**XÂY DỰNG MÔ HÌNH CHẤM ĐIỂM TÍN DỤNG**

**CỦA KHÁCH HÀNG**

***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.**Bài báo đề xuất các mô hình học cho dự đoán chấm điểm tín dụng của khách hàng; với năm mô hình học máy Decision Tree, K-Nearest Neighbor, Random Forest, Naive Bayes và hai mô hình học sâu Convolutional Neural Network, Recurrent Neural Network. Để chuẩn bị dữ liệu cho huấn luyện mô hình, chúng tôi sử dụng các kỹ thuật tiền xử lý dữ liệu khác nhau chẳng hạn như trích xuất tính năng, xử lý các giá trị bị thiếu, quản lý các ngoại lệ và áp dụng các phương pháp cân bằng dữ liệu. Kết quả các thử nghiệm cho thấy Random Forest đã thu được kết quả tốt với các chỉ số accuracy là 88.58%, recall là 88.78%, precision là 88.58% và F1-score là 88.27%.*

# GIỚI THIỆU

Điểm tín dụng là điểm các ngân hàng, tổ chức tài chính sử dụng để đánh giá sự uy tín của khách hàng khi các tổ chức đó cung cấp cho khách hàng các dịch vụ tài chính hay các hình thức cho vay để họ sử dụng. Khi một người hoặc doanh nghiệp đề xuất vay tiền từ một tổ chức tín dụng, tổ chức đó cần phải đánh giá xem liệu họ có đủ khả năng trả nợ hay không. Một vấn đề thường gặp trong lĩnh vực tài chính và ngân hàng là rủi ro tín dụng, được hiểu là khi người sử dụng tín dụng không thực hiện nghĩa vụ đối với ngân hàng. Nguyên nhân của rủi ro này thường là do các ngân hàng áp dụng tiêu chuẩn cho vay quá lỏng lẻo mà không xem xét kỹ lưỡng độ đáng tin cậy của khách hàng tiềm năng [1]. Trong tình hình này, một vấn đề cần phải được giải quyết là phân tích rủi ro tín dụng để xác định người vay có thể gặp khó khăn trong việc hoàn trả khoản vay hay không. Để đánh giá rủi ro này, các tổ chức tài chính và ngân hàng thường tính toán thông qua việc chấm điểm tín dụng của khách hàng.

Chấm điểm tín dụng là một quy trình đánh giá khả năng thực hiện các nghĩa vụ tài chính của một khách hàng đối với ngân hàng như trả nợ gốc và lãi vay khi đến hạn hoặc các điều kiện tín dụng khác nhằm xác định mức định rủi ro trong hoạt động tín dụng của ngân hàng. Sự thiếu kinh nghiệm của những người cho vay cùng với sự thiếu thông tin hoặc không chắc chắn về lịch sử tín dụng của người vay có thể tăng nguy cơ rủi ro trong các nền tảng cho vay, do đó, việc đánh giá rủi ro tín dụng một cách chính xác trở nên cần thiết. Đánh giá rủi ro tín dụng đóng một vai trò quan trọng để hỗ trợ cho các tổ chức tài chính trong việc phân tích khách hàng và từ đó xác định chính sách của ngân hàng và chiến lược thương mại của họ [2]. Ngân hàng thường sử dụng phương pháp chấm điểm tín dụng dựa trên các mô hình được phát triển từ bộ dữ liệu giàu tính năng để đánh giá rủi ro tín dụng đối với khách hàng hiện tại, các mô hình này bao gồm dữ liệu nhân khẩu học, lịch sử tín dụng, thu nhập và các yếu tố khác để dự đoán khả năng trả nợ của cá nhân hoặc tổ chức. Phương pháp chấm điểm tín dụng thường hiệu quả trong việc dự đoán khả năng vỡ nợ và không vỡ nợ của khách hàng trong tương lai. Điều này chủ yếu nhờ vào khả năng truy cập vào thông tin lịch sử tín dụng của khách hàng, bao gồm thông tin về quá khứ, các khoản thanh toán quá hạn, nợ đọng và tình trạng chồng chéo nợ [3]. Các biến giải thích được tạo ra trên cơ sở những thông tin như vậy rất quan trọng để phát triển các mô hình chấm điểm tín dụng hiệu quả, có khả năng dự đoán chính xác các khoản vỡ nợ trong tương lai. Sự sáng tạo mô hình chấm điểm tín dụng mạnh mẽ, hiệu quả và có khả năng thích ứng có tác động đáng kể về khả năng sinh lời của các tổ chức tài chính [4]. Mỗi mô hình chấm điểm rủi ro tín dụng phải tuân thủ các quy định nghiêm ngặt và bất kỳ hành vi vi phạm nào cũng có thể dẫn đến hậu quả nghiêm trọng chi phí theo quy định. Vì vậy, việc tạo ra các mô hình chấm điểm tín dụng có khả năng thích ứng, hiệu quả và mạnh mẽ trong việc dự đoán chính xác các khoản nợ không trả được nợ là rất quan trọng. Các mô hình chấm điểm tín dụng được phát triển để tăng cường quá trình ra quyết định đặc biệt cho các tổ chức tài chính để đối phó với rủi ro liên quan đến ứng cử viên tín dụng trong khi đăng ký sản phẩm tín dụng mới. Để khắc phục những vấn đề như vậy, trong những năm qua lĩnh vực nghiên cứu chấm điểm tín dụng của khách hàng đang rất được quan tâm và đem lại được kết quả đáng mong đợi.

Các mô hình dự đoán về chấm điểm tín dụng có thể được phân thành hai loại chính phương pháp thống kê và phương pháp trí tuệ nhân tạo [5]. Về các phương pháp thống kê, chúng đã được đề xuất nhưng gặp phải các vấn đề về phạm vi bao phủ vốn có của các hiệu ứng phi tuyến tính giữa các biến liên quan. Đánh giá rủi ro tín dụng được đặc trưng bởi các đặc tính sau: sự phụ thuộc, độ phức tạp và tính liên kết, do đó việc ước tính điểm tín dụng rất phức tạp vì nó phụ thuộc vào các thông số khác nhau.

Sự phát triển của ngành khoa học dữ liệu trong những năm gần đây đã giải quyết nhiều bài toán khác nhau, và đây là một lựa chọn rất đúng đắn cho các ngân hàng trong việc giải quyết bài toán chấm điểm tín dụng. Mục tiêu của nghiên cứu này là tìm ra phương pháp tốt nhất để dự đoán điểm tín dụng của khách hàng, từ đó giúp các ngân hàng áp dụng biện pháp kịp thời để duy trì sự gắn bó của khách hàng với dịch vụ của họ.

Trong phần tiếp theo tôi giới thiệu một số Nghiên cứu liên quan ở Phần 2; Phương pháp đề xuất được đưa ra ở Phần 3; Phần 4 là Thực nghiệm, kết quả và phân tích; Kết luận và nghiên cứu tương lai ở Phần 5.

# NGHIÊN CỨU LIÊN QUAN

Cho đến nay nhiều nghiên cứu đã đề xuất các phương pháp nhằm giải quyết bài toán chấm điểm khách hàng bằng nhiều mô hình khác nhau. Trong [4] đề xuất một mô hình đánh giá rủi ro tín dụng tích hợp các mô hình XGBoost và Support Vector Machine (SVM) để nâng cao độ chính xác và độ tin cậy của các mô hình chấm điểm tín dụng đánh giá rủi ro. Phương pháp này sử dụng các kỹ thuật học máy để nhận ra các mẫu và xu hướng từ dữ liệu trong quá khứ để dự đoán các sự kiện xảy ra trong tương lai. Nghiên cứu đề xuất mô hình Adaptive and Dynamic Heterogeneous Ensemble (ADHE) để chấm điểm tín dụng. ADHE là một kỹ thuật học máy kết hợp nhiều mô hình để cải thiện độ chính xác và mạnh mẽ của dự đoán. Nó liên quan đến việc tạo ra một tập hợp các mô hình đa dạng bổ sung cho điểm mạnh và điểm yếu của nhau. Khía cạnh thích ứng và năng động đề cập đến khả năng điều chỉnh bộ quần áo để đáp ứng với những thay đổi về dữ liệu và môi trường. Trong bối cảnh chấm điểm tín dụng, phương pháp tiếp cận ADHE tích hợp các mô hình XGBoost và Support Vector Machine để tạo ra một mô hình chấm điểm tín dụng chính xác và đáng tin cậy hơn. Bài viết so sánh hiệu suất dự đoán mô hình ADHE với các mô hình đồng nhất khác. Kết quả của hệ thống đề xuất cho thấy mô hình ADHE thường hoạt động tốt hơn các mô hình đồng nhất.

Trong [6] sử dụng bộ dữ liệu bao gồm lịch sử giao dịch thẻ tín dụng và hồ sơ khách hàng, đã được đề xuất và thực hiện thử nghiệm bằng nhiều thuật toán học máy, bao gồm logistic regression (LR), Decision Tree (DT), random forest (RF), multi layer perceptron (MLP), XGBoost và LightGBM. Kết quả thử nghiệm chứng minh rằng MLP vượt trội hơn LR, DT, RF, LightGBM và XGBoost về hiệu suất dự đoán trong tỷ lệ dương tính thực sự, đạt được diện tích ấn tượng dưới đường cong (AUC) là 86,7% và chỉ số accuracy là 91,6%, với recall vượt quá 80%. Những kết quả này cho thấy sự vượt trội của MLP trong việc dự đoán khách hàng mặc định và đánh giá các rủi ro tiềm ẩn. Trong [7] sử dụng các mô hình machine learning (ML) khác nhau để phát triển một mô hình mới để dự đoán điểm tín dụng. Các thuật toán ML support vector machine (SVM), neural networks (NNs), decision trees (DT), random forest (RF) và logistic regression (LR) được sử dụng ở đây cho các thí nghiệm cùng với các phương pháp lựa chọn tính năng IG, GR và chi-square để dự đoán tín dụng trên các bộ dữ liệu của Úc và Đức. Nghiên cứu cung cấp sự hiểu biết về quá trình ra quyết định cho các đặc điểm thông tin và chức năng của học máy (ML) trong các nhiệm vụ dự đoán tín dụng. Phân tích thực nghiệm cho thấy trong trường hợp tập dữ liệu của Đức, mô hình DT với lựa chọn tính năng GR tối ưu hóa tham số vượt trội hơn SVM và NN với accuracy là 99,78%, recall là 99,39% và precision là 99,43%. Đối với tập dữ liệu của Úc, mô hình SVM với lựa chọn tính năng GR vượt trội hơn NN và DT với accuracy là 99,98%, recall là 97,67% và precision là 97.67%.

Nghiên cứu [8] phân tích độ chính xác của các thuật toán rừng ngẫu nhiên (RF) và thuật toán hồi quy tuyến tính (LR) được sử dụng để phê duyệt các khoản vay ngân hàng. Biểu đồ kết quả giải thích việc so sánh các giá trị chính xác trung bình của các thuật toán RF và LR trong đó độ chính xác trung bình của rừng ngẫu nhiên là khoảng 70,5% và giá trị chính xác trung bình của hồi quy tuyến tính là khoảng 69,5%.

Phân tích chấm điểm tín dụng [9] nghiên cứu này đi sâu vào phân tích toàn diện các thuật toán học máy khác nhau, nhấn mạnh nền tảng toán học của chúng và khả năng ứng dụng của chúng trong phân loại điểm tín dụng. Đánh giá toàn diện được thực hiện trên một loạt các thuật toán, bao gồm hồi quy logistic, cây quyết định, máy vectơ hỗ trợ và mạng thần kinh, sử dụng bộ dữ liệu tín dụng có sẵn công khai. Trọng tâm của cuộc điều tra là giải thích LIME (Giải thích mô hình bất khả tri có thể diễn giải cục bộ). Nghiên cứu này cung cấp một mô hình toán học toàn diện bằng cách sử dụng trình giải thích LIME, làm sáng tỏ vai trò quan trọng của nó trong việc làm sáng tỏ sự phức tạp của các mô hình học máy phức tạp. Những phát hiện thực nghiệm của nghiên cứu này cung cấp bằng chứng thuyết phục về hiệu quả của các phương pháp này trong chấm điểm tín dụng, với độ chính xác đáng chú ý lần lượt là 88,84%, 78,30% và 77,80% cho các bộ dữ liệu của Úc, Đức và Nam Đức.

**TÀI LIỆU THAM KHẢO**

1. Almajid, Adi Sakti. "Multilayer Perceptron Optimization on Imbalanced Data Using SVM-SMOTE and One-Hot Encoding for Credit Card Default Prediction." Journal of Advances in Information Systems and Technology (2022).
2. Moscato, Vincenzo, Antonio Picariello and Giancarlo Sperlí. "A benchmark of machine learning approaches for credit score prediction." Expert Syst. Appl. 165 (2021): 113986.
3. Hjelkrem, Lars Ole, Petter Eilif de Lange and Erik Nesset. "The Value of Open Banking Data for Application Credit Scoring: Case Study of a Norwegian Bank." Journal of Risk and Financial Management (2022).
4. Mokheleli, Tsholofelo and Tinofirei Museba. "Machine Learning Approach for Credit Score Predictions." Journal of Information Systems and Informatics (2023).
5. Amato, Flora, Antonino Ferraro, Antonio Galli, Francesco Moscato, Vincenzo Moscato and Giancarlo Sperlí. "Credit Score Prediction Relying on Machine Learning." Trong Sistemi Evoluti per Basi di Dati (2022).
6. Arram, Anas, Masri Ayob, Musatafa Abbas Abbood Albadr, Alaa Sulaiman and Dheeb Albashish. "Credit card score prediction using machine learning models: A new dataset." ArXiv (2023): abs/2310.02956.
7. Jakka, Geethamanikanta, Amrutanshu Panigrahi, Abhilash Pati, Manmath Nath Das and Jyotsnarani Tripathy. "A novel credit scoring system in financial institutions using artificial intelligence technology." Journal of Autonomous Intelligence (2023).
8. Sandeep, Ch. Venkata. "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." Journal of Pharmaceutical Negative Results (2022).
9. Aljadani, Abdussalam, Bshair Alharthi, Mohammed A. Farsi, Hossam Magdy Balaha, Mahmoud Badawy and Mostafa A. Elhosseini. "Mathematical Modeling and Analysis of Credit Scoring Using the LIME Explainer: A Comprehensive Approach." Mathematics (2023).

|  |  |  |
| --- | --- | --- |
| **id** | **Cite** | **Abstract** |
|  | @article{Almajid2022MultilayerPO,    title={Multilayer Perceptron Optimization on Imbalanced Data Using SVM-SMOTE and One-Hot Encoding for Credit Card Default Prediction},    author={Adi Sakti Almajid},    journal={Journal of Advances in Information Systems and Technology},    year={2022},    url={https://api.semanticscholar.org/CorpusID:253334436}  } | ***Multilayer Perceptron Optimization on Imbalanced Data Using SVM-SMOTE and One-Hot Encoding for Credit Card Default Prediction***  Credit risk assessment analysis by classifying potential users is an important process to reduce the occurrence of default users. The problems faced from the classification process using real-world datasets are imbalanced data that causes bias-to-majority in model training outcomes. These problems cause the algorithm to only focus on the majority class and ignore the minority class, even though both classes have the same important role. To overcome this problem, a combination of One-hot encoding (OHE) and SVM-Synthetic minority oversampling technique (SVM-SMOTE) techniques are used for the optimization process of the MLP classification algorithm. OHE is used to encode values categorical nominal and SVM-SMOTE for the oversampling. The results of the measurement of the ability of the model generated from the optimized MLP are then compared with the baseline using the AUC score. The data used is the default of credit card client dataset from Taiwan which has 30000 instances. The result of the highest AUC score of the MLP that has gone through optimization is 0.7184, an increase of 0.2179 compared to the baseline. |
|  | @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***  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. 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 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 and leverage several classifiers among the most used in the literature, which are based on different sampling techniques. |
|  | @article{Hjelkrem2022TheVO,    title={The Value of Open Banking Data for Application Credit Scoring: Case Study of a Norwegian Bank},    author={Lars Ole Hjelkrem and Petter Eilif de Lange and Erik Nesset},    journal={Journal of Risk and Financial Management},    year={2022},    url={https://api.semanticscholar.org/CorpusID:254613848}  } | ***The Value of Open Banking Data for Application Credit Scoring: Case Study of a Norwegian Bank***  Banks generally use credit scoring models to assess the creditworthiness of customers when they apply for loans or credit. These models perform significantly worse when used on potential new customers than existing customers, due to the lack of financial behavioral data for new bank customers. Access to such data could therefore increase banks’ profitability when recruiting new customers. If allowed by the customer, Open Banking APIs can provide access to balances and transactions from the past 90 days before the score date. In this study, we compare the performance of conventional application credit scoring models currently in use by a Norwegian bank with a deep learning model trained solely on transaction data available through Open Banking APIs. We evaluate the performance in terms of the AUC and Brier score and find that the models based on Open Banking data alone are surprisingly effective in predicting default compared to the conventional credit scoring models. Furthermore, an ensemble model trained on both traditional credit scoring data and features extracted from the deep learning model further outperforms the conventional application credit scoring model for new customers and narrows the performance gap between application credit scoring models for existing and new customers. Therefore, we argue that banks can increase their profitability by utilizing data available through Open Banking APIs when recruiting new customers. |
|  | @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***  This 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. |
|  | @inproceedings{Amato2022CreditSP,    title={Credit Score Prediction Relying on Machine Learning},    author={Flora Amato and Antonino Ferraro and Antonio Galli and Francesco Moscato and Vincenzo Moscato and Giancarlo Sperl{\'i}},    booktitle={Sistemi Evoluti per Basi di Dati},    year={2022},    url={https://api.semanticscholar.org/CorpusID:251770996}  } | ***Credit Score Prediction Relying on Machine Learning***  Financial institutions use a variety of methodologies to define their commercial and strategic policies, and a significant role is played by credit risk assessment. In recent years, different credit risk assessment services arose, providing Social Lending platforms to connect lenders and borrowers in a direct way without assisting of financial institutions. Despite the pros of these platforms in supporting fundraising process, there are different stems from multiple factors including lack of experience of lenders, missing or uncertain information about the borrower’s credit history. In order to handle these problems, credit risk assessments of financial transactions are usually modeled as a binary problem based on debt repayment, going to apply Machine Learning (ML) techniques. The paper represents an extended abstract of a recent work, where some of the authors performed a benchmarking among the most used credit risk assessment ML models in the field of predicting whether a loan will be repaid in a P2P platform. The experimental analysis is based on a real dataset of Social Lending (Lending Club), going to evaluate several evaluation metrics including AUC, sensitivity, specificity and explainability of the models.  IMG_256 |
|  | @article{Arram2023CreditCS,    title={Credit card score prediction using machine learning models: A new dataset},    author={Anas Arram and Masri Ayob and Musatafa Abbas Abbood Albadr and Alaa Sulaiman and Dheeb Albashish},    journal={ArXiv},    year={2023},    volume={abs/2310.02956},    url={https://api.semanticscholar.org/CorpusID:263620728}  } | ***Credit card score prediction using machine learning models: A new dataset***  The use of credit cards has recently increased, creating an essential need for credit card assessment methods to minimize potential risks. This study investigates the utilization of machine learning (ML) models for credit card default prediction system. The main goal here is to investigate the best-performing ML model for new proposed credit card scoring dataset. This new dataset includes credit card transaction histories and customer profiles, is proposed and tested using a variety of machine learning algorithms, including logistic regression, decision trees, random forests, multi-layer perceptron (MLP) neural network, XGBoost, and LightGBM. To prepare the data for machine learning models, we perform data pre-processing, feature extraction, feature selection, and data balancing techniques. Experimental results demonstrate that MLP outperforms logistic regression, decision trees, random forests, LightGBM, and XGBoost in terms of predictive performance in true positive rate, achieving an impressive area under the curve (AUC) of 86.7% and an accuracy rate of 91.6%, with a recall rate exceeding 80%. These results indicate the superiority of MLP in predicting the default customers and assessing the potential risks. Furthermore, they help banks and other financial institutions in predicting loan defaults at an earlier stage. |
|  | @article{Jakka2023ANC,    title={A novel credit scoring system in financial institutions using artificial intelligence technology},    author={Geethamanikanta Jakka and Amrutanshu Panigrahi and Abhilash Pati and Manmath Nath Das and Jyotsnarani Tripathy},    journal={Journal of Autonomous Intelligence},    year={2023},    url={https://api.semanticscholar.org/CorpusID:261147882}  } | ***A novel credit scoring system in financial institutions using artificial intelligence technology***  In order to evaluate a person’s or a company’s creditworthiness, financial institutions must use credit scoring. Traditional credit scoring algorithms frequently rely on manual and rule-based methods, which can be tedious and inaccurate. Recent developments in artificial intelligence (AI) technology have opened up possibilities for creating more reliable and effective credit rating systems. The data are pre-processed, including scaling using the 0–1 normalization method and resolving missing values by imputation. Information gain (IG), gain ratio (GR), and chi-square are three feature selection methodologies covered in the study. While GR normalizes IG by dividing it by the total entropy of the feature, IG quantifies the reduction in total entropy by adding a new feature. Based on chi-squared statistics, the most vital traits are determined using chi-square. This research employs different ML models to develop a hybrid model for credit score prediction. The ML algorithms support vector machine (SVM), neural networks (NNs), decision trees (DTs), random forest (RF), and logistic regression (LR) classifiers are employed here for experiments along with IG, GR, and chi-square feature selection methodologies for credit prediction over Australian and German datasets. The study offers an understanding of the decision-making process for informative characteristics and the functionality of machine learning (ML) in credit prediction tasks. The empirical analysis shows that in the case of the German dataset, the DT with GR feature selection and hyperparameter optimization outperforms SVM and NN with an accuracy of 99.78%. For the Australian dataset, SVM with GR feature selection outperforms NN and DT with an accuracy of 99.98%. |
|  | @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***  Credit 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. |
|  | @article{Tripathi2021ExperimentalAO,    title={Experimental analysis of machine learning methods for credit score classification},    author={Diwakar Tripathi and Damodar Reddy Edla and Annushree Bablani and Alok Kumar Shukla and B. Ramachandra Reddy},    journal={Progress in Artificial Intelligence},    year={2021},    volume={10},    pages={217 - 243},    url={https://api.semanticscholar.org/CorpusID:233650527}  } | ***Experimental analysis of machine learning methods for credit score classification***  Credit scoring concerns with emerging empirical model to assist the financial institutions for financial decision-making process. Credit risk analysis plays a vital role for decision-making process; statistical and machine learning approaches are utilized to estimate the risk associated with a credit applicant. Enhancing the performance of credit scoring model, particularly toward non-trustworthy “or non-creditworthy” group, may result incredible effect for financial institution. However, credit scoring data may have excess and unimportant data and features which degrades the performance of model. So, selection of important features (or reduction in irrelevant and redundant features) may play the key role for improving the effectiveness and reducing the complexity of the model. This study presents a experimental results analysis of various combinations of feature selection approaches with various classification approaches and impact of feature selection approaches. For experimental results analysis, nine feature selection and sixteen classification state-of-the-art approaches have been applied on seven benched marked credit scoring datasets. |
|  | @article{Roy2021ModellingAS,    title={Modelling a sustainable credit score system (SCSS) using BWM and fuzzy TOPSIS},    author={Pranith Kumar Roy and Krishnendu Shaw},    journal={International Journal of Sustainable Development \& World Ecology},    year={2021},    volume={29},    pages={195 - 208},    url={https://api.semanticscholar.org/CorpusID:237886994}  } | ***Modelling a sustainable credit score system (SCSS) using BWM and fuzzy TOPSIS***  Sustainable development has emerged as a critical agenda for all organisations around the world. Despite the fact that profitability and sustainability are inorganically linked to financial institutions, sustainable lending has been a constant focus of attention. Due to regulatory pressure and stakeholder concerns, financial institutions are forced to implement a variety of sustainable measures; they are also gradually thinking to give more support to socially impactful and sustainable projects. Financial institutions can play a major role in establishing sustainable development by adopting the green lending policy. However, there have been few studies on sustainability credit score systems (SCSS) that take into account social and environmental factors. To fill the gaps of existing literature, this study proposes a multi-criteria SCSS that takes into account the environment and social aspects in addition to financial and managerial aspects. A combined Best-Worst Method (BWM) and the fuzzy-Technique for Order Preferences by Similarity to an Ideal Solution (TOPSIS) method are used in this study to create a credit scoring system. The BWM is used to determine the weight of factors, and the fuzzy-TOPSIS is used to evaluate applicants. The ambiguity while evaluating borrowers has been captured by applying fuzzy set theory. A real-life case study is used to demonstrate the efficacy of the proposed model. The model is unique in terms of the number of social and environmental factors considered. This research will assist financial institutions in identifying borrowers who engage in sustainable business practices. Borrowers can be holistically prioritised by applying the model.  IMG_257 |
|  | @article{Yao2022EnterpriseCR,    title={Enterprise credit risk prediction using supply chain information: A decision tree ensemble model based on the differential sampling rate, Synthetic Minority Oversampling Technique and AdaBoost},    author={Gang Yao and Xiaojian Hu and Taiyun Zhou and Yue Zhang},    journal={Expert Systems},    year={2022},    volume={39},    url={https://api.semanticscholar.org/CorpusID:246715994}  } | ***Enterprise credit risk prediction using supply chain information: A decision tree ensemble model based on the differential sampling rate, Synthetic Minority Oversampling Technique and AdaBoost***  The spread of enterprise credit risk in the supply chain may lead to large‐scale bankruptcy and credit crises, which are related to national economic and social stability and financial system security. Therefore, enterprise credit risk in the supply chain context is not only a concern for banking financial institutions, credit rating agencies and enterprise managers but also the focus of governments. This article develops a DTE‐DSA (decision tree [DT] ensemble model using the differential sampling rate, Synthetic Minority Oversampling Technique [SMOTE] and AdaBoost) prediction framework integrating supply chain information to predict enterprise credit risk. The empirical test shows that using supply chain information can significantly improve the prediction score. The DTE‐DSA model has the best prediction effect in dealing with class imbalance problems. Compared with single classifier models—such as logistic regression, k‐nearest neighbours, support vector machine, DT and DT using the SMOTE—as well as ensemble models—such as extremely randomized trees, random forest, rotation forest, extreme gradient boosting, gradient boosting DT and DT ensemble model using AdaBoost—the DTE‐DSA model not only has the best prediction score but also has a more stable performance. The comprehensive use of supply chain information and the DTE‐DSA model can result in the highest prediction score, with an area under the curve of 0.9016 and a Kolmogorov–Smirnov statistic of 0.7369. Further analysis of the variables of importance enhances the interpretability of the model and obtains relevant management insights. |
|  | @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***  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. |
|  | @article{Patel2021KiRTiAB,    title={KiRTi: A Blockchain-Based Credit Recommender System for Financial Institutions},    author={Shivani Bharatbhai Patel and Pronaya Bhattacharya and Sudeep Tanwar and Neeraj Kumar},    journal={IEEE Transactions on Network Science and Engineering},    year={2021},    volume={8},    pages={1044-1054},    url={https://api.semanticscholar.org/CorpusID:226563789}  } | ***KiRTi: A Blockchain-Based Credit Recommender System for Financial Institutions***  In this paper, we propose KiRTi, a deep-learning-based credit-recommender scheme for public blockchain to facilitate smart lending operations between prospective borrowers (PB) and prospective lenders (PL) to eliminate the need of third party credit-rating agencies (CRAs) for credit-score (CS) generation. Thus loan grants to PB from PL is secured, authorized, and automated so as to expedite the disbursement process. KiRTi stores PB historical transactions, current assets, and liabilities as time-series sequenced data in a public blockchain. The sequenced data is fetched from blockchain by a long-short term memory (LSTM) model that generates CS for loan recommendations based on proposed lending algorithms for PB and PL. To ensure real-time updation of CS, edge-weights are updated based on boolean indicators from PB and PL, which indicates the successful repayments and loan-defaults. The process is iterated to improve the accuracy of edge-weights and generated CS to ensures the correct credibility of PB for future lending. Smart contracts (SC) are proposed for automatic setup of loan repayments between PB and PL. To model the LSTM recommender scheme, a German credit dataset from UCI repository is considered with 1000 credit-histories of PB, with 700 successful repayments and 300 defaults. KiRTi achieves an accuracy of 97.5% in comparison to conventional approaches with an F-measure of 0.98304. The security evaluation of KiRTi shows that it has computation cost of 20.96 ms and communication cost of 121 bytes compared to other state-of-the-art approaches. |
|  | @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***  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. |
|  | @article{Faraji2022ARO,    title={A Review of Machine Learning Applications for Credit Card Fraud Detection with A Case study},    author={Zahra Faraji},    journal={SEISENSE Journal of Management},    year={2022},    url={https://api.semanticscholar.org/CorpusID:246894540}  } | ***A Review of Machine Learning Applications for Credit Card Fraud Detection with A Case study***  Purpose - This paper aims to highlight the widely used supervised techniques applied for fraud detection. In addition, this paper aims to apply some techniques to evaluate their performance on real-world data and develop an ensemble model as a potential solution for this problem. Design/Methodology - Different techniques applied in this study for fraud detection purposes are logistic regression, decision tree, random forest, KNN, and XGBoost. The confusion matrix gives information about the assignment of inputs to the different classes. This study uses precision and recall to evaluate the performance, calculated based on the confusion matrix. Findings- XGBoost is the fastest and is expected to have the best performance; however, it is only outperforming the random forest in terms of accuracy, precision, recall, and f1-score. In general, the KNN and logistic regression have better performance, which means they better detect fraudulent transactions. Practical Implications- The new model can be applied to new data instead of the previous techniques. |
|  | @article{Li2023CreditRP,    title={Credit Risk Prediction Model for Listed Companies Based on CNN-LSTM and Attention Mechanism},    author={Jingyuan Li and Caosen Xu and Bing Feng and Hanyu Zhao},    journal={Electronics},    year={2023},    url={https://api.semanticscholar.org/CorpusID:257899057}  } | ***Credit Risk Prediction Model for Listed Companies Based on CNN-LSTM and Attention Mechanism***  The financial market has been developing rapidly in recent years, and the issue of credit risk concerning listed companies has become increasingly prominent. Therefore, predicting the credit risk of listed companies is an urgent concern for banks, regulators and investors. The commonly used models are the Z-score, Logit (logistic regression model), the kernel-based virtual machine (KVM) and neural network models. However, the results achieved could be more satisfactory. This paper proposes a credit-risk-prediction model for listed companies based on a CNN-LSTM and an attention mechanism, Our approach is based on the benefits of the long short-term memory network (LSTM) model for long-term time-series prediction combined with a convolutional neural network (CNN) model. Furthermore, the advantages of being integrated into a CNN-LSTM model include reducing the complexity of the data, improving the calculation speed and training speed of the model and solving the possible lack of historical data in the long-term sequence prediction of the LSTM model, resulting in prediction accuracy. To reduce problems, we introduced an attention mechanism to assign weights independently and optimize the model. The results show that our model has distinct advantages compared with other CNNs, LSTMs, CNN-LSTMs and other models. The research on the credit-risk prediction of the listing formula has significant meaning. |
|  | @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***  Credit 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. |
|  | @article{Khan2022DevelopingAC,    title={Developing a Credit Card Fraud Detection Model using Machine Learning Approaches},    author={Shahnawaz Khan and Abdullah Alourani and Bharavi Mishra and Ashraf Tahseen Ali and Mustafa Kamal},    journal={International Journal of Advanced Computer Science and Applications},    year={2022},    url={**https://api.semanticscholar.org/CorpusID:247867641**}  } | ***Developing a Credit Card Fraud Detection Model using Machine Learning Approaches***  The growing application and usage of e-commerce applications have given an exponential rise to the number of online transactions. Though there are several methods for completing online transactions, however, credit cards are most commonly used. The increased number of transactions has given the opportunity to the fraudsters to mislead the customers and make them execute fraudulent transactions. Therefore, there is a need for such a method that can automatically classify detect fraudulent transactions. This research study aims to develop a credit-card fraud detection model that can effectively classify an online transaction as fraudulent or genuine. Three supervised machine learning approaches have been applied to develop a credit-card fraud classifier. These techniques include logistic regression, artificial intelligence and support vector machine. The classification accuracy achieved by all the classifiers is almost similar. This research has used the confusion matrix and area under the curve to demonstrate the score of the different performance measures and evaluate the overall performance of the classifiers. Several performance measures such as accuracy, precision, recall, F1-measure, Matthews correlation coefficient, receiver operating characteristic curve have been computed and analysed to evaluate the performance of the credit-card fraud detection classifiers. The analysis demonstrates that the support vector machine-based classifier outperforms the other classifiers. |
|  | @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***  In 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. |
|  | @article{Jammalamadaka2022ResponsibleAI,    title={Responsible AI in automated credit scoring systems},    author={Krishna Ravali Jammalamadaka and Srikanth Itapu},    journal={AI and Ethics},    year={2022},    pages={1-11},    url={https://api.semanticscholar.org/CorpusID:249491866}  } | ***Responsible AI in automated credit scoring systems***  In the last few years, Artificial Intelligence (AI) has achieved a notable momentum that, may deliver the expectations over many application sectors across the field. For this to occur, expert systems and rule-based models need to overcome the limitation of fairness and interpretability. Paradigms underlying this problem fall within the so-called explainable AI (XAI) field. This report presents the work on German credit card dataset to overcome the challenges of fairness, bias and in return, deem the machine learning models giving a responsible expectation. This is defined as responsible AI in practice. Since the dataset we dealt with, is to classify credit score of a user as good or bad, using fair ML modelling approach, the key metric of interest is the F1-score to reduce share of misclassifications. It is observed that hyper parameter tuned XGBoost model (GC2) gives optimal performance in terms of both F1-score, accuracy and fairness for the case of both gender and age as protected variable through Disparate Impact Remover, a pre-processing bias mitigation technique. The same is deployed using both Heroku through Flask API (for age). The Disparate Impact Analysis (DIA) using H2O.AI helped to identify optimum threshold levels at which the fairness metrics are observed at legally acceptable/permissible levels for both age and gender. Overall, fairness, bias responsibility and explainability have been established for the dataset considered. |
|  | @article{Du2022ExplorationOF,    title={Exploration of Financial Market Credit Scoring and Risk Management and Prediction Using Deep Learning and Bionic Algorithm},    author={Peng Du and Hong Shu},    journal={J. Glob. Inf. Manag.},    year={2022},    volume={30},    pages={1-29},    url={https://api.semanticscholar.org/CorpusID:245301019}  } | ***Exploration of Financial Market Credit Scoring and Risk Management and Prediction Using Deep Learning and Bionic Algorithm***  The purpose is to effectively manage the financial market, comprehensive assess personal credit, reduce the risk of financial enterprises. Given the systemic risk problem caused by the lack of credit scoring in the existing financial market, a credit scoring model is put forward based on the deep learning network. The proposed model uses RNN (Recurrent Neural Network) and BRNN (Bidirectional Recurrent Neural Network) to avoid the limitations of shallow models. Afterward, to optimize path analysis, bionic optimization algorithms are introduced, and an integrated deep learning model is proposed. Finally, a financial credit risk management system using the integrated deep learning model is proposed. The probability of default or overdue customers is predicted through verification on three real credit data sets, thus realizing the credit risk management for credit customers. |
|  | @article{Atif2022TheME,    title={The Most Effective Strategy for Incorporating Feature Selection into Credit Risk Assessment},    author={Dalia Atif and Mabrouka Salmi},    journal={SN Computer Science},    year={2022},    volume={4},    pages={1-18},    url={https://api.semanticscholar.org/CorpusID:254851653}  } | ***The Most Effective Strategy for Incorporating Feature Selection into Credit Risk Assessment***  This paper aims to identify the most effective strategy for incorporating feature selection (FS) into credit risk classification, employing three classifiers: Logistic regression (Logreg), Random Forests (RF), and Support Vector Machine (SVM) with the linear kernel through various embedded and wrapper strategies existing in the literature. We performed a comparative analysis on the German Credit dataset using three criteria: classification error rate, stability of selection, and calculation time. According to the Welsh t-test, RFE-RF (Recursive Feature Elimination for RF) outperformed RFE-SVM and penalized Logistic regression, with no significant difference in F1-score for RFE-SVM and suffers from the long-running computation. Conversely, RFE-SVM offers the best stability of 71% with a significantly shorter computation time. Furthermore, the paper intends to introduce a new classification of feature selection strategies in credit risk assessment in light of recent developments. Based on this new classification, a comparison with related literature reveals that the one-stage FS (RFE-RF and RFE-SVM) provides roughly the same accuracy as the two-stage FS and the two-stage classification model and, in some cases, outperforms. |
|  | @article{Baker2022EnsembleLW,    title={Ensemble Learning with Supervised Machine Learning Models to Predict Credit Card Fraud Transactions},    author={Mohammed Rashad Baker and Zuhair Norii Mahmood and Ehab Hashim Shaker},    journal={Revue d'Intelligence Artificielle},    year={2022},    url={https://api.semanticscholar.org/CorpusID:252630717}  } | ***Ensemble Learning with Supervised Machine Learning Models to Predict Credit Card Fraud Transactions***  In recent years, the highly boosting development in e-commerce technologies made it possible for people to select the most desirable items from shops and stores worldwide while being at home. Credit card frauds transactions are common nowadays because of online payments. Online transactions are the root cause of fraudulent credit card activity, bringing enormous financial losses. Financial institutions must install an automatic deterrent mechanism to check these fraudulent actions. The fraudulent transactions do not follow a specific pattern and continuously change their shape and behavior. This paper aims to use ensemble learning with supervised Machine Learning (ML) models to predict the occurrence of fraud transactions. The experimental study has been evaluated on the open-source Kaggle credit card fraud detection dataset. The performance of the proposed model is measured in terms of accuracy score, confusion matrix, and classification report. The results were state-of-the-art using the voting ensemble learning technique shows that it can be get the best results using PCA with 100.0% accuracy, 97.3% precision, 73.5% recall, and 83.7% f1-score against other ML classifiers. |
|  | @article{Leevy2023ThresholdOA,    title={Threshold optimization and random undersampling for imbalanced credit card data},    author={Joffrey L. Leevy and Justin M. Johnson and John T. Hancock and Taghi M. Khoshgoftaar},    journal={Journal of Big Data},    year={2023},    volume={10},    pages={1-22},    url={https://api.semanticscholar.org/CorpusID:258509782}  } | ***Threshold optimization and random undersampling for imbalanced credit card data***  utput thresholding is well-suited for addressing class imbalance, since the technique does not increase dataset size, run the risk of discarding important instances, or modify an existing learner. Through the use of the Credit Card Fraud Detection Dataset, this study proposes a threshold optimization approach that factors in the constraint True Positive Rate (TPR) ≥  True Negative Rate (TNR). Our findings indicate that an increase of the Area Under the Precision–Recall Curve (AUPRC) score is associated with an improvement in threshold-based classification scores, while an increase of positive class prior probability causes optimal thresholds to increase. In addition, we discovered that best overall results for the selection of an optimal threshold are obtained without the use of Random Undersampling (RUS). Furthermore, with the exception of AUPRC, we established that the default threshold yields good performance scores at a balanced class ratio. Our evaluation of four threshold optimization techniques, eight threshold-dependent metrics, and two threshold-agnostic metrics defines the uniqueness of this research. |
|  | @article{Agosto2023BayesianLM,    title={Bayesian learning models to measure the relative impact of ESG factors on credit ratings},    author={Arianna Agosto and Paola Cerchiello and Paolo Giudici},    journal={International Journal of Data Science and Analytics},    year={2023},    pages={1-12},    url={https://api.semanticscholar.org/CorpusID:259326371}  } | ***Bayesian learning models to measure the relative impact of ESG factors on credit ratings***  Artificial intelligence methods, based on machine learning models, are rapidly changing financial services, and credit lending in particular, complementing traditional bank lending with platform lending. While financial technologies improve user experience and possibly lower costs, they may increase risks and, in particular, the model risks that derive from inaccurate credit rating assessments. In this paper, we will show how to reduce such model risks, using a S.A.F.E. statistical learning model, which improves: Sustainability, taking environmental, social and governance factors into account; Accuracy, building a model which maximises predictive accuracy; Fairness, merging ESG scores from different data providers, improving their representativeness; Explainability, clarifying the relative contribution of each ESG score to predictive accuracy. |
|  | @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***  Aim: 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. |
|  | @article{Aftab2023FraudDO,    title={Fraud Detection of Credit Cards Using Supervised Machine Learning Techniques},    author={Ammar Aftab and Iqra Shahzad and Amna Sajid and Maira Anwar and Nosheen Anwar},    journal={Pakistan Journal of Emerging Science and Technologies (PJEST)},    year={2023},    url={https://api.semanticscholar.org/CorpusID:259579277}  } | ***Fraud Detection of Credit Cards Using Supervised Machine Learning Techniques***  Credit card fraud encompasses illicit activities aimed at unlawfully obtaining confidential information to enable unauthorized individuals to engage in illegal transactions. As technology advances, fraudsters have honed their skills in evading security measures, presenting a formidable challenge in fraud detection. To address this issue, an array of algorithms and analytical techniques has emerged to identify and mitigate instances of fraud. This research aimed to identify the most appropriate supervised machine learning algorithm for credit card fraud detection. Logistic Regression, Random Forest, Support Vector Machine, and Decision Trees were implemented and compared. Due to the imbalanced nature of the dataset, the SMOTE (Synthetic Minority Oversampling Technique) technique was employed to rectify the data imbalance by oversampling the minority class. The performance of the trained models was evaluated using various metrics, including the confusion matrix, accuracy, precision, recall, f1-score, Matthews Correlation Coefficient (MCC), and Area Under the Curve (AUC). The results of the analysis revealed that Random Forests exhibited exceptional performance, achieving an impressive recall score of 84% and surpassing other algorithms. This research provides the groundwork for future investigations involving diverse deep-learning techniques applied to real-time and dynamic datasets, enabling continuous enhancements in fraud detection and prevention mechanisms. |
|  | @article{Yarahmadi2023MultiAgentCA,    title={Multi-Agent Credit Assignment and Bankruptcy Game for Improving Resource Allocation in Smart Cities},    author={Hossein Yarahmadi and Mohammad Ebrahim Shiri and Moharram Challenger and Hamidreza Navidi and Arash Sharifi},    journal={Sensors (Basel, Switzerland)},    year={2023},    volume={23},    url={https://api.semanticscholar.org/CorpusID:256658447}  } | ***Multi-Agent Credit Assignment and Bankruptcy Game for Improving Resource Allocation in Smart Cities***  In recent years, the development of smart cities has accelerated. There are several issues to handle in smart cities, one of the most important of which is efficient resource allocation. For the modeling of smart cities, multi-agent systems (MASs) can be used. In this paper, an efficient approach is proposed for resource allocation in smart cities based on the multi-agent credit assignment problem (MCA) and bankruptcy game. To this end, the resource allocation problem is mapped to MCA and the bankruptcy game. To solve this problem, first, a task start threshold (TST) constraint is introduced. The MCA turns into a bankruptcy problem upon introducing such a constraint. Therefore, based on the concept of bankruptcy, three methods of TS-Only, TS + MAS, and TS + ExAg are presented to solve the MCA. In addition, this work introduces a multi-score problem (MSP) in which a different reward is offered for solving each part of the problem, and we used it in our experiments to examine the proposed methods. The proposed approach is evaluated based on the learning rate, confidence, expertness, efficiency, certainty, and correctness parameters. The results reveal the better performance of the proposed approach compared to the existing methods in five parameters. |
|  | @article{SalomoLeuwol2023ImplementationOG,    title={Implementation of Gradient Boosted Tree, Support Vector Machinery and Random Forest Algorithm to Detecting Financial Fraud in Credit Card Transactions},    author={Ferdinand Salomo Leuwol and Asri Ady Bakri and Muhsin N. Bailusy and Hari Setia Putra and Ni Ketut Sukanti},    journal={Jurnal Informasi dan Teknologi},    year={2023},    url={https://api.semanticscholar.org/CorpusID:263643475}  } | ***Implementation of Gradient Boosted Tree, Support Vector Machinery and Random Forest Algorithm to Detecting Financial Fraud in Credit Card Transactions***  According to Google Trends data, machine learning-based credit card identification has grown over the last five years, at the very least, across all nations. In order to detect credit card fraud in this study, the authors will use machine learning methods such random forests, support vector machines, and gradient-boosted trees. The authors used the Synthetic Minority Oversampling Technique (SMOTE) and Random Under Sampling (RUS) sampling methods in each algorithm to compare because there was a class imbalance in this investigation. The research findings demonstrate that the author's algorithm and sample technique were successfully used, as shown by the AUC values obtained for each being > 0.7. The top score in RUS was 0.7835 using the Random Forest algorithm, whereas the greatest score in SMOTE was 0.73 with the Gradient Boosted Trees approach. The Random Forest algorithm and the Random Under Sampling (RUS) technique are developed as a result of this research, and they are useful for identifying fraudulent credit card transactions. |
|  | @article{Xia2022CreditCF,    title={Credit Card Fraud Detection Based on Support Vector Machine},    author={Jia-zhi Xia},    journal={Highlights in Science, Engineering and Technology},    year={2022},    url={https://api.semanticscholar.org/CorpusID:254709190}  } | ***Credit Card Fraud Detection Based on Support Vector Machine***  Due to the increasing popularity cashless transactions, credit card fraud has become one of the most common frauds and caused huge harm to the financial institutions and individuals in real life. In this academic paper, the algorithm Support Vector Machine (SVM) is used to build models to deal with the credit card fraud detection problem with the performance metrics AUC and F1-score. The experiment dataset is named Credit Card Transactions Fraud Detection Dataset from the Kaggle website. After the step of preprocessing, the dataset is split into the training, testing and validation dataset with 11 numerical features and a label feature called “is\_fraud”. The inner parameter “class\_weight” of the SVM algorithm in Python is set as “balanced” to deal with the imbalanced datasets. The main method to find the optimized models is using the GridSearchCV function in Python library sklearn. After tuning the hyperparameters and handling the overfitting phenomenon, the optimized models for the two metrics are found. The parameter values of the best model for AUC are C=10, class\_weight= “balanced”, g =0.01, kernel = “rbf”. The training AUC is 0.87 and testing AUC is 0.90. The parameter values of the final optimized model for F1-score are C=0.8, class\_weight= “balanced”, g =0.06, kernel = “rbf”. The final training F-score is 0.305 and testing F-score is 0.260. |
|  | @article{Altman2023TheOS,    title={The Omega Score: An improved tool for SME default predictions},    author={Edward I. Altman and Marco Balzano and Alessandro Giannozzi and Stjepan Srhoj},    journal={Journal of the International Council for Small Business},    year={2023},    volume={4},    pages={362 - 373},    url={https://api.semanticscholar.org/CorpusID:258067962}  } | ***The Omega Score: An improved tool for SME default predictions***  The Omega Score, a novel small and medium-sized enterprise (SME) default predictor developed by Altman et al. in 2022, combines indicators related to financial ratios, payment behavior, and management and employees variables that play an important role in predicting SME defaults. Built with machine-learning techniques and rich dataset information, the Omega Score can be used to categorize an SME into one of the following three groups: healthy, moderate-risk, and high-risk. The Omega Score can be utilized by financial institutions to reduce lending errors and minimize loan defaults, support policy makers in implementing effective restructuring policies, assist credit analytics firms in assessing creditworthiness, assist investors in allocating funds, and asset managers to support decision-making processes. |
|  | @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***  Determination 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. |