|  |  |  |
| --- | --- | --- |
| **stt** |  |  |
| **1.** | @article{Liu2019ResearchOP,  title={Research on Personal Credit Scoring Model Based on Artificial Intelligence},  author={Chengyong Liu and Huang-Chen Huang and Shangqian Lu},  journal={Application of Intelligent Systems in Multi-modal Information Analytics},  year={2019},  } | Artificial intelligence is considered to be the technological commanding height of the next era. At present, after the development of China’s artificial intelligence industry ranks in the United States, its application in the financial field is also in a new stage of rapid development, and affects many aspects of the financial industry, thus strengthening its research is of great significance. The continuous development of artificial intelligence technology has been widely used in many aspects of financial services, which is of great significance for the realization of its modeling, standardization and intelligent development. However, there are still security risks hidden in the application, which requires attention to this. aspects of the research to identify effective measures for risk prevention, this paper analyzes the application of artificial intelligence in the financial sector in the personal credit score. |
| **2.** | @article{Gahlaut2017PredictionAO,  title={Prediction analysis of risky credit using Data mining classification models},  author={Archana Gahlaut and Tushar and Prince Kumar Singh},  journal={2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)},  year={2017},  pages={1-7},  } | Gaining as many good credit scores are beneficial for customers in numerous ways and it also allows banks to analyse their clients and to give credit loans to them accordingly. In this paper, we look whether data mining techniques are useful to predict and classify the customer's credit score (good/bad) to overcome the future risks giving loans to clients who cannot repay. We use historical given dataset of a bank for our predictive modelling (general models), banks can use them for the better outcome of their overall credit system. For example, if a customer is assigned a bad credit score after applying these predictive classification models, then the bank will not allow giving that customer a future credit and will quickly analyse all the other risky credits. |
| **3.** | @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},  } | 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 |
| **4.** | @article{Brevoort2016CreditIA,  title={Credit Invisibles and the Unscored},  author={Kenneth P. Brevoort and Philipp H. Grimm and Michelle Kambara},  journal={Household Finance eJournal},  year={2016},  } | Having a credit record and a credit score can be an important determinant of credit access. Yet, surprisingly little is known about people who lack credit records or scores. This paper provides the first documented analysis of the characteristics of consumers without credit records, called "credit invisibles," and of consumers whose records are treated as "unscorable" by a widely used credit-scoring model. Our estimates suggest that 26 million adults, representing about 11 percent of the adult population, lack credit records. An additional 8.3 percent, or 19.6 million adults have credit records that are unscored. The incidence of having a credit record is not evenly distributed. Young, elderly, minority, and lower-income consumers are more likely to be credit invisible or have an unscored record. Additionally, our analysis finds that observable credit performance is not widely available for such consumers, which may hinder the ability of alternative data to expand credit access for these consumers. |
| **5.** | @article{Burke2019CreditBO,  title={Credit Building or Credit Crumbling? A Credit Builder Loan's Effects on Consumer Behavior, Credit Scores and Their Predictive Power},  author={Jeremy Burke and Julian C. Jamison and Dean S. Karlan and Kata Mihaly and Jonathan Zinman},  journal={Household Finance eJournal},  year={2019},  } | There is little evidence on how the large market for credit score improvement products affects consumers or credit market efficiency. A randomized encouragement design on a standard credit builder loan (CBL) identifies null average effects on whether consumers have a credit score and the score itself, with important heterogeneity: those with loans outstanding at baseline fare worse, those without fare better. Selection, treatment effect, and prediction models indicate the CBL reveals valuable information to markets, inducing positive selection and making credit histories more precise, while keeping credit scores' predictive power intact. With modest targeting changes, CBLs could work as intended. |
| **6.** | @article{Herkenhoff2020TheIO,  title={The Impact of Consumer Credit Access on Self-Employment and Entrepreneurship},  author={Kyle F. Herkenhoff and G. Phillips and Ethan Cohen-Cole},  journal={Tuck School of Business at Dartmouth Research Paper Series},  year={2020},  } | Abstract We examine how consumer credit affects entrepreneurship by linking three million earnings and pass-through tax records to credit reports. In the cross-section, we show that self-employment without employees and employer firm ownership increase monotonically with credit limits and credit scores. We then isolate individuals who have had discrete increases in credit limits after the exogenous removal of bankruptcy flags to measure the effects of personal credit on entrepreneurship. Following bankruptcy flag removal, individuals are more likely to start a new employer business and borrow extensively. Those who own businesses with employees borrow $40,000 more after bankruptcy flag removal, a 33% gain relative to the sample average. |
| **7.** | @article{Herkenhoff2020TheIO,  title={The Impact of Consumer Credit Access on Self-Employment and Entrepreneurship},  author={Kyle F. Herkenhoff and G. Phillips and Ethan Cohen-Cole},  journal={Tuck School of Business at Dartmouth Research Paper Series},  year={2020},  } | Abstract We examine how consumer credit affects entrepreneurship by linking three million earnings and pass-through tax records to credit reports. In the cross-section, we show that self-employment without employees and employer firm ownership increase monotonically with credit limits and credit scores. We then isolate individuals who have had discrete increases in credit limits after the exogenous removal of bankruptcy flags to measure the effects of personal credit on entrepreneurship. Following bankruptcy flag removal, individuals are more likely to start a new employer business and borrow extensively. Those who own businesses with employees borrow $40,000 more after bankruptcy flag removal, a 33% gain relative to the sample average. |
| **8.** | @article{Wang2022ADL,  title={A Deep Learning Approach for Credit Scoring Using Feature Embedded Transformer},  author={Chongren Wang and Zhuoyi Xiao},  journal={Applied Sciences},  year={2022},  } | In this paper, we introduce a transformer into the field of credit scoring based on user online behavioral data and develop an end-to-end feature embedded transformer (FE-Transformer) credit scoring approach. The FE-Transformer neural network is composed of two parts: a wide part and a deep part. The deep part uses the transformer deep neural network. The output of the deep neural network and the feature data of the wide part are concentrated in a fusion layer. The experimental results show that the FE-Transformer deep learning model proposed in this paper outperforms the LR, XGBoost, LSTM, and AM-LSTM comparison methods in terms of area under the receiver operating characteristic curve (AUC) and the Kolmogorov–Smirnov (KS). This shows that the FE-Transformer deep learning model proposed in this paper can accurately predict user default risk. |
| **9.** | @article{Bussmann2019ExplainableML,  title={Explainable Machine Learning in Credit Risk Management},  author={Niklas Bussmann and Paolo Giudici and Dimitri Marinelli and Jochen Papenbrock},  journal={Computational Economics},  year={2019},  volume={57},  pages={203 - 216},  } | The paper proposes an explainable Artificial Intelligence model that can be used in credit risk management and, in particular, in measuring the risks that arise when credit is borrowed employing peer to peer lending platforms. The model applies correlation networks to Shapley values so that Artificial Intelligence predictions are grouped according to the similarity in the underlying explanations. The empirical analysis of 15,000 small and medium companies asking for credit reveals that both risky and not risky borrowers can be grouped according to a set of similar financial characteristics, which can be employed to explain their credit score and, therefore, to predict their future behaviour. |
| **10.** | @article{Zeng2020OnTC,  title={On the confusion matrix in credit scoring and its analytical properties},  author={Guoping Zeng},  journal={Communications in Statistics - Theory and Methods},  year={2020},  volume={49},  pages={2080 - 2093},  } | Abstract Confusion Matrix is an important measure to evaluate the accuracy of credit scoring models. However, the literature about Confusion Matrix is limited. The analytical properties of Confusion Matrix are ignored. Moreover, the concept of Confusion Matrix is confusing. In this article, we systematically study Confusion Matrix and its analytical properties. We enumerate 16 possible variants of Confusion Matrix and show that only 8 are reasonable. We study the relationship between Confusion Matrix and 2 other performance measures: the receiver operating characteristic curve (ROC) and Kolmogorov-Smirnov statistic (KS). We show that an optimal cutoff score can be attained by KS. |
| **11.** | @article{Bhutta2021HowMD,  title={How Much Does Racial Bias Affect Mortgage Lending? Evidence from Human and Algorithmic Credit Decisions},  author={Neil Bhutta and Aurel Hizmo and Daniel R. Ringo},  journal={ERN: Microeconometric Studies of Housing Markets (Topic)},  year={2021},  } | We assess racial discrimination in mortgage approvals using new data on mortgage applications. Minority applicants tend to have significantly lower credit scores, higher leverage, and are less likely than white applicants to receive algorithmic approval from race-blind government automated underwriting systems (AUS). Observable applicant-risk factors explain most of the racial disparities in lender denials. Further, we exploit the AUS data to show there are risk factors we do not directly observe, and our analysis indicates that these factors explain at least some of the residual 1-2 percentage point denial gaps. Overall, we find that differential treatment has played a limited role in generating denial disparities in recent years. |
| **12.** | @article{Alali2012TheEO,  title={The Effect of Corporate Governance on Firm’s Credit Ratings: Further Evidence Using Governance Score in the United States},  author={Fatima Alali and Asokan Anandarajan and Wei Jiang},  journal={Auditing},  year={2012},  } | We investigate whether corporate governance affects firms’ credit ratings and whether improvement in corporate governance standards is associated with improvement in investment grade rating. We use the Gov‐score of Brown and Caylor (2006), the Gomper’s G index and an entrenchment score of Bebchuk et al. (2009) to proxy for corporate governance. Using a sample of US firms, we find that firms characterized by stronger corporate governance have a significantly higher credit rating, and that this association is accentuated for smaller firms relative to larger firms. We find that an improvement in corporate governance is associated with improvement in bond rating. |
| **13.** | @article{Hassija2020SecureLB,  title={Secure Lending: Blockchain and Prospect Theory-Based Decentralized Credit Scoring Model},  author={Vikas Hassija and Gaurang Bansal and Vinay Chamola and Neeraj Kumar and Mohsen Guizani},  journal={IEEE Transactions on Network Science and Engineering},  year={2020},  volume={7},  pages={2566-2575},  } | Credit scoring is a rigorous statistical analysis carried out by lenders and other third parties to access an individual's creditworthiness. Lenders use credit scoring to estimate the degree of risk in lending money to an individual. However, credit score evaluation is primarily based on a transaction record, payment history, professional background, etc. sourced from different credit bureaus. So, evaluating a credit score is a laborious and tedious task involving a lot of paperwork. In this paper, we propose how blockchain can provide the solution to decentralized credit scoring evaluation and reducing the amount of dependence of paperwork. Lending money is not always objective but subjective to every lender. The decision of lending involves different levels of risk and uncertainty, depending on their perspective. This paper uses the prospect theory to model the optimal investment strategy for different risk vs. return scenarios. |
| **14.** | @article{Ampountolas2021AML,  title={A Machine Learning Approach for Micro-Credit Scoring},  author={Apostolos Ampountolas and Titus Nyarko Nde and Paresh Date and Corina Constantinescu},  journal={Risks},  year={2021},  } | In micro-lending markets, lack of recorded credit history is a significant impediment to assessing individual borrowers’ creditworthiness and therefore deciding fair interest rates. This research compares various machine learning algorithms on real micro-lending data to test their efficacy at classifying borrowers into various credit categories. We demonstrate that off-the-shelf multi-class classifiers such as random forest algorithms can perform this task very well, using readily available data about customers (such as age, occupation, and location). This presents inexpensive and reliable means to micro-lending institutions around the developing world with which to assess creditworthiness in the absence of credit history or central credit databases. |
| **15.** | @article{Robb2018TestingFR,  title={Testing for racial bias in business credit scores},  author={Alicia Robb and David T. Robinson},  journal={Small Business Economics},  year={2018},  volume={50},  pages={429-443},  } | We develop a novel empirical test of racial bias based on comparisons between forward-looking, expectations-based credit scores and backward-looking, repayment-history-based credit scores. We then test for racial bias using confidential-access data from the Kauffman Firm Survey. Businesses founded by disadvantaged minorities have much lower average business credit scores, but these scores show no evidence of racial bias. If anything, forward-looking credit-score models under-predict the rate of payment delinquency among minority-owned businesses. |
| **16.** | @article{Oh2009EvaluationOC,  title={Evaluation of credit guarantee policy using propensity score matching},  author={Inha Oh and Jeong-Dong Lee and Almas Heshmati and Gyoung-Gyu Choi},  journal={Small Business Economics},  year={2009},  volume={33},  pages={335-351},  } | In this article, we evaluate the effect of the credit guarantee policy by comparing a large sample of guaranteed firms and matched non-guaranteed firms from 2000 to 2003. The sample firms are compared with respect to growth rates of different performance indicators including: productivity, sales, employment, investment, R&D, wage level, and the survival of firms in the post crisis period. In order to avoid the selectivity problem, propensity score matching methodologies are adopted. Results suggest that credit guarantees influenced significantly firms’ ability to maintain their size, and increase their survival rate, but not to increase their R&D and investment and hence, their growth in productivity. Moreover, due to the adverse selection problem, firms with lower productivity were receiving guarantees. |
| **17.** | @article{Jiang2020DecipheringBD,  title={Deciphering Big Data in Consumer Credit Evaluation},  author={Jinglin Jiang and Li Liao and Xi Lu and Zhengwei Wang and Hongyu Xiang},  journal={International Political Economy: Investment \& Finance eJournal},  year={2020},  } | Abstract This paper examines the impact of large-scale alternative data on predicting consumer delinquency. Using a proprietary double-blinded test from a traditional lender, we find that the big data credit score predicts an individual’s likelihood of defaulting on a loan with 18.4% greater accuracy than the lender’s internal score. Moreover, the impact of the big data credit score is more significant when evaluating borrowers without public credit records. We also provide evidence that big data have the potential to correct financial misreporting. |
| **18.** | @article{Puri2018OnTR,  title={On the Rise of FinTechs – Credit Scoring Using Digital Footprints},  author={Manju Puri and Tobias Berg and Valentin Burg and Ana Gombovi{\'c}},  journal={Risk Management eJournal},  year={2018},  } | We analyze the information content of a digital footprint—that is, information that users leave online simply by accessing or registering on a Web site—for predicting consumer default. We show that even simple, easily accessible variables from a digital footprint match the information content of credit bureau scores. A digital footprint complements rather than substitutes for credit bureau information and affects access to credit and reduces default rates. We discuss the implications for financial intermediaries’ business models, access to credit for the unbanked, and the behavior of consumers, firms, and regulators in the digital sphere. (JEL G20, G21, G29) |
| **19.** | @article{Addi2020AnOM,  title={An Ontology-Based Model for Credit Scoring Knowledge in Microfinance: Towards a Better Decision Making},  author={Khaoula Ben Addi and Nissrine Souissi},  journal={2020 IEEE 10th International Conference on Intelligent Systems (IS)},  year={2020},  pages={380-385},  } | In developing countries, microfinance actors mobilized during the covid-19 pandemic to support the activities of their most vulnerable clients. In this context, the main concern of microfinance institutions is to minimize credit risk by adopting the most reliable scoring system possible. There are many dimensions to consider. In the literature, credit-scoring models essentially base on the financial dimension and neglect others deemed relevant. The study presented in this paper is based on a review of several models to identify aspects related to credit score in a microfinance context, in order to build an ontological model presenting the dimensions having an impact on credit score and their interrelations. The proposed model will help these institutions in their decision-making and in particular in the evaluation of the granting of loans. |
| **20.** | @article{Addi2020AnOM,  title={An Ontology-Based Model for Credit Scoring Knowledge in Microfinance: Towards a Better Decision Making},  author={Khaoula Ben Addi and Nissrine Souissi},  journal={2020 IEEE 10th International Conference on Intelligent Systems (IS)},  year={2020},  pages={380-385},  } | In developing countries, microfinance actors mobilized during the covid-19 pandemic to support the activities of their most vulnerable clients. In this context, the main concern of microfinance institutions is to minimize credit risk by adopting the most reliable scoring system possible. There are many dimensions to consider. In the literature, credit-scoring models essentially base on the financial dimension and neglect others deemed relevant. The study presented in this paper is based on a review of several models to identify aspects related to credit score in a microfinance context, in order to build an ontological model presenting the dimensions having an impact on credit score and their interrelations. The proposed model will help these institutions in their decision-making and in particular in the evaluation of the granting of loans. |
| **21.** | @article{Qi2021ApplicationOE,  title={Application of explainable machine learning based on Catboost in credit scoring},  author={Ji Qi and Ruicheng Yang and Pucong Wang},  journal={Journal of Physics: Conference Series},  year={2021},  volume={1955},  } | Credit scoring is the core part of an institution’s lending. As artificial intelligence is used in various fields, credit rating is also under the same topic of accepting technological changes. Combining credit evaluation and machine learning can incorporate relatively comprehensive features into the credit evaluation process. Through the excellent performance of Catboost, while ensuring accuracy, it demonstrates the explainability of the model as much as possible, avoiding the traditional trust problem of the black-box model. Explainability is proposed to the machine learning model, which reduces the difficulty of processing large amounts of data and the threshold for non-professionals to understand the model. In this article, the dataset is the personal loan data of LendingClub obtained through python. By analyzing the data through Catboost, we can derive excellent results in applying the explainability of machine learning in personal credit evaluation. |
| **22.** | @article{Kyeong2021CanSL,  title={Can System Logs Enhance the Performance of Credit Scoring? – Evidence from an Internet Bank in Korea},  author={Sunghyun Kyeong and Daehee Kim and Jinho Shin},  journal={Sustainability},  year={2021},  } | This study is the first to examine whether the performance of credit rating, one of the most important data-based decision-making of banks, can be improved by using banking system log data that is extensively accumulated inside the bank for system operation. This study uses the log data recorded for the mobile app system of Kakaobank, a leading internet bank used by more than 14 million people in Korea. After generating candidate variables from Kakaobank's vast log data, we develop a credit scoring model by utilizing variables with high information values. Consequently, the discrimination power of the new model compared to the credit bureau grades was significantly improved by 1.84% points based on the Kolmogorov–Smirnov statistics. Therefore, the results of this study imply that if a bank utilizes its log data that have already been extensively accumulated inside the bank, decision-making systems, including credit scoring, can be efficiently improved at a low cost. |
| **23.** | @article{Njuguna2021PosterAS,  title={Poster: A Scoping Review of Alternative Credit Scoring Literature},  author={Rebecca G Njuguna and Karen Sowon},  journal={ACM SIGCAS Conference on Computing and Sustainable Societies},  year={2021},  } | This paper covers a scoping review to establish the breadth of alternative credit scoring literature. The field is nascent and gaining popularity due to the crucial role alternative data is playing to accelerate financial inclusion. Historically, evaluating creditworthiness required availability of past financial activity such as loan repayment. Such stringent requirements rendered people with little or no financial history ‘credit invisible’. Advancements in Artificial Intelligence and Machine Learning have enabled scoring algorithms to work with non-financial data such as digital footprints from mobile devices and psychometric data to compute credit scores. Although the largest portion of ‘credit invisibles’ are in developing economies, research in the area is predominantly originating from developed economies and most alternative credit scoring models are trained with data from developed economies. There is need for more research from developing contexts and utilization of alternative data from populations with a smaller digital footprint. |
| **24.** | @article{Levinger2011TheCO,  title={The Cost of Not Knowing the Score: Self-Estimated Credit Scores and Financial Outcomes},  author={Benjamin Levinger and Marques Benton and Stephan Meier},  journal={Journal of Family and Economic Issues},  year={2011},  volume={32},  pages={566-585},  } | This study analyzes consumers’ knowledge of their own credit situation and tests whether a lack of knowledge affects financial outcomes. The unique dataset from survey and credit report data includes self-estimates of credit scores and actual scores from a low-to-moderate income sample. We argue and show empirically that many respondents don’t know their credit score and generally underestimate their creditworthiness. Furthermore, our evidence suggests that this biased self-assessment may explain differences in perceived credit constraints and credit contracts, specifically credit card interest rates. Our research suggests that an important aspect of financial literacy is self-assessment, and that it is important to encourage consumers to regularly check their credit reports and scores so as to better understand their actual creditworthiness. |
| **25.** | @article{Anderson2021RetailC,  title={Retail Credit},  author={Raymond A. Anderson},  journal={Credit Intelligence \& Modelling},  year={2021},  } | This chapter covers retail credit, which has different data and modelling needs than wholesale. (1) Scorecard terminology—presented is a points-based model (other forms are acknowledged). The goal is to identify rare events, e.g. loan defaults, liquidations, bankruptcies or other undesirable outcomes. (2) Retail models—types across the credit cycle {solicitation, origination, collection, recovery, fraud}, what is being measured {risk, response, retention, revenue}, whose data is used {bespoke, generic, pooled, borrowed} and how it is done {empirical, hybrid, expert judgment}. (3) Data sources—focus is on credit bureaux and credit registries, their spread across various countries, ownership types of credit bureaux and some behind their establishment and spread. (4) Risk indicators—presentation of scores to end-users or downstream processes, as distinct from risk grades. (5) FICO scores—provided by major credit bureaux, with details of different versions and types, plus an imperfect formula for converting their scores into probabilities. |
| **26.** | @article{Stango2016BorrowingHV,  title={Borrowing High versus Borrowing Higher: Price Dispersion and Shopping Behavior in the U.S. Credit Card Market},  author={Victor Stango and Jonathan Zinman},  journal={Review of Financial Studies},  year={2016},  volume={29},  pages={979-1006},  } | We document substantial cross-individual dispersion in U.S. credit card borrowing costs, even after controlling for borrower risk and card characteristics. That remaining dispersion arises because cross-lender pricing heterogeneity generates dispersion in annual percentage rate (APR) offers to borrowers, and borrowers vary in shopping intensity. Our empirics match administrative data to self-reported card shopping intensity and use instruments suggested by fair lending law to account for the endogeneity between APRs and search. The results show that shoppers versus nonshoppers pay APRs as different as those paid by borrowers in the best versus worst credit score deciles. We discuss implications for policy and practice. Received August 2, 2014; accepted July 7, 2015 by Editor Philip Strahan. |
| **27.** | @article{Giudici2020NetworkBC,  title={Network based credit risk models},  author={Paolo Giudici and Branka Hadji-Misheva and Alessandro Spelta},  journal={Quality Engineering},  year={2020},  volume={32},  pages={199 - 211},  } | Abstract Peer-to-Peer lending platforms may lead to cost reduction, and to an improved user experience. These improvements may come at the price of inaccurate credit risk measurements, which can hamper lenders and endanger the stability of a financial system. In the article, we propose how to improve credit risk accuracy of peer to peer platforms and, specifically, of those who lend to small and medium enterprises. To achieve this goal, we propose to augment traditional credit scoring methods with “alternative data” that consist of centrality measures derived from similarity networks among borrowers, deduced from their financial ratios. Our empirical findings suggest that the proposed approach improves predictive accuracy as well as model explainability. |
| **28.** | @article{Agarwal2019FinancialIA,  title={Financial Inclusion and Alternate Credit Scoring: Role of Big Data and Machine Learning in Fintech},  author={Sumit Agarwal and Shashwat Alok and Pulak Ghosh and Sudip Gupta},  journal={FEN: Behavioral Finance (Topic)},  year={2019},  } | We use unique and proprietary data from a large Fintech lender to analyze whether alternative data captured from an individual’s mobile phone (mobile/social footprint) can substitute for traditional credit bureau scores. Variables that measure a borrowers’ digital presence, such as the number and types of apps installed, crude measures of social connections, and measures of borrowers’ “deep social footprints” based on call logs, significantly improve default prediction and outperform the credit bureau score. Using machine learning-based prediction counterfactual analysis, we show that alternate credit scoring based on the mobile and social footprints can expand credit access for individuals who lack credit scores without adversely impacting the default outcomes. Our analysis suggests that the marginal benefit of using alternative data for credit decisions are likely to be higher for borrowers with low levels of income and education, as well as borrowers residing in regions with low levels of financial inclusion. |
| **29.** | @article{Jang2020ESGSA,  title={ESG Scores and the Credit Market},  author={Ga‐Young Jang and Hyoung-Goo Kang and Ju-Yeong Lee and Kyounghun Bae},  journal={Sustainability},  year={2020},  } | This study analyzes the relationship between Environmental, Social and Governance (ESG) scores and bond returns using the corporate bond data in Korea during the period of 2010 to 2015. We find that ESG scores include valuable information about the downside risk of firms. This effect is particularly salient for the firms with high information asymmetry such as small firms. Interestingly, of the three ESG criteria, only environmental scores show a significant impact on bond returns when interacted with the firm size, suggesting that high environmental scores lower the cost of debt financing for small firms. Finally, ESG is complementary to credit ratings in assessing credit quality as credit ratings cannot explain away ESG effects in predicting future bond returns. This result suggests that credit rating agencies should either integrate ESG scores into their current rating process or produce separate ESG scores which bond investors integrate with the existing credit ratings by themselves. |
| **30.** | @article{Smith2010StabilityIC,  title={Stability in Consumer Credit Scores: Level and Direction of FICO Score Drift as a Precursor to Mortgage Default and Prepayment},  author={Brent C. Smith},  journal={LSN: Other Consumer Reporting (Sub-Topic)},  year={2010},  } | This article represents an extension of the expansive credit risk and credit migration literature, prominent in the corporate bond and securities risk pricing literature, to an analysis of the drift of consumer credit scores. A rich data set of residential mortgages is used to observe credit score migration post loan origination and in a test of the ability of credit score transition to serve as a precursor to potential default and prepayment. The results indicate credit scores provide signals and information to investors and servicing agents in a fashion similar to credit ratings on commercial paper as to default potential. |

**NÂNG CAO HIỆU QUẢ PHÁT HIỆN RỦI RO TÍN DỤNG VỚI MÔ HÌNH MÁY HỌC**

**1. Giới thiệu bài toán**

Các ngân hàng sẽ tiếp tục cung cấp cho vay tín dụng cho mọi loại khách hàng và tổ chức trong bối cảnh nền kinh tế đang phát triển ngày càng nhanh chóng và sôi động. Các nguồn tài trợ để hỗ trợ phục hồi kinh tế sẽ tăng lên, đặc biệt là sau khi đại dịch COVID-19 kết thúc. Số liệu thống kê được Ngân hàng Nhà Nước Việt Nam cung cấp vào ngày 10 tháng 3 năm 2022 cho thấy tỷ lệ nợ xấu tại các ngân hàng đã có xu hướng tăng trở lại kể từ khi dịch bệnh bắt đầu, với tỷ lệ nợ xấu nội bảng tăng 1,6% và 4,4% vào năm 2019 lên mức 1,9% và 7,3% năm 2021 [1].

Tín dụng vẫn chiếm tỷ trọng cao nhất trong tổng tài sản của ngân hàng ở Việt Nam. Tín dụng cũng mang lại nguồn thu nhập lớn nhất cho ngân hàng nhưng cũng là hoạt động mang lại rủi ro cao nhất. Do đó, để tối đa hóa hoạt động của họ, các ngân hàng cần có những biện pháp nhằm hạn chế tối đa những rủi ro tín dụng này. Các ngân hàng và các tổ chức tín dụng khác nên quan tâm nhiều hơn đến việc đánh giá ban đầu rủi ro và khả năng trả nợ của khách hàng để đưa ra quyết định có cho khách hàng vay hay không và mức độ an toàn của khoản vay.

Các thông tin mà chúng tôi thu thập được từ khách hàng sẽ dựa trên khả năng trả nợ của họ, có thể bao gồm tuổi, thu nhập, lịch sử vay nợ trước đây, hôn nhân, tài sản, v.v. Chúng tôi sẽ đề xuất trong nghiên cứu này xây dựng một mô hình học máy để nâng cao khả năng phát hiện rủi ro tín dụng.

**2. Tổng quan nghiên cứu**

Nhiều nghiên cứu trước đây đã đề xuất các phương pháp và mô hình để giải quyết bài toán rủi ro cho vay tín dụng. Sau đây là một số kết quả của các bài nghiên cứu trước đây:

Trong bài nghiên cứu "Phân tích rủi ro tín dụng bằng cách sử dụng công cụ phân loại học máy (Credit Risk Analysis using Machine Learning Classifiers) của Trilok Nath Pandey và đồng tác giả (2017), nhóm nghiên cứu đã tiến hành một phân tích sâu sắc về rủi ro tín dụng bằng cách xây dựng và đánh giá các mô hình phân loại trên bộ dữ liệu liên quan. Bộ dữ liệu này cung cấp thông tin về các khía cạnh quan trọng của khách hàng và các yếu tố tiềm ẩn liên quan đến khả năng trả nợ của họ.

Kết quả nghiên cứu cho thấy rằng mô hình Extreme Learning Machine (ELM) là phương pháp dự đoán rủi ro tín dụng hiệu quả nhất. Chỉ số accuracy đã được sử dụng để đo accuracy của mô hình này và những con số đáng chú ý đã được ghi nhận. Chỉ số accuracy của bộ dữ liệu Australian và German là 0.9633 và 0.9632. Điều này làm nổi bật sức mạnh của mô hình ELM trong việc phân loại khách hàng và dự đoán khả năng trả nợ của họ; nó cung cấp cho các ngân hàng và các tổ chức tài chính một công cụ quan trọng để quản lý rủi ro và đưa ra quyết định về tín dụng. [2]

Nhóm tác giả của nghiên cứu "Phương pháp học máy để dự đoán mặc định khoản vay của người Trung Quốc trên thị trường P2P (Loan default prediction of Chinese P2P market: a machine learning methodology) đã tạo ra một cấu trúc mạnh mẽ để dự đoán khả năng mặc định trong thị trường P2P của Họ đã thực hiện một phân tích chuyên sâu để đánh giá hiệu suất của mỗi mô hình bằng cách sử dụng bốn mô hình phân loại phổ biến: Forest Random, XGBoost, Gradient Boost và Neural Network.

Khi tất cả bốn mô hình đều đạt được độ chính xác cao hơn 93%, kết quả nghiên cứu đã phản ánh điều đó. Điều này cho thấy các mô hình có khả năng dự đoán mặc định cao trên thị trường P2P của Trung Quốc. Mô hình Random Forest nổi bật với độ chính xác cao nhất, đạt được trên 98%. Sự tiến bộ này được coi là một bước tiến quan trọng trong việc sử dụng học máy để dự đoán rủi ro tín dụng trong môi trường P2P, đặc biệt là trên thị trường Trung Quốc, nơi mô hình này có thể cung cấp cho các nhà đầu tư và tổ chức tài chính thông tin quan trọng để quản lý [2]

Nghiên cứu được gọi là "Áp dụng mô hình học máy để dự đoán tính đủ điều kiện của khoản vay ngân hàng" của Ugochukwu .E. Orji và đồng nghiệp (2022), nhóm tác giả đã phát triển và đánh giá sáu mô hình phân loại khác nhau để dự đoán tính đủ điều kiện của Random Forest, Gradient Boost, Decision Tree, Support Vector Machine, K-Nearest Neighbor và Logistic Regression là một số trong số các mô hình này. Đáng chú ý, nhóm tác giả đã sử dụng thuật toán SMOTE để cải thiện hiệu suất của mô hình trong quá trình xử lý mất cân bằng dữ liệu..

Kết quả nghiên cứu cho thấy hiệu suất mô hình cao. Mô hình Random Forest có độ chính xác cao nhất trong số các mô hình được kiểm tra, lên đến 95%. Các mô hình khác có hiệu suất tốt, với độ chính xác trên 80%. Điều này cho thấy tiềm năng của việc sử dụng học máy trong việc dự đoán tính đủ điều kiện cho vay của khách hàng ngân hàng và cung cấp cho khách hàng ngân hàng một công cụ hữu ích để đưa ra quyết định cho vay. [4]

Nhóm tác giả của nghiên cứu "Ứng dụng một số mô hình học máy để dự đoán khả năng trả nợ của mỗi người khi nộp đơn đi vay tín dụng" (2022) đã xây dựng và đánh giá nhiều mô hình phân loại kết hợp với thuật toán Synthetic Minority Over-sampling Technique (SMOTE). Logistic Regression, Random Forest, XGBoost và LightGBM đã được xây dựng cụ thể. Kỹ thuật SMOTE cải thiện hiệu suất dự đoán của các mô hình và hỗ trợ cân bằng dữ liệu.

Sự ảnh hưởng đến việc dự đoán khả năng trả nợ được chứng minh bởi kết quả của nghiên cứu. Mỗi mô hình đều đạt được độ chính xác đáng kể; Logistic Regression đạt 70.8%, Random Forest đạt 95.2%, XGBoost đạt 95.5% và LightGBM đạt 95.4%. Điều này cho thấy rằng các mô hình học máy có thể dự đoán một cách chính xác và hiệu quả khả năng trả nợ của mọi người khi nộp đơn vay tín dụng, cung cấp một công cụ hữu ích cho quá trình ra quyết định cho vay tín dụng. [5]