purpose loans. Even tiny performance improvements in small businesses may yield a positive impact on the economy as they generate more than 60% of the value. The method presented in this paper hybridizes the Genetic Algorithm (GA) and the Support Vector Machine (SVM) in a bi-level feeding mechanism for increased prediction accuracy. The first level is to determine the parameters of SVM and the second is to find a feature set that increases classification accuracy. To test the proposed approach, we have investigated three different data sets; UCI Australian data set for preliminary works, Lending Club data set for large training and testing, and UCI German and Australian datasets for benchmarking against some other notable methods that use GA. Our computational results show that our proposed method using a feedback mechanism under the hybrid bi-level GA-SVM structure outperforms other classification algorithms in the literature, namely Decision Tree, Random Forests, Logistic Regression, SVM and Artificial Neural Networks, effectively improves the classification accuracy. |
| 13 | @article{Kurniawan2021PREDICTIONFC,  title={PREDICTION FOR COOPERATIVE CREDIT ELIGIBILITY USING DATA MINING CLASSIFICATION WITH C4.5 ALGORITHM},  author={Yogiek Indra Kurniawan and Annastalia Fatikasari and Muhammad Luthfi Hidayat and Mohamad Waluyo},  journal={Jurnal Teknik Informatika (Jutif)},  year={2021},  url={https://api.semanticscholar.org/CorpusID:234893051}  } | *Prediction For Cooperative Credit Eligibility Using Data Mining Classification With C4.5 Algorithm* AbstractBMT Artha Mandiri is a cooperative that provides savings and loans services. In providing credit, BMT Artha Mandiri still uses the manual method, namely by looking at the ledger and history of each customer, to find out whether the applicant is worthy or not worthy of credit so that it is not effective and efficient. The purpose of this research is to make an application that can predict whether a prospective customer is eligible or not to be given credit. Predictions are made using the data mining classification method, namely the C4.5 algorithm based on the supporting data each customer has to classify which factors have the most influence on the level of credit payments in the cooperative. In a built application, the C4.5 algorithm produces a decision tree that is easy to interpret based on the existing variables. In the application, there are features that can be used to make decisions about customers who will apply for credit at the cooperative. The blackbox test results on the application show that the application has been able to run as expected, while the results of the algorithm test also show that the application has been able to implement the C4.5 algorithm correctly. In addition, the results of testing for accuracy show that the maximum average value of Accuracy is 79.19%. |
| 14 | @article{Yu2021AHL,  title={A high-dimensionality-trait-driven learning paradigm for high dimensional credit classification},  author={Lean Yu and Lihang Yu and Kai Yu},  journal={Financial Innovation},  year={2021},  volume={7},  url={https://api.semanticscholar.org/CorpusID:233736434}  } | *A high-dimensionality-trait-driven learning paradigm for high dimensional credit classification*AbstractTo solve the high-dimensionality issue and improve its accuracy in credit risk assessment, a high-dimensionality-trait-driven learning paradigm is proposed for feature extraction and classifier selection. The proposed paradigm consists of three main stages: categorization of high dimensional data, high-dimensionality-trait-driven feature extraction, and high-dimensionality-trait-driven classifier selection. In the first stage, according to the definition of high-dimensionality and the relationship between sample size and feature dimensions, the high-dimensionality traits of credit dataset are further categorized into two types: 100 < feature dimensions < sample size, and feature dimensions ≥ sample size. In the second stage, some typical feature extraction methods are tested regarding the two categories of high dimensionality. In the final stage, four types of classifiers are performed to evaluate credit risk considering different high-dimensionality traits. For the purpose of illustration and verification, credit classification experiments are performed on two publicly available credit risk datasets, and the results show that the proposed high-dimensionality-trait-driven learning paradigm for feature extraction and classifier selection is effective in handling high-dimensional credit classification issues and improving credit classification accuracy relative to the benchmark models listed in this study. |
| 15 | @inproceedings{Dushimimana2020UseOM,  title={Use of Machine Learning Techniques to Create a Credit Score Model for Airtime Loans},  author={Bernard Dushimimana and Yvonne Wambui and Timothy Lubega and Patrick E. McSharry},  year={2020},  url={https://api.semanticscholar.org/CorpusID:225423168}  } | *Use of Machine Learning Techniques to Create a Credit Score Model for Airtime Loans*AbstractAirtime lending default rates are typically lower than those experienced by banks and microfinance institutions (MFIs) but are likely to grow as the service is offered more widely. In this paper, credit scoring techniques are reviewed, and that knowledge is built upon to create an appropriate machine learning model for airtime lending. Over three million loans belonging to more than 41 thousand customers with a repayment period of three months are analysed. Logistic Regression, Decision Trees and Random Forest are evaluated for their ability to classify defaulters using several cross-validation approaches and the latter model performed best. When the default rate is below 2%, it is better to offer everyone a loan. For higher default rates, the model substantially enhances profitability. The model quadruples the tolerable level of default rate for breaking even from 8% to 32%. Nonlinear classification models offer considerable potential for credit scoring, coping with higher levels of default and therefore allowing for larger volumes of customers. |
| 16 | @article{Sadgali2021HumanBS,  title={Human behavior scoring in credit card fraud detection},  author={Imane Sadgali and Nawal Sael and Faouzia Benabbou},  journal={IAES International Journal of Artificial Intelligence},  year={2021},  volume={10},  pages={698-706},  url={https://api.semanticscholar.org/CorpusID:237157434}  } | *Human behavior scoring in credit card fraud detection*AbstractNow days, the analysis of the behavior of cardholders is one of the important fields in electronic payment. This kind of analysis helps to extract behavioral and transaction profile patterns that can help financial systems to better protect their customers. In this paper, we propose an intelligent machine learning (ML) system for rules generation. It is based on a hybrid approach using rough set theory for feature selection, fuzzy logic and association rules for rules generation. A score function is defined and computed for each transaction based on the number of rules, that make this transaction suspicious. This score is kind of risk factor used to measure the level of awareness of the transaction and to improve a card fraud detection system in general. The behavior analysis level is a part of a whole financial fraud detection system where it is combined to intelligent classification to improve the fraud detection. In this work, we also propose an implementation of this system integrating the behavioral layer. The system results obtained are very convincing and the consumed time by our system, per transaction was 6 ms, which prove that our system is able to handle real time process. |
| 17 | @article{Strelcenia2023ImprovingCP,  title={Improving Classification Performance in Credit Card Fraud Detection by Using New Data Augmentation},  author={Emilija Strelcenia and Simant Prakoonwit},  journal={AI},  year={2023},  url={https://api.semanticscholar.org/CorpusID:256534483}  } | *Improving Classification Performance in Credit Card Fraud Detection by Using New Data Augmentation*AbstractIn many industrialized and developing nations, credit cards are one of the most widely used methods of payment for online transactions. Credit card invention has streamlined, facilitated, and enhanced internet transactions. It has, however, also given criminals more opportunities to commit fraud, which has raised the rate of fraud. Credit card fraud has a concerning global impact; many businesses and ordinary users have lost millions of US dollars as a result. Since there is a large number of transactions, many businesses and organizations rely heavily on applying machine learning techniques to automatically classify or identify fraudulent transactions. As the performance of machine learning techniques greatly depends on the quality of the training data, the imbalance in the data is not a trivial issue. In general, only a small percentage of fraudulent transactions are presented in the data. This greatly affects the performance of machine learning classifiers. In order to deal with the rarity of fraudulent occurrences, this paper investigates a variety of data augmentation techniques to address the imbalanced data problem and introduces a new data augmentation model, K-CGAN, for credit card fraud detection. A number of the main classification techniques are then used to evaluate the performance of the augmentation techniques. These results show that B-SMOTE, K-CGAN, and SMOTE have the highest Precision and Recall compared with other augmentation methods. Among those, K-CGAN has the highest F1 Score and Accuracy. |
| 18 | @article{Lin2022AnEP,  title={An Efficient Privacy-Preserving Credit Score System Based on Noninteractive Zero-Knowledge Proof},  author={Chao Lin and Min Luo and Xinyi Huang and Kim-Kwang Raymond Choo and De-biao He},  journal={IEEE Systems Journal},  year={2022},  volume={16},  pages={1592-1601},  url={https://api.semanticscholar.org/CorpusID:232333058}  } | *An Efficient Privacy-Preserving Credit Score System Based on Noninteractive Zero-Knowledge Proof*AbstractCredit system is generally associated with the banking and financial institutions, although it has far reaching implications for residents of countries, such as U.S., particularly for those with a poor credit history. Specifically, a credit score computation (CSC) quantifies an individual’s credit value or credit risk, which is used by banking and financial institutions, as well as other entities (e.g., during purchasing of insurance policies and application of rental properties), to facilitate their decision-making (e.g., whether to approve the insurance policy purchase or the level of premium). Although a number of CSC models have been proposed in the literature for supporting different application scenarios, privacy protection of CSC is rarely considered despite the potential for leakage of user private information (e.g., registration, hobbies, credit, relationships, and inquiry). Such information can then be abused for other nefarious activities, such as identity theft and credit card fraud. Thus, in this article, we first analyze the privacy strength of existing CSC models, prior to presenting the formal definition of a privacy-preserving CSC system alongside its security requirements. Then, we propose a concrete construction based on Paillier encryption with three proposed noninteractive zero-knowledge schemes. To demonstrate feasibility of our proposal, we evaluate both its security and performance. |
| 19 | @article{Prassanna2020SecrecyProtectorAN,  title={SecrecyProtector: A Novel Data Analytics based credit score management system},  author={J. Prassanna and Abdul Quadir and J ChristyJackson and R. Prabakaran and Sakkaravarthi Ramanathan},  journal={International Journal of Scientific \& Technology Research},  year={2020},  volume={9},  pages={1215-1218},  url={https://api.semanticscholar.org/CorpusID:216291604}  } | *SecrecyProtector: A Novel Data Analytics based credit score management system*Abstract: This work gives an account of the Credit Score web service application and the primary purpose of a credit score is to help lenders assess individuals' risk of not repaying a loan. Credit scoring assessment, despite the fact that a moderately new idea in the Indian money related business sector, have increased wide acknowledgment among financial specialists. In the meantime, easy-going and narrative confirmation recommends that there are worries among speculators and controllers about the execution of rating offices in India. This paper looks at financial specialists' mindfulness, discernment, understanding level and use of Credit scoring assessment through a poll-based example overview covering individual and additionally institutional speculators. We find high dissemination of rating use among all class of financial specialists, however, there is a recognizable upsetting with the dependability of appraisals, inclination of ensuing minimizing and opportuneness of rating reconnaissance. The review additionally uncovers that the institutional financial specialists have predominant information and comprehension about evaluations than individual speculators. In this way, the review underlines the requirement for rating offices to take a shot at instructing the basic speculators to engender appropriate comprehension and use of Credit score.. |
| 20 | @article{Adisa2022CreditSP,  title={Credit Score Prediction using Genetic Algorithm-LSTM Technique},  author={Juliana Adeola Adisa and Samuel Olusegun Ojo and Pius Adewale Owolawi and Agnieta Pretorius and S. Ojo},  journal={2022 Conference on Information Communications Technology and Society (ICTAS)},  year={2022},  pages={1-6},  url={https://api.semanticscholar.org/CorpusID:247961496}  } | *Credit Score Prediction using Genetic Algorithm-LSTM Technique*AbstractIn data mining, the goal of prediction is to develop a more effective model that can provide accurate results. Prior literature has studied different classification techniques and found that combining multiple classifiers into ensembles outperformed most single classifier approaches. The performance of an ensemble classifier can be affected by some factors. How to determine the best classification technique? Which combination method to employ? This paper applies Long Short-Term Memory (LSTM), one of the most advanced deep learning algorithms which are inherently appropriate for the financial domain but rarely applied to credit scoring prediction. The research presents an optimization approach to determine the optimal parameters for a deep learning algorithm. The LSTM parameters are determined using an optimization algorithm. The LSTM parameters include epochs, batch size, number of neurons, learning rate and dropout. The results show that the optimized LSTM model outperforms both single classifiers and ensemble models. |
| 21 | @article{Parvin2020AnEC,  title={An Ensemble Classifier Model to Predict Credit Scoring - Comparative Analysis},  author={A. Safiya Parvin and B. Saleena},  journal={2020 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS)},  year={2020},  pages={27-30},  url={https://api.semanticscholar.org/CorpusID:234478544}  } | *An Ensemble Classifier Model to Predict Credit Scoring - Comparative Analysis*AbstractCredit scoring is a way of analyzing statistical data used in financial organizations and banks to acquire a person’s creditworthiness. The bestowers generally manipulate it to decide to widen or retract credit. The score plays a significant role in determining the creditworthiness of a person and if he/she can be sanctioned a loan or not. Machine learning techniques help us to predict the credit score more accurately using classification algorithms. Few base and ensemble classification algorithms were used in this research to perform a comparative analysis. The ensemble method incorporates several base classification algorithms like Decision trees, Logistic Regression, Nearest neighbor, Support Vector Machine, etc. to achieve better results. The objective of this paper is to predict the credit score based on different classifier models and evaluate the performance of each model based on the metrics. A comparative analysis is done to identify the best classifier to predict the credit score. The evaluation metrics used for evaluating the model are Recall, Precision, F-measure, and Accuracy. Error measures like MAE and RMSE of the model were also used to evaluate the model. This helps us to improve the decision in identifying the more accurate classifier model. The dataset used for this analysis is the Australian credit dataset from the UCI Machine learning repository. Experimental results prove that the Random Forest and Extratree classifier model produces better accuracy in ensemble classifiers and the SVM model furnishes better accuracy in the base classifier. |
| 22 | @article{Li2021ANM,  title={A novel method for credit scoring based on feature transformation and ensemble model},  author={Hongxiang Li and Ao Feng and Bin Lin and Houcheng Su and Zixi Liu and Xuliang Duan and Haibo Pu and Yifei Wang},  journal={PeerJ Computer Science},  year={2021},  volume={7},  url={https://api.semanticscholar.org/CorpusID:235479482}  } | *A novel method for credit scoring based on feature transformation and ensemble model*AbstractCredit scoring is a very critical task for banks and other financial institutions, and it has become an important evaluation metric to distinguish potential defaulting users. In this paper, we propose a credit score prediction method based on feature transformation and ensemble model, which is essentially a cascade approach. The feature transformation process consisting of boosting trees (BT) and auto-encoders (AE) is employed to replace manual feature engineering and to solve the data imbalance problem. For the classification process, this paper designs a heterogeneous ensemble model by weighting the factorization machine (FM) and deep neural networks (DNN), which can efficiently extract low-order intersections and high-order intersections. Comprehensive experiments were conducted on two standard datasets and the results demonstrate that the proposed approach outperforms existing credit scoring models in accuracy. |
| 23 | @article{Li2021CreditSB,  title={Credit scoring by one-class classification driven dynamical ensemble learning},  author={Hao Li and Hao Qiu and Shu Sun and Jun Chang and Wenting Tu},  journal={Journal of the Operational Research Society},  year={2021},  volume={73},  pages={181 - 190},  url={https://api.semanticscholar.org/CorpusID:237749067}  } | *Credit scoring by one-class classification driven dynamical ensemble learning*AbstractAbstract It is very useful to endow machines with the ability to measure credit scores of loan applicants. Conventional methodologies always train Credit Scoring (CS) models by using data from clients who passed previous credit examination (i.e. who were considered adequately creditworthy and took out a loan). However, the CS models trained on data from the applicants who with good credit background may not work well for new applicants with plain or ambiguous credit backgrounds. Previous work always alleviates this by techniques of rejected inference and semisupervised learning. In this article, we propose a novel approach called as “One-class Classification Driven Dynamical Ensemble Learning” (abbreviated as OCDDEL). Different from rejected inference or semisupervised learning, OCDDEL does not use inferred labels of past rejected applications. Instead, OCDDEL only relies on past accepted applications and their true labels. It builds a dynamical ensemble model which deal with different test applications in different ways. To determine the ensemble weights for a specific test case, OCDDEL will learn a one-class classifier to separate test applications into groups, according to their similarities with training applicants. An experimental evaluation with 2 real-world datasets demonstrates the effectiveness of our approach. |
| 24 | @article{Mokheleli2023MachineLA,  title={Machine Learning Approach for Credit Score Predictions},  author={Tsholofelo Mokheleli and Tinofirei Museba},  journal={Journal of Information Systems and Informatics},  year={2023},  url={https://api.semanticscholar.org/CorpusID:259551401}  } | *Machine Learning Approach for Credit Score Predictions*AbstractThis paper addresses the problem of managing the significant rise in requests for credit products that banking and financial institutions face. The aim is to propose an adaptive, dynamic heterogeneous ensemble credit model that integrates the XGBoost and Support Vector Machine models to improve the accuracy and reliability of risk assessment credit scoring models. The method employs machine learning techniques to recognise patterns and trends from past data to anticipate future occurrences. The proposed approach is compared with existing credit score models to validate its efficacy using five popular evaluation metrics, Accuracy, ROC AUC, Precision, Recall and F1\_Score. The paper highlights credit scoring models’ challenges, such as class imbalance, verification latency and concept drift. The results show that the proposed approach outperforms the existing models regarding the evaluation metrics, achieving a balance between predictive accuracy and computational cost. The conclusion emphasises the significance of the proposed approach for the banking and financial sector in developing robust and reliable credit scoring models to evaluate the creditworthiness of their clients. |
| 25 | @article{Sandeep2022ANA,  title={A Novel Approach for Bank Loan Approval by Verifying Background Information of Customers through Credit Score and Analyze the Prediction Accuracy using Random Forest over Linear Regression Algorithm},  author={Ch.Venkata Sandeep and T. Kalavathi Devi},  journal={Journal of Pharmaceutical Negative Results},  year={2022},  url={https://api.semanticscholar.org/CorpusID:252893023}  } | *A Novel Approach for Bank Loan Approval by Verifying Background Information of Customers through Credit Score and Analyze the Prediction Accuracy using Random Forest over Linear Regression Algorithm*AbstractAim: To analyze the accuracy of Novel Random Forest (RF) and Linear Regression Algorithm (LR) algorithms used to approve bank loans. Materials and Methods: The existing model uses Linear Regression Algorithm (LR) and the proposed model employs a Novel Random Forest (RF). The Random Forest is a supervised learning model, it constructs solutions for different regression problems. It provides a high rate of accuracy by cross-validation. The 20 sample values are used to find out the Mean, Std. Deviation and Std. error means. The sample size was measured as 40 per group using G power (80%). Results: The resultant graph explains the comparison of the mean accuracy values of algorithms Novel Random Forest (RF) and Linear Regression (LR) where the mean accuracy of the Novel random forest is about 70.5% and the mean accuracy value of the Linear Regression is about 69.5%. The significance obtained is p=1.0 that is p>0.05, it shows insignificance between the groups based on independent sample T-Test. Conclusion : The mean accuracy rate of the Novel Random Forest algorithm has been improved to 70.5% compared to Linear Regression which is having around 69.5% mean accuracy. |
| 26 | @article{Chen2022ACS,  title={A Credit Scoring Ensemble Framework using Adaboost and Multi-layer Ensemble Classification},  author={R. Chen and C. H. Ju and F. Shen. Tu.},  journal={Proceedings of the 2022 International Conference on Pattern Recognition and Intelligent Systems},  year={2022},  url={https://api.semanticscholar.org/CorpusID:251694449}  } | *A Credit Scoring Ensemble Framework using Adaboost and Multi-layer Ensemble Classification*AbstractThe accuracy of classification plays a crucial role in the financial industry, and an increase of 1% in the accuracy of credit scoring in credit customer selection, risk measurement, etc. would significantly reduce the losses of financial institutions. Sometimes one particular classifier perform better than others for a given dataset, while the performance may worse than other classifiers for other datasets. Many studies have shown that classifier ensemble method is a more effective approach. a multi-level weighted voting classification algorithm based on the combination of classifier ranking and Adaboost algorithm is proposed in this paper.. Four feature selection methods are used to select the features, and then seven commonly used heterogeneous classifiers are used to select five classifiers and calculate their ranks, and then AdaBoost is used to boost the performance of the selected base classifiers and calculate the updated F1 and ranks. The effects of ensemble framework Majority Voting (MV), Weighted Voting (WV), Layered Majority Voting (LMV), Layered Weighted Voting (LWV) were all evaluated from the aspects of accuracy, sensitivity, specificity, and G-measure. In addition, the ROC curves of each ensemble framework are plotted for analysis, and the outcome of the experiments shows that our presented method achieves significant results on Australian credit score data and some progress on the German loan approval data. |
| 27 | @article{Radovi2021CreditSW,  title={Credit scoring with an ensemble deep learning classification methods – comparison with traditional methods},  author={Ognjen Radovi{\'c} and Srđan Marinkovi{\'c} and Jelena Radoji{\vc}i{\'c}},  journal={Facta Universitatis. Series: Economics and Organization},  year={2021},  url={https://api.semanticscholar.org/CorpusID:233377257}  } | *Credit scoring with an ensemble deep learning classification methods – comparison with traditional methods*AbstractCredit scoring attracts special attention of financial institutions. In recent years, deep learning methods have been particularly interesting. In this paper, we compare the performance of ensemble deep learning methods based on decision trees with the best traditional method, logistic regression, and the machine learning method benchmark, support vector machines. Each method tests several different algorithms. We use different performance indicators. The research focuses on standard datasets relevant for this type of classification, the Australian and German datasets. The best method, according to the MCC indicator, proves to be the ensemble method with boosted decision trees. Also, on average, ensemble methods prove to be more successful than SVM. |
| 28 | @article{Kazemi2023EstimationOO,  title={Estimation of optimum thresholds for binary classification using genetic algorithm: An application to solve a credit scoring problem},  author={Hamid Reza Kazemi and Kaveh Khalili Damghani and Soheil Sadi-Nezhad},  journal={Expert Systems},  year={2023},  volume={40},  url={https://api.semanticscholar.org/CorpusID:256630574}  } | *Estimation of optimum thresholds for binary classification using genetic algorithm: An application to solve a credit scoring problem*AbstractThe main issue in a classification problem is classifying observations into various disjoint classes. Different classification techniques generate a continuous number between a and b, usually between 0 and 1; thus, the optimal cut‐off value(s) must be carefully selected to discriminate classes precisely. The decision is about setting a threshold value and transforming the continuous score into a binary output. Therefore, in addition to using the so‐called sophisticated classification methods to have a more accurate classification, there is a need to identify and choose the optimal threshold value(s). However, the latter has not been thoroughly investigated. Hence, this study proposes an approach based on a Genetic Algorithm (GA) and Neural Networks (NNs) to automatically find customized cut‐off values, considering different performance criteria and given datasets. Since credit scoring is a binary classification problem, two popular credit scoring datasets, namely “Australian” and “German” credit datasets, are used to test the proposed approach. Our numerical results revealed that the proposed GA‐NN model could successfully find customized acceptance thresholds, considering predetermined performance criteria, including Accuracy, Estimated Misclassification Cost (EMC), and Area under ROC Curve (AUC) for the tested datasets. Furthermore, the best‐obtained results and the paired‐samples t‐test results show that utilizing the customized cut‐off points leads to a more accurate classification than the commonly‐used threshold value of 0.5. |
| 29 | @article{Aljadani2023MathematicalMA,  title={Mathematical Modeling and Analysis of Credit Scoring Using the LIME Explainer: A Comprehensive Approach},  author={Abdussalam Aljadani and Bshair Alharthi and Mohammed A. Farsi and Hossam Magdy Balaha and Mahmoud Badawy and Mostafa A. Elhosseini},  journal={Mathematics},  year={2023},  url={https://api.semanticscholar.org/CorpusID:263012527}  } | *Mathematical Modeling and Analysis of Credit Scoring Using the LIME Explainer: A Comprehensive Approach*AbstractCredit scoring models serve as pivotal instruments for lenders and financial institutions, facilitating the assessment of creditworthiness. Traditional models, while instrumental, grapple with challenges related to efficiency and subjectivity. The advent of machine learning heralds a transformative era, offering data-driven solutions that transcend these limitations. This research delves into a comprehensive analysis of various machine learning algorithms, emphasizing their mathematical underpinnings and their applicability in credit score classification. A comprehensive evaluation is conducted on a range of algorithms, including logistic regression, decision trees, support vector machines, and neural networks, using publicly available credit datasets. Within the research, a unified mathematical framework is introduced, which encompasses preprocessing techniques and critical algorithms such as Particle Swarm Optimization (PSO), the Light Gradient Boosting Model, and Extreme Gradient Boosting (XGB), among others. The focal point of the investigation is the LIME (Local Interpretable Model-agnostic Explanations) explainer. This study offers a comprehensive mathematical model using the LIME explainer, shedding light on its pivotal role in elucidating the intricacies of complex machine learning models. This study’s empirical findings offer compelling evidence of the efficacy of these methodologies in credit scoring, with notable accuracies of 88.84%, 78.30%, and 77.80% for the Australian, German, and South German datasets, respectively. In summation, this research not only amplifies the significance of machine learning in credit scoring but also accentuates the importance of mathematical modeling and the LIME explainer, providing a roadmap for practitioners to navigate the evolving landscape of credit assessment. |
| 30 | @article{Melyanti2021OnlineDO,  title={Online Determination of Credit Score (PAK) Application Functional Teachers},  author={Rika Melyanti and M. Giatman and Riri Mayliza},  journal={International Journal of Management and Humanities},  year={2021},  url={https://api.semanticscholar.org/CorpusID:235551539}  } | *Online Determination of Credit Score (PAK) Application Functional Teachers*AbstractDetermination of Teacher Credit Numbers (PAK) is proposed by the teacher, which are evaluated and evaluated by the Assessment Team. Calculation of credit numbers in PAK still uses manual methods using Microsoft Excel for inputting and Microsoft results as reports, errors in input such as typos and risk of accidental deletion of data still occur frequently. The DUPAK report that will be input into the PAK system is also still waiting for the Assessment Team to send the file to the Pelalawan Regency Education Office so that it takes more time to complete the functional teacher promotion report. To overcome this problem, the credit score calculation process is fast and accurate. A new web-based system, which includes all the elements that are valued by credit numbers. In the old system procedure and the new system it is not much different, the fundamental difference between the new system uses a web-based computer technology in data management that can shorten the data entry process and can overcome the obstacles of the old system. |