Journal of the Operational Research Society

ISSN: 0160-5682 (Print) 1476-9360 (Online) Journal homepage: www.tandfonline.com/journals/tjor20

Integrating AI and OR for investment decision-making in emerging digital lending businesses: a risk-return multi-objective optimization approach

Vajiheh Torkian, Shahrooz Bamdad & Amir Homayoun Sarfaraz

To cite this article: Vajiheh Torkian, Shahrooz Bamdad & Amir Homayoun Sarfaraz (05 May 2025): Integrating AI and OR for investment decision-making in emerging digital lending businesses: a risk-return multi-objective optimization approach, Journal of the Operational Research Society, DOI: 10.1080/01605682.2025.2498652

To link to this article: https://doi.org/10.1080/01605682.2025.2498652

JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY
https://doi.org/10.1080/01605682.2025.2498652

Integrating AI and OR for investment decision-making in emerging digital lending businesses: a risk-return multi-objective optimization approach

Vajiheh Torkian[a], Shahrooz Bamdad[b] and Amir Homayoun Sarfaraz[a]

[a]Department of Industrial Engineering, ST.C., Islamic Azad University, Tehran, Iran; [b]Business School, Birmingham City University, UK

ABSTRACT

This study investigates the application of operational research techniques to optimize investment decisions in peer-to-peer (P2P) lending platforms, focusing on balancing risk and return for investors. The study proposes a multi-objective decision-making model that leverages data from the Lending Club, the largest P2P marketplace in the United States, to minimize risk and maximize returns. To address the data imbalance, the model uses classification techniques including logistic regression, decision trees, random forests, and light gradient boosting machines (LGBM), which are supported by the synthetic minority oversampling technique (SMOTE). While a convolutional neural network (CNN) predicts net present value (NPV), logistic regression is used to assess risk. The nondominated sorting genetic algorithm II (NSGA-II) is then used for portfolio optimization, producing returns of over 7% with risk levels that are comparable with conventional methods. Sensitivity analysis highlights the importance of investment allocation strategies by emphasizing that portfolio returns are more sensitive to changes in investments than risk. This study contributes to the operational research literature on risk management, investment modeling, and practical decision support systems in financial services by integrating advanced AI-based computational methods and optimization tools.

HIGHLIGHTS

- The multi-objective model seeks to balance risk reduction with return maximization.
- The LGBM, logistic regression, random forest, and decision tree models are assessed.
- The NSGA-II algorithm is used to optimize the portfolio model.
- A sensitivity analysis is used to evaluate the investment amounts.
- The results provide wisdom on return optimization and risk reduction in P2P lending.

ARTICLE HISTORY
Received 17 November 2024
Accepted 21 April 2025

KEYWORDS
P2P lending platforms; AI-driven investment decision-making; portfolio performance; risk-return multi-objective optimization; deep learning

1. Introduction

Financial technologies, or fintechs as they are sometimes called, have become increasingly prevalent in the global economy and have revolutionized almost every aspect of business and finance (Alfonso-Sánchez et al., 2024). Fintech businesses are in a competitive market with traditional financial institutions, striving to transform financial services through innovation and contemporary technology (Porfírio et al., 2024). Many cutting-edge financial services like digital banking, wealth management, insurance, mobile payments, and digital currencies go hand in hand with this (Gozman & Willcocks, 2019; Yap, 2023). The digital revolution has led to the rapid replacement of various traditional types of borrowing by platforms, commonly referred to as online lending P2P platforms (Sunardi et al., 2022; Zhu & Wu, 2025). Lending Club and Bondora are two examples of platforms that facilitate the matching of lenders and borrowers in this lending system (Zhao et al., 2017). Borrowers give personal information while registering on this platform, and their identification is verified by their ID card number. A credit rating is given to the borrower based on the information submitted in the registration form, which often contains personal, financial, and business information. In addition, certain details concerning the loan, like the amount of each monthly payment and the length of the repayment period, are typically requested or optional. Following the

CONTACT Shahrooz Bamdad sh.bamdad2000@yahoo.com Business School, Birmingham City University, UK.
Supplemental data for this article is available online at https://doi.org/10.1080/01605682.2025.2498652

registration process, borrowers specify the maximum interest rate they are ready to pay as well as the quantity of money they wish to borrow. Lenders who are prepared to offer money are placed on the loan list once the application is completed (Wang et al., 2015; Zhang et al., 2022).

The primary obstacle for individual investors in the P2P lending business is effectively allocating their funds across several loans while meticulously assessing the creditworthiness of each one (Ha et al., 2019). The credit rating of loans is essential to reduce default risk and increase profitability for lenders and platforms because many P2P loans are unsecured personal loans (Wang et al., 2021). There is a chance of nonpayment of the loan in this kind of lending system, which results in losses for the business. Therefore, several studies tried to lower the default risk by creating a default prediction classification model with an emphasis on accuracy (Muslim et al., 2023). A well-executed loan evaluation can support lenders in making wise investment choices. The association between loan repayments and final loan repayment outcomes is not discovered by existing approaches since loan repayments are not considered during the main learning stage (Wu et al., 2022). It is a difficult challenge for the lending institution to determine if a borrower will be a debtor or a defaulter before giving loans. With the ultimate goal of enhancing borrower classification in loan decision-making, developing an autonomous default loan classification system is an optimization and machine learning problem (Babo & Beyene, 2023).

Current approaches to P2P lending investment optimization face several challenges. Traditional risk prediction models, such as machine learning classifiers and logistic regression, have trouble handling imbalanced datasets and are unable to identify intricate, nonlinear borrower–risk relationships. Risk assessments are skewed because P2P lending datasets are frequently big and highly skewed, with a far larger percentage of successful loans than defaults. Hence, these models may fail to accurately identify high-risk loans, prioritizing the majority class instead.

To mitigate these issues, data resampling techniques are used in this article to improve classification accuracy and rebalance the dataset prior to model creation. To ensure that risk prediction models learn from a more representative dataset, methods like oversampling the minority class and undersampling the majority class help reduce the detrimental consequences of skewed data distributions. Synthetic data generation with SMOTE is a useful method that increases model robustness by producing artificial but realistic instances of under-represented categories.

The knowledge asymmetry between lenders and borrowers, which is more noticeable in P2P lending than in traditional banking, is another significant obstacle. Standardized credit reporting, guarantees, and collateral all help lenders in structured risk assessment and decrease asymmetric information in traditional financial institutions (Lu et al., 2024). These risk-reduction strategies, however, are less successful in P2P lending because of increased transaction costs and the absence of direct borrower-lender contacts. This emphasizes the necessity of sophisticated AI-driven approaches to close this gap and improve the precision of credit evaluations.

To get over these restrictions, this study suggests a hybrid investment recommender system that combines personalized risk-based suggestions, deep learning (DL), and portfolio optimization. Using deep learning models—which successfully capture high-dimensional borrower-risk patterns—the approach improves risk prediction and return estimation. SMOTE is also used to address problems with data imbalance, which enhances the model's capacity to forecast results for high-risk loans. Superior portfolio optimization is made possible by the application of NSGA-II in contrast to traditional mean-variance models. In comparison to earlier research, this study integrates financial indicators such as NPV into a thorough investment framework to concurrently optimize risk and return. Additionally, by considering the risk-averse and risk-tolerant investor profiles, the model tailors investment suggestions guarantee flexibility to accommodate a wide range of investment preferences. Ultimately, this study fills important research gaps and improves the usefulness of AI-driven portfolio optimization in P2P lending.

In this work, we offer an AI-driven multi-objective investment recommendation model for P2P lending that enables investors to optimize their investment selections. To the best of our knowledge, few studies have examined the issue of multi-objective P2P lending using deep learning. First, in this work to improve the model's dependability in social lending, the chosen dataset has been balanced using the SMOTE approach. Also, NPV has been used to compute and forecast the return on loans by comparing several convolutional neural networks. Then, logistic regression has been utilized to anticipate portfolio risk, and NSGA-II has been used to optimize the multi-objectives portfolio optimization model with return maximization and risk minimization.

1.1. Innovations

This study significantly advances the field of digital platform investment decision strategies, specifically in P2P lending, through the advancement of the integration of AI-driven methodology and portfolio optimization. In particular, we have made the following contributions:

i. *Improved Risk Management through Data Balancing*: To efficiently handle data imbalances in P2P lending datasets, we use classification models in conjunction with synthetic minority oversampling to increase the accuracy of risk forecasts. More accurate risk assessments are, thus, guaranteed, especially for high-risk loan types.
ii. *AI-Driven Return and Risk Prediction*: By combining deep learning models, such as CNN, for predicting net present value and logistic regression for risk assessment, the study improves portfolio optimization and provides a more accurate estimate of possible returns and risk exposure.
iii. *Dynamic Investment Strategy Analysis*: Through sensitivity analysis across various investment levels, we demonstrate the interaction between risk mitigation, investment decisions, and portfolio performance, further refining the decision-making process for investors.
iv. *Practical Implications for Digital Investment Strategies*: This study offers an approach for integrating meta-heuristic models with AI algorithms to identify the best possible portfolio in P2P lending. Through the multi-objective optimization of recommendations for both risk-tolerant and risk-averse investors, the model provides a customized investing framework that considers a range of investor attributes. It helps investors make data-driven, risk-adjusted investing decisions by giving them a deeper understanding of real investment environments.

The remainder of this research is structured as follows: In Section 2, the review of the literature is displayed. The techniques we employ in our model are demonstrated in Section 3, and the suggested investment model's empirical findings are presented in Section 4. Section 5 presents an examination of the experimental results of the suggested model, and a detailed discussion including managerial insights, practical implications, and research limitations are presented in Section 6. Section 7 conclude this study.

2. Related works

A critical issue in management engineering and decision-making is portfolio selection. However, investors find it difficult to accomplish predetermined goals due to the intrinsic complexity of the capital market and their frequently irrational behavior (Wang et al., 2023). Previous works have been studied in this section.

2.1. Integrating portfolio optimization techniques

To address portfolio optimization, García et al. (2020) extended a random mean-semi-variance model into a fuzzy multi-objective framework. The NSGA-II algorithm was used to solve the resulting constrained issue. Using a dataset of assets from the integrated Latin America market, their empirical investigation proved the model's efficacy and practicality. To maximize profits, Shih et al. (2022) developed a three-stage market basket approach that uses artificial intelligence to pinpoint high-potential product markets. To maximize overall market share and growth while reducing competition and market risk, their model employed a fuzzy multi-objective mathematical programming technique, which raised portfolio profitability. To find Pareto-optimal solutions for bi-objective and three-objective portfolio problems respectively, Awad et al. (2022) used NSGA-II and its variation, NSGA-III. Their research demonstrated that NSGA-III works best for three-objective optimization, which enables the best possible selection of stocks or assets to maximize returns and minimize risk, whereas NSGA-II efficiently manages two-objective problems. By examining 255 businesses registered on the Lima Stock Exchange, Flores-Fernandez et al. (2022) created a genetic algorithm in conjunction with an artificial neural network to optimize investment portfolios in the Peruvian market. According to their findings, the genetic algorithm produced better returns and less volatility, especially with a much lower volatility percentage. Using a genetic algorithm to create a Pareto front at each step and choose the best points depending on investor profile parameters, Kely de Melo et al. (2022) presented a multi-objective predictive control model for portfolio selection. The simulation showed how well this model balances risk-return trade-offs, transaction costs, and constraints while managing the dynamic character of the investment environment.

A thorough analysis of portfolio selection models is provided by Ziane et al. (2024), who look at a variety of mathematical frameworks used in optimization, including conventional mean-variance optimization and more recent techniques like value-at-risk (VaR) and conditional value-at-risk (CVaR) optimization. Investigating the effectiveness of CVaR as a measurement of systemic risk, Mba (2024) utilized statistical methods including vine copulas and APARCH-DCC models. To better understand dependence structures and the dynamics of tail risk in financial portfolios, the study assessed the sensitivity and robustness of the CVaR for two distinct portfolios under five allocation strategies. By combining pairs trading principles with a threshold-based approach, Bağcı and Soylu (2024) present a high-frequency rebalancing algorithm (HFRA) that offers a more flexible approach than conventional periodic and threshold rebalancing techniques. According to their findings, improving portfolio management techniques in high-frequency trading environments requires the use of real-time market data and volatility measures.

2.2. Risk models

Machado and Karray (2022) employed hybrid approaches combining supervised and unsupervised machine learning to investigate commercial customers' credit scores. Combination models outperform independently supervised models in forecasting commercial clients' credit scores, according to an analysis and comparison of their respective performances. A convolutional neural network architecture for automatic feature extraction and enhanced payback prediction performance in P2P social lending was presented by Kim and Cho (2019). Results from experiments using fivefold cross-validation demonstrated that the technique is efficient at predicting repayment in lending club data and can automatically extract complicated features.

To lower default risks and asymmetric information in P2P lending platforms, Ko et al. (2022) created predictive models and used the under-sampling method to pre-process a large amount of data that was taken from Lending Club. Following that, they employed five artificial intelligence models (decision tree, random forest, LGBM, artificial neural network, and convolutional neural network) in addition to three statistical models (logistic regression, Bayesian classifier, and linear discriminant analysis). Ultimately, the statistical significance of the difference between the models was demonstrated by the Student's *t*-test. Wei et al. (2023) used machine learning methods and the loan contracts of one of the biggest P2P websites online in China to forecast credit default probabilities for P2P lending. The outcomes of the experiment shown that the suggested model performs well in terms of minimum mistake rate, recall, and prediction accuracy.

Huang and Ma (2023) examined the effects of inflation on investments in risky and risk-free assets to solve a portfolio selection problem that takes inflation into account as a key background risk multiplier. Their results show that, in comparison to current strategies, the suggested strategy—which makes use of probability theory—produces more profit for investors. In a study, Sifrain (2023) used data from the Lending Club and an under-sampling methodology to forecast the likelihood of borrower default using three techniques: random forest, neural network, and logistic regression. In research, Asencios et al. (2023) created six profit scoring models that serve as a tool for credit analysts by predicting the internal rate of return of credit programs using XGBoost, TabNet, and multilayer perceptron algorithms. The XGBoost algorithm was recommended as the optimal method based on the outcomes of the profit scoring models' performance evaluation. Furthermore, the model with the best performance was the one that considered all the features and employed the XGBoost algorithm.

Novel methodologies for risk assessment are particularly helpful in light of the growing complexity of portfolio management. A roadmap for optimizing portfolio strategies in banking contexts is provided by Albehery et al. (2025), who explore inverse multi objective optimization for portfolio allocation in commercial banks. They demonstrate how a well-calibrated model can manage risks and returns more effectively.

2.3. Implementation of recommender systems based on artificial intelligence

Ahmadian et al. (2022) proposed a novel recommendation approach that models the labeling of information and the display of trust relationships using deep neural networks. To do this, hidden features from user-user trust connections and user label matrices are extracted using a sparse autoencoder. As the extracted hidden features have smaller dimensions than the original data, the suggested method can address the issue of data sparsity and lower the computational complexity of recommender systems. The suggested approach outperforms the most advanced recommender systems, as demonstrated by experimental findings on two benchmark datasets.

An investment recommendation model for P2P lending was suggested by Babaei and Bamdad (2020) to examine the effectiveness of several artificial neural network models using an under sampling approach in which NPV is considered as return. Based on real-world datasets, experimental results showed that the suggested strategy improves both risk and return on investment. A deep learning model based on a series of additive matrix factorization techniques was presented by Liu et al. (2021). This model suggests that the problem of commencing the maturity of loans can be efficiently solved by creating an incremental matrix factorization model based on time series. Considering the risk assessment, a neural network has then been utilized to offer investors personal investment advice services.

In their paper, Lee and Sohn (2021) demonstrated how to use self-supervised learning to build a deep neural network that can learn meaningful representations of high-dimensional and heterogeneous categorized features from patent data. They suggested a recommendation mechanism based on each company's technology portfolio for businesses searching for convergence opportunities. The findings demonstrated the potential utility of deep representation learning's recent theoretical and experimental advancements for the extraction of important structured data. Using an online optimization framework and a real-world dataset from an online financial platform, Hu et al. (2023) compared previous methods and ascertained investors' preferences for more financial product recommendations. Updating investor preferences for fresh datasets and handling scenarios when a significant number of values are missing from investor records are made feasible by the suggested approach. Furthermore, Gotardelo and Goliatt (2024) addressed the integration of behavioral economics into recommender systems by proposing a multiperiod fuzzy portfolio optimization model that takes investors' loss aversion into account.

2.4. AI-driven investment decision-making

Research on the role and impact of artificial intelligence on conventional business analysis was conducted by Chintala and Thiyagarajan (2023). Through case studies and real instances, they also demonstrated, using conventional methods, how artificial intelligence may help businesses better harness data, make more informed decisions, and accomplish their objectives. The mechanics and advantages of artificial intelligence-based customization in fintech were examined in a study by Abba (2022), which demonstrated how sophisticated algorithms and data analysis are utilized to customize payment solutions to the preferences of individual users. The substantial advantages of tailored recommendations are also covered in this article, such as enhanced user satisfaction, better financial decision-making, and higher conversion rates. The intricacy of artificial intelligence in finance was examined in a study by Bhat (2024), which concentrated on locating and eliminating barriers to its successful application. The report urges stakeholders to work together for sustainable innovation while predicting future developments and obstacles in AI-driven finance.

Zhang (2024) demonstrated how swarm intelligence techniques have advanced to solve complicated portfolio-related problems by presenting a novel application of multi-objective optimization for investment portfolios using an artificial fish school algorithm. Other tools such as decision-making and simulation tools have also been used in this field. For instance, by combining simulation approaches with multicriteria decision-making methodologies, Kaya et al. (2024) investigate the financial performance of firms in the Borsa Istanbul Sustainability Index, ensuring accuracy through expert reviews. To improve investment decisions in sustainability-focused indexes, their study highlights the usage of sophisticated decision-making frameworks and the integration of both qualitative and quantitative assessments.

2.5. Comparison with the state of the art and research gap

Several research, including those by García et al. (2020), Shi et al. (2019), Awad et al. (2022), and Kely de Melo et al. (2022), have focused on portfolio optimization. However, others have studied the use of AI to minimize investment risk, such as Kim and Cho (2019), Ko et al. (2022), and Wei et al. (2023). No study has systematically assessed how various machine learning algorithms perform when handling particular data-related issues, such as data imbalance and feature complexity. Chintala and Thiyagarajan (2023) reviewed research on the role and impact of AI on traditional business analytics, Abba (2022) reviewed the benefits of AI-based customization in fintech, and Bhat (2024) reviewed the complexity of AI in finance.

Regarding comparison with the state of the art, the literature review highlights progression in credit risk modelling and AI recommender system implementation for P2P lending platforms. Research has demonstrated the efficacy of hybrid approaches, which include supervised and unsupervised machine learning techniques, in the prediction of default probability in P2P lending scenarios and the forecasting of commercial credit scores, such as CNNs. Moreover, advances in using a range of AI models, such as decision trees, random forests, LGBM, and neural networks, and managing data imbalance through under-sampling techniques have greatly increased the accuracy of credit risk assessment. Furthermore, new studies have shown the value of AI-driven recommender systems in delivering individualized investment suggestions based on user trust relationships and portfolio evaluations. These systems make use of deep learning architectures and self-supervised learning techniques. The developments highlight a trend toward increasingly complex and all-encompassing AI-driven models for P2P lending credit risk assessment and investment suggestion, opening the door for more precise forecasts and improved decision-making capabilitiess in the financial business.

The research gap in the existing literature on AI-driven investment decision models for digital platforms, particularly in the context of P2P lending, is the absence of comprehensive studies that integrate multi-objective risk assessment with state-of-the-art machine learning techniques. Although several studies have investigated artificial intelligence-based recommender systems and credit risk models in P2P lending, further study is required to create comprehensive investment decision models by combining these approaches. Further research on the creation of hybrid recommender systems, which combine several recommendation techniques, such as deep learning-based approaches, may be possible. To be more precise, current research frequently concentrates on discrete elements like credit scoring or recommendation algorithms rather than considering the concurrent optimization of risk and return goals. Furthermore, little research has been done to systematically assess how well different machine learning algorithms perform when tackling the particular difficulties associated with P2P lending data, such as data imbalance and feature complexity. Moreover, not much research has been done on how investment choices, risk mitigation techniques, and portfolio performance interact dynamically in the digital era, especially in light of newly emerging digital platforms. Further research might therefore fill in these gaps by creating thorough AI-driven models for investment decision-making that consider sophisticated machine learning methods, multi-objective risk assessment, and the dynamic nature of investments in digital platforms. This study tries to fill this gap, and its superiority and uniqueness compared to existing models in the literature can be summarized as follows:

i. This article addresses the aforementioned research gap by providing an extensive study that combines sophisticated deep learning approaches with multi-objective risk assessment in the context of P2P lending on digital platforms.
ii. The work intends to close the gap between current research on credit risk models and recommender systems by putting forth an AI-powered investment decision model that simultaneously addresses risk and return objectives.
iii. The study assesses the efficacy of different machine learning algorithms, such as logistic regression, random forest, decision tree, and LGBM, in handling the complexities of P2P lending data by carefully preprocessing data.

Additionally, the research provides a comprehensive approach to investment decision-making by utilizing methods like CNNs for NPV forecasting and logistic regression for risk assessment, in addition to SMOTE for handling data imbalance. The study's contribution to comprehending the dynamic relationship between investment decisions, risk management, and portfolio performance is further enhanced by the application of the NSGA-II for model optimization and sensitivity analysis to evaluate dependency on different investment levels.

3. Methodology

3.1. Research design

This study employs a multi-stage research design that combines deep learning, risk assessment, and multi-objective portfolio optimization to improve P2P lending investment decision-making. To resolve data imbalance, the dataset is first preprocessed through feature selection, variable transformation, and class balancing using SMOTE. Subsequently, completed loans are used to train deep learning models, particularly CNN, which forecast NPV, a metric for investment returns. The probability of default (PD) is estimated using logistic regression, guaranteeing a thorough assessment of risk and return. To guarantee reliability and accuracy, performance measurements including sensitivity analysis and confusion matrices are used to validate prediction models.

Following risk-return assessment, a framework for portfolio optimization is used, leveraging NSGA-II to determine the optimal investment allocations. By minimizing PD and maximizing NPV, this multi-objective optimization seeks to give investors a balanced trade-off between risk and return. Both risk-averse and risk-tolerant investors can benefit from the final investment suggestions, which are customized to fit various risk profiles. This study offers a solid, data-driven method for making investment decisions in digital lending markets by fusing artificial intelligence with operational research methodologies.

3.2. Structure of neural network

The three components of a typical neural network are its topology, activation function, and learning rules. The way neurons respond to the activity output of other neurons is determined by a function called the activation function. The connection mode is determined by the connection weights between neurons and learning rules; these weights are tuned and changed during training (Rahmati et al., 2024; Ren et al., 2016). Deep learning algorithms are made to gradually extract relevant information from the raw input by optimizing with numerous layers and learning from data using neural networks to build a powerful model (Makridakis et al., 2023; Borisov et al., 2022; Zao et al., 2024). CNN is a deep learning system made up of multiple layers of neurons, each with its own set of weights and biases, that simulates the operation of the human visual cortex (Hoogstra et al., 2024; Qin et al., 2024). An array of integers representing tabulated data is fed into the grid as input (Yan et al., 2015). The first layer of neurons is typically a convolution layer, which applies a convolution operation to the input data to generate a set of feature maps. Then, these feature maps are processed by several additional neuron layers, each with a unique number of neurons and activation function. Ultimately, a set of labels or categories typically represent the network's output (Hamad et al., 2020). A type of deep learning architecture known as recurrent neural networks, or RNNs, can process sequential input by maintaining information across time steps. Another well-known deep learning architecture is the long short-term memory (LSTM) network, which is a type of RNN. Because LSTMs are specifically designed to manage sequential data by storing information across time, they are particularly well-suited for applications with long-term dependencies.

3.3. Investment proposal model

In the proposed model, we first use the SMOTE method to balance the dataset, and then, use the loans disbursed in the dataset to estimate loan repayment. After training the deep neural network using completed loans, which can forecast the return of loans or new listings, the NPV investment evaluation index is used as a return variable to evaluate the profit from investing in various loans, and logistic regression is used to predict the probability of default of the lists. After training the deep neural network using completed loans, which can forecast the return of loans or new listings, the NPV_i investment evaluation index is used as a return variable to evaluate the profit from investing in different loans. Next, the objective functions to be maximized and minimized are NPV and PD, respectively. The next stage is portfolio optimization, whose primary goal is to determine the ideal weights associated with specific investment decisions. Its concepts are derived from Markowitz's mean-variance model (Markowitz, 1959). The two-objective model is solved using a genetic algorithm, and the return and risk values of the investment portfolio are then computed based on the weights that were determined. The suggested model for this study is shown in Figure 1.

By clearly defining each stage of the workflow, Figure 1 clarifies the transition from data preprocessing to model training and portfolio optimization. It begins with the dataset, which undergoes feature selection to refine input variables. After addressing class imbalances with SMOTE, the original dataset is divided into train and test sets. Predicting returns using CNN, which estimates NPV, and predicting risk using logistic regression, which determines the probability of default, are the two concurrent components of model training. These forecasts inform a systematic risk-return analysis, in which performance evaluation evaluates the model's predicted accuracy and portfolio optimization refines investment decisions. Understanding how different risk levels affect returns is summarized in the analysis of results. By visually distinguishing the decision boundaries and interactions

Algorithm 1. Investment decision-making model.

Start

Data Preprocessing $\rightarrow$ Fully paid and charged-off loans $\rightarrow$ Balancing with SMOTE $\rightarrow$ 70% Train, %30 Test data

Returns Calculation $\rightarrow \mu = \sum_i \lambda_i NPV_i \quad \rightarrow \left\{ \begin{array}{c} \text{Training a DL model based on fully paid loans} \\ \text{Find the best DL model based on MSE, MAE and RMSE} \\ \text{Predicting NPV using the best DL} \end{array} \right.$

Risk Calculation $\rightarrow \delta^2 = \sum_i \lambda_i^2 PD_i \rightarrow$ Predicting $PD_i \rightarrow \log it\left(\widehat{PD}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_K x_k$

Performance Evaluation $\rightarrow$ Confusion matrix

Portfolio Optimization $\rightarrow$

$$\min \delta^2 = \sum_i \lambda_i^2 PD_i$$

$$\max \mu = \sum_i \lambda_i NPV_i$$

$$st:$$

$$\sum_i \lambda_i = 1$$

$$\lambda_i \geq 0$$

$$m \leq \lambda_i M \leq \text{loan amount}$$

Analysis of results $\rightarrow$ Sensitivity analysis

End

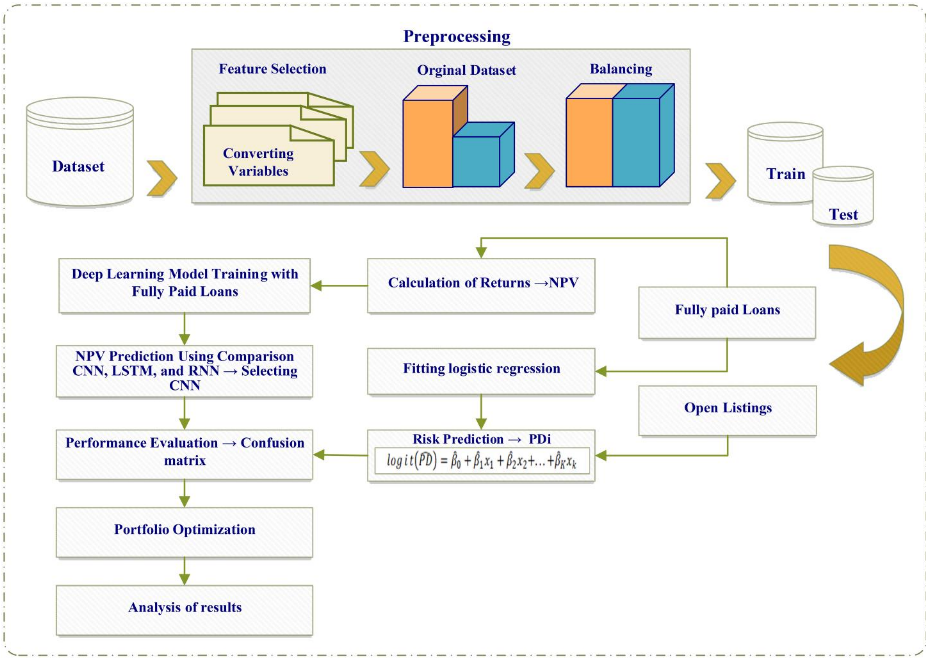

Figure 1. The proposed model.

between different risk profiles, the figure highlights important trade-offs between risk and return.

Algorithm 1 presents a framework for proposed P2P lending investment decision-making powered by AI that combines data preprocessing, predictive modelling, performance assessment, and portfolio optimization. The dataset is first divided into 70% training and 30% testing subsets after being balanced using SMOTE to address class imbalances. Risk is evaluated by assessing the probability of default by logistic regression, while returns are computed by forecasting NPV using deep learning models optimized based on MSE, MAE, and RMSE. Reliable risk predictions are ensured by evaluating model performance using a confusion matrix. The objective of portfolio optimization is to reduce risk while reaching a predetermined return, R^*, subject to constraints on allocation and loan limits. Finally, sensitivity analysis improves decision-making for various investor profiles by analyzing how investment alterations affect risk-return trade-offs. In dynamic lending environments, investors can make data-driven, risk-adjusted portfolio decisions using this structured method.

3.4. Designing the return model with NPV

The expected return (μ) is calculated in the form of a weighted average by using finished loans, each of which has a historical return NPV_i and weight λ_i, through the following equation:

$$\mu = \sum_i \lambda_i NPV_i. \tag{1}$$

3.5. Designing a risk model with logistic regression

To evaluate the risk of new loans (δ^2), the weighted average of past loans PD_i and logistic regression are used to predict the variance of the lists according to the following equation:

$$\delta^2 = \sum_i \lambda_i^2 PD_i. \tag{2}$$

3.6. Constraints of the model

Equation (3) is the limit of weights, which states that the total weight of loans in the capital portfolio must be equal to one. Equation (4) considers the minimum weight of each asset in the portfolio to be greater than zero and rejects negative numbers.

$$\sum_i \lambda_i = 1, \tag{3}$$

$$\lambda_i \geq 0. \tag{4}$$

The total asset M that an investor has for a selected loan, the minimum investment amount (m) and loan amount for P2P lending platforms are shown in Equation (5). In Lending Club, the minimum investment amount for each loan is \$25.

$$m \leq \lambda_i M \leq \text{loan amount} . \tag{5}$$

4. Experimental results

4.1. Research data

The largest P2P platform, Lending Club, made its statistical data available to the public during the first quarter of 2017, which is available on its website

Table 1. Description of independent variables used in this study.

Attribute	Definition
Annual income	The annual income that the borrower disclosed during registration.
Delinquency_2 years	The quantity of past-due instances of 30 days or more that have occurred in the borrower's credit history throughout the last two years.
Employment length	Years worked during employment. There are ten possible values: 0 denotes a value of less than a year, and 10 denotes a value of ten years or more.
Homeownership	Rent, own, and mortgage.
Inquiry last 6 months	Inquiries into credit in the previous six months in number.
Loan amount	The listed loan amount that the borrower has applied for.
Loan purpose	Credit card, house purchase, home improvement, major purchase, debt consolidation, car, medical, small business, vacation, moving, renewable energy, and others.
Open account	The number of open credit accounts within the borrower's credit history
fico_range_high	The upper bound range that the loan's originator's FICO score falls into.
fico_range_low	The lower bound range into which a borrower's FICO is situated when the loan is first originated.
LC subgrade	For borrowers, there are 35 loan subgrades ranging from A1 to G5, with the most secure being A1.
Dti	A ratio determined by dividing the borrower's total monthly debt payments by the total amount of debt obligations (excluding mortgages and the proposed LC loan) and then dividing the result by the borrower's reported monthly income.
Revolving utilization	The proportion of the borrower's credit is used to the total amount of available revolving credit.
Interest rate	The interest rate incurred by the borrower on a loan.
Loan amount to annual income	Loan amount to yearly income
Annual installment to income	The annual debt repayment made by the borrower is divided by the annual income that the borrower stated at enrollment.
Public records	Number of negative public records.
Month since last delinquency	The length of time in months since the borrower's previous late payment.
Grade	Borrowers are categorized into seven loan grades by Lending Club, ranging from A to G, with A being the safest.

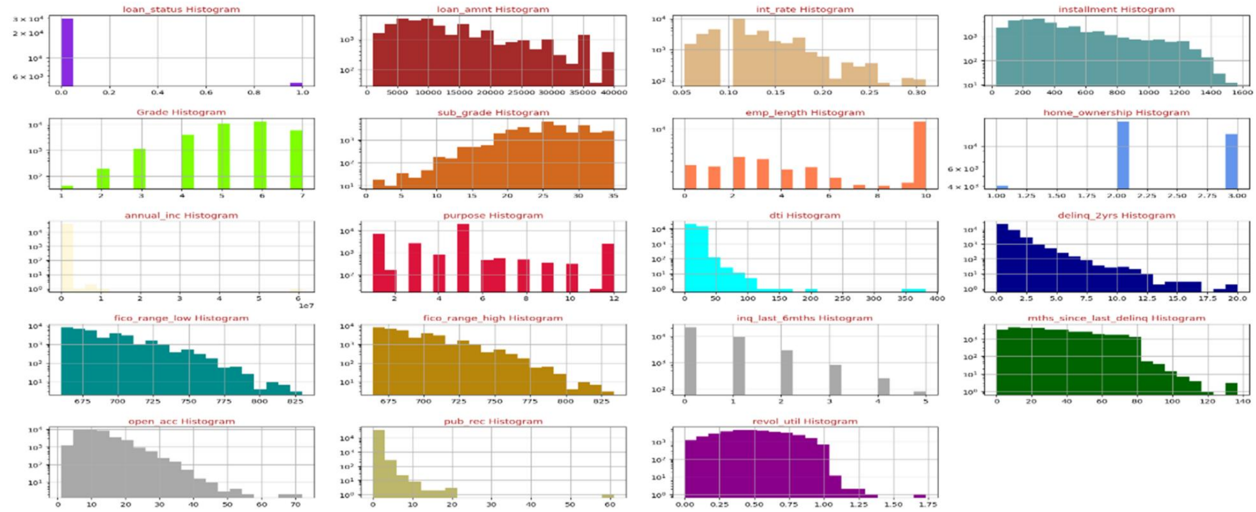

Figure 2. Schematic illustrating the variables used in this study.

(https://www.lendingclub.com/). This information served as the basis for the research. Based on 150 criteria, the pertinent database has 96,779 distinct loan types with 36- and 60-month repayment terms. Since we are unsure of the final status of the loans that are currently in this database, we have only examined completely paid and charged-off loans with a 36-month payback duration. As preprocessing the data to shorten calculation times disrupted the model's functionality, the majority of the study's features with empty values as well as rows with empty, incomplete, and outlier values were eliminated. After processing, selected features were determined to be independent variables, and Table 1 displays descriptions of the chosen variables.

The histogram of research variables is shown in Figure 2.

Then, non-numerical variables were transformed into numerical data using coding techniques (Lopez-Arevalo et al., 2020). For example, a loan that is fully paid is assigned to zero, while a loan that fails is assigned to one to train the model. It is required to adjust or scale up features due to the disparity in data size as shown in Figure 3.

A min–max normalization scaling method has been employed in this research. In Equation (6), x represents the initial value of the feature, x_{min} represents the lowest value of the feature in the data, and x_{max} represents the highest value of the feature. By applying this formula to each feature value, its

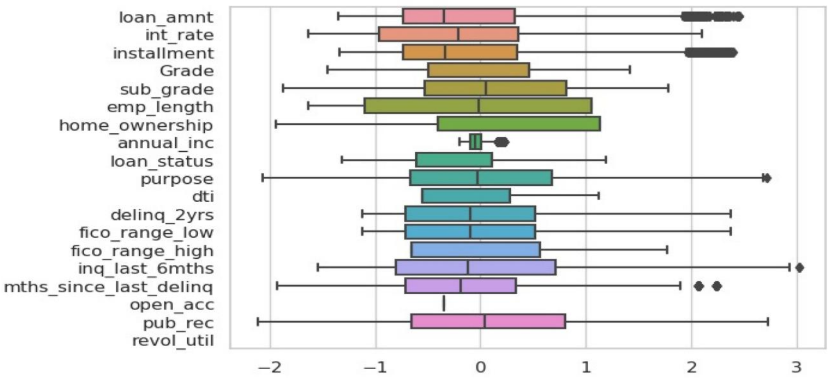

Figure 3. Box plot of research variables.

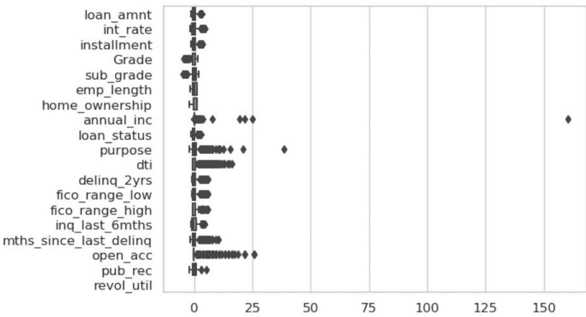

Figure 4. Scaling of studied characteristics.

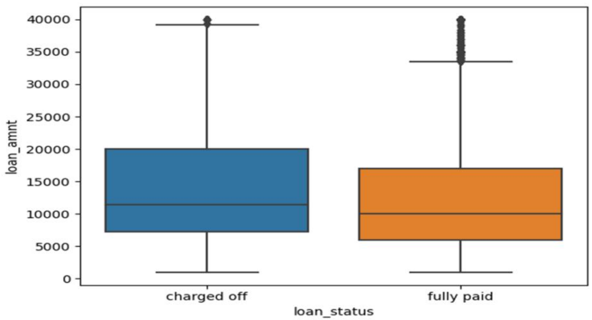

Figure 5. Data distribution plot for charged-off and fully paid loans.

scaled value will be in the range of 0 to 1.

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \tag{6}$$

The scaling of the features is shown in Figure 4.

4.2. Balancing

Compared to traditional banks that have been in the business for centuries, P2P lending platforms confront more information asymmetry because of their coverage of broader financial issues (Syzdykov, 2020). Class imbalance is a common issue in machine learning because of the lack of data from one class. Problems with imbalanced classification occur when there are more notable examples of certain classes than others (Grina et al., 2023; Koyyada & Singh, 2023; Zhang et al., 2024). Conventional classifiers frequently produce inaccurate predictions and are biased toward majority classes. Oversampling is a popular and effective method for creating a balanced sample from imbalanced data (Cefis & Carpita, 2025; Ren et al., 2023). Another well-liked technique for balancing out imbalanced datasets is the SMOTE method, which comes in a variety of forms to improve efficiency (Liu & Wang, 2025; Thejas et al., 2022). It balances the dataset by artificially generating points from minority data. Using this technique, a new artificial sample is created by randomly choosing one of each minority sample's closest neighbors. The characteristics of its closest neighbors are randomly combined to create the new sample (Mishra, 2017). Figure 5 shows the distribution of fully paid and failed loans. The phrase "Fully Paid" implies that the loan has been entirely repaid at maturity, whereas charged-off denotes a failed loan.

Table 2. Number of samples in fully paid and failed classes.

Loan status	Quantity	Proportion
Charged-off	4,970	14.27
Fully-paid	29,850	85.73
Total	34,820	100%

Table 2 shows that 29,850 of the loans in our dataset are fully paid, compared to 4970 that are not. This creates a class imbalance in the dataset, which needs to be balanced.

4.3. Classification model design

At this point, according to Table 3 the dataset under study has been divided into 70% for training and 30% for testing. It has been assessed using SMOTE oversampling and random under sampling. 24,374 entries make up the training dataset, whereas 10,446 records make up the test dataset. There are 3496 records for the minority class and 20,878 records for the majority class in the training dataset.

According to Table 4, the efficacy of the classification models for logistic regression, random forest, decision trees, and LGBM in random oversampling (ROS) and SMOTE oversampling has been evaluated using six criteria: accuracy, AUC-ROC, log-loss, precision, recall, and F1_score.

These machine learning methods are chosen based on their computational efficiency, predictive accuracy, and capacity to manage class imbalance. Because of their unique advantages in managing financial data, logistic regression, decision trees, random forests, and LGBM are selected for classification. For binary classification, logistic regression is a fundamental yet effective model offering interpretability and ease of implementation. A straightforward, rule-based approach for understanding decision boundaries is offered by decision trees. Compared to single decision trees, random forests reduce overfitting and improve prediction stability through ensemble learning. The gradient boosting framework LGBM is chosen due to its effectiveness in handling large financial datasets at faster training speeds. To address the problem of financial data imbalance and guarantee fair comparisons across models, the random oversampling techniques and SMOTE are used. So, these methods effectively address the challenges of risk-return optimization in P2P lending by balancing predictive accuracy, interpretability, and computational efficiency.

The outcomes demonstrate that improved classification performance arises from combining the random forest model with the SMOTE method. Figure 6 illustrates the significance of the features utilizing the SMOTE approach in conjunction with the random forest model.

Table 3. Summary of data partitioning of the proposed model.

Data	Dataset number	Percentage, %
Training data	24,374	70
Test data	10,446	30
Total	34,820	100

4.4. Experimental results

Before explicitly discussing the approach, we provide a small illustrative example. First, the dataset is balanced using SMOTE, creating synthetic samples to address the imbalance between defaulted and fully paid loans. Logistic regression then is used to predict the PD for each loan, for instance, Loan A, which has a PD of 0.15, and Loan B, which has a PD of 0.30. Next, CNN provides a reliable measure of projected returns by estimating the NPV for each loan based on historical repayment patterns. For example, it predicts that Loan A's NPV will be $1,200 and Loan B's NPV will be $800. Then, we have two objective functions of $\min 0.15\lambda_1 + 0.3\lambda_2$ and $\max 1200\lambda_1 + 800\lambda_2$, with these constraints: $25 \le 3000\lambda_1 \le 2000$, $25 \le 3000\lambda_2 \le 2500$, $\lambda_1 + \lambda_2 = 1$, and $\lambda_1, \lambda_2 \ge 0$ taking into account $m_1 = 2000$ and $m_2 = 2500$ for the first and

Table 4. Comparison results of models.

Model	Resampling	Accuracy	AUC-ROC	Log-loss	Precision	Recall	F1_score
Logistic regression	ROS	62.64	**0.6226**	13.4637	0.2201	0.5345	0.3118
	SMOTE	61.88	0.6218	13.7397	0.2262	0.5061	0.3127
Random forest	ROS	83.77	0.5126	5.8485	**0.2395**	0.4624	0.2990
	SMOTE	**85.52**	0.5072	**5.2171**	0.2087	**0.5491**	0.2870
Decision tree	ROS	76.50	0.5199	8.4674	0.1730	0.1797	0.1763
	SMOTE	74.27	0.5253	9.2748	0.1721	0.2123	0.1901
LGBM	ROS	67.19	0.6134	11.8247	0.2226	0.5318	**0.3139**
	SMOTE	74.13	0.5219	9.3231	0.1721	0.2184	0.1925

Accuracy: Classification accuracy.
AUC-ROC: Showing the performance of the classification model at all classification thresholds.
Log-loss: The amount of prediction uncertainty based on its difference from the true label.
Precision: It calculates the proportion of positive predictions that are actually correct.
Recall: It calculates the proportion of actual positives that the model correctly identified.
*F*1_score: It provides a balanced measure of how well a model performs in positive instances.

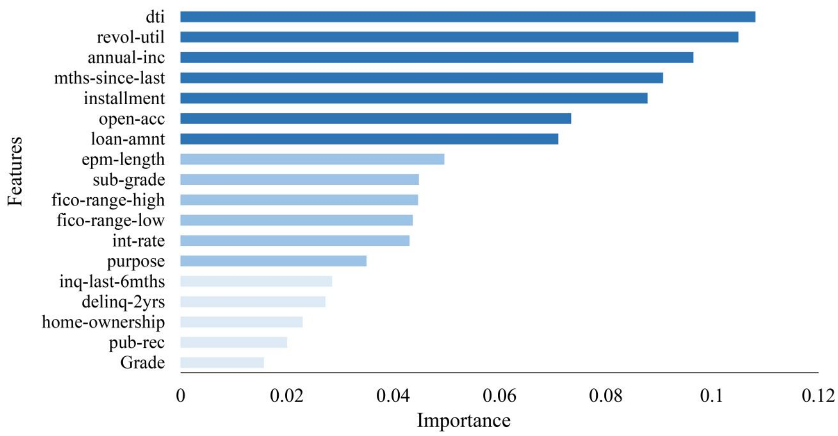

Figure 6. The features of importance from random forest model.

Table 5. The computational results of the CNN, LSTM, and RNN model for NPV predictions.

Model	Epochs	Batch size	MSE	MAE	RMSE
CNN	10	128	0.1418	0.1461	0.3766
		256	0.1429	**0.1432**	0.3781
		512	**0.1300**	0.1820	**0.3606**
	100	128	0.1343	0.1703	0.3664
		256	0.1343	0.1663	0.3665
		512	0.1310	0.2958	0.3620
LSTM	10	128	0.1479	0.2044	0.3875
		256	0.1429	0.1548	0.3762
		512	**0.1300**	0.2035	0.3624
	100	128	0.1385	0.1958	0.3752
		256	0.1363	0.1854	0.3680
		512	0.1335	0.2658	0.3645
RNN	10	128	0.1684	0.1872	0.3905
		256	0.1665	0.1722	0.3743
		512	0.1553	0.1849	0.3724
	100	128	0.1465	0.2977	0.3842
		256	0.1551	0.2965	0.3853
		512	0.1545	0.3047	0.3730

second loan amounts, respectively, total asset of investor amount $M = 3000$, and minimum investment amount $m = 25$ that is required by Lending Club. Finally, NSGA-II is used for portfolio optimization, assigning investment amounts λ_1 to Loan A and λ_2 to Loan B to maximize returns while minimizing total portfolio risk.

4.4.1. Investment return forecast

The weighted average of the expected loan return using a convolutional neural network is the portfolio return, or investment return, as defined by Equation (7). The interest rate, loan amount, and annual payment are the inputs and N is the number of assets. NPV_i is used as a return measure in this study.

$$\mu = \sum\nolimits_{i=1}^{N} \lambda_i NPV_i. \tag{7}$$

Net present value, which is calculated to identify the difference between the cost of the project and its outgoing or incoming cash flow, is one of the economic evaluation indicators on which economic judgments are based (Castro, 2022; Jones & David Smith, 1982; Yu et al., 2024). According to Equation (8), the NPV of an investment project is the sum of the present values of all incoming and outgoing cash flows, where t is the year or calculation period, T is the length of the investment period, C_t is the total of financial receipts or payments in the period t, and r is the expected discount rate, inflation rate or profit. When $NPV > 0$, investment in a project is profitable. The larger the NPV, the more profitable the project (Yoomak et al., 2019).

$$NPV = \sum\nolimits_{t=1}^{T} \frac{C_t}{(1+r)^t} - C_0. \tag{8}$$

Table 5, which was chosen by contrasting the outcomes of the optimal network architecture, prediction results using the CNN, LSTM, and RNN for evaluation criteria. With the lowest MSE of 0.1300 and the lowest RMSE of 0.3606, the CNN model with 10 epochs and a batch size of 512 is chosen based on the computational results, suggesting optimal performance among the evaluated configurations.

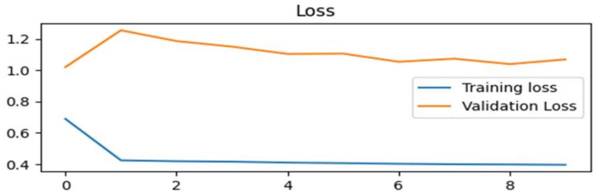

Figure 7. CNN model loss for training and evaluation model.

Table 6. Hyperparameters of the proposed model.

Model	Name	Values	Name	Values
CNN	Layer	8	Hidden size	512
	Filter	32,64	Epoch	10, 100
	Kernel size	3	Batch size	256, 512
	Pool size	2	Optimizer	Adam
	Dropout	0.5	Activation	RELU

The model error for the convolutional neural network's training and evaluation modes is displayed in Figure 7. It is observed that as they pass through the early phases, they stabilize.

Table 6 lists the hyperparameters used in the proposed model.

CNN's ability to extract hierarchical features from structured financial data led to its selection for return estimation. CNNs are ideal for estimating NPV because, in contrast to typical machine learning models, they can grasp intricate, nonlinear correlations in financial patterns. CNN's capacity to detect spatial dependencies and reduce feature engineering efforts justifies its selection over alternative deep learning models like LSTM or RNN. More accurate return predictions are made possible by the convolutional layers' assistance in spotting trends in loan repayment behavior. CNN is also a strong candidate for modelling financial return estimation in P2P lending due to its scalability and optimization capabilities.

4.4.2. Investment risk prediction

Equation (9), which defines portfolio risk or investment risk, uses PD as a risk metric in the context of portfolio optimization. The statistical model of logistic regression is used to categorize a binary dependent variable and use the results to forecast the likelihood of a loan default.

$$\delta^2 = \sum\nolimits_{i=1}^{N} \lambda_i^2 PD_i. \tag{9}$$

Logistic regression is employed to estimate risk by forecasting the probability of default (PD_i) because of its interpretability and transparency. In financial decision-making, particularly in risk

Table 7. The results of logistic regression.

Variables	$\hat{\beta}_k$	SE Coef	*Z*-value	*p*-value
Constant	−7.91	0.933	−8.47	0
loan amount	−0.000018	0.000025	−0.72	0.473
int_rate	10.39	2.28	4.57	0
installment	−0.000085	0.000721	−0.12	0.906
emp_length	0.00977	0.0043	2.27	0.023
annual_inc	0.000001	0	3.51	0
dti	−0.00724	0.00184	−3.94	0
delinq_2yrs	−0.0152	0.0147	−1.03	0.302
fico_range_low	1.9844	0.2339	8.4848	0
fico_range_high	−1.9775	0.2333	−8.4747	0
inq_last_6mths	−0.1114	0.0185	−6.01	0
mths_since_last_delinq	0.001538	0.000869	1.77	0.077
open_acc	−0.0025	0.00298	−0.84	0.401
pub_rec	−0.0468	0.0177	−2.64	0.008
revol_util	0.0649	0.0807	0.8	0.421
Grade	0.0486	0.0544	0.89	0.371
sub_grade	0.1638	0.0208	7.88	0
home_ownership	−0.1539	0.0247	−6.23	0
purpose	−0.01395	0.00583	−2.39	0.017

assessment, model explainability is essential for investor confidence and regulatory compliance. Logistic regression provides probabilistic outputs, which are essential for determining the likelihood of loan default. In contrast to more intricate models, it prevents overfitting when used on structured financial datasets with clearly defined feature relationships. Furthermore, it is a practical choice for large-scale P2P lending datasets where risk assessment must be performed quickly and precisely because of its computational efficiency and ease of implementation.

To determine risk, a binary logistic regression model is constructed, with fully paid debts represented by 0 and charged-off loans represented by 1. The logistic regression model (Chawla et al., 2004; Dong et al., 2024) in the form of Equation (10) is a mathematical representation that is fitted to the estimates of $\hat{\beta}_k$ ($k = 0, 1, \ldots, 19$) to calculate the PD of lists in the form of Equation (11).

$$\log\left[\frac{p}{1-p}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k, \tag{10}$$

$$\log it\left(\widehat{PD}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_K x_k. \tag{11}$$

The results of the logistic regression fit are displayed in Table 7, from which p-values at a significance level of 0.05 can be drawn conclusions. The model achieved an R-squared of 63.6%, and only significant coefficients ($\hat{\beta}_k$) are taken into account in the model's final equation (i.e., $\hat{\beta}_0, \hat{\beta}_2, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6, \hat{\beta}_8, \hat{\beta}_9, \hat{\beta}_{10}, \hat{\beta}_{13}, \hat{\beta}_{16}, \hat{\beta}_{17}, \hat{\beta}_{18}$).

To determine the number of appropriate variables in the model, we employ 10-fold cross-validation, to assess the fit performance of the logistic regression model once it has been trained. A portion of the training data used for modeling is split into two parts in the cross-validation method during an iterative procedure. Part of the data is used for training and another part is utilized for testing the model each time the cross-validation process is repeated; this procedure is regarded as a resampling approach to estimate the model loss.

Table 8. Confusion matrix.

	Prediction	
Loan status	Fully paid	Charged-off
Fully paid charged-off	True Positive (TP) False Positive (FP)	False Negative (FN) True Negative (TN)

Table 9. Empirical results of 10-fold cross-validation.

	Accuracy	Sensitivity	Specificity
Fold number	(TP + TN)/ (TP + FP + TN + FN)	TP/ (TP + FN)	TN/ (TN + FP)
1	0.8585	0.0095	0.3043
2	0.8511	0.0045	0.25
3	0.8556	0.0073	0.3928
4	0.8565	0.0127	0.4523
5	0.857	0.0074	0.3666
6	0.8568	0.0047	0.2916
7	0.8536	0.0079	0.3076
8	0.8622	0.0105	0.3658
9	0.8604	0.0096	0.4375
10	0.8602	0.0082	0.4

Table 8 displays the confusion matrix, while Table 9 displays the findings of the evaluation. The findings show that the model successfully finds high-risk loans with a favorable trade-off between accuracy, sensitivity, and specificity.

5. Analysis of the proposed model's findings

Sensitivity analysis has been used in this part to assess the dynamics of the expected return and calculate the risk of new loans by analyzing the performance of previous loans. Owing to the substantial volume of information and computation time, 100 loans were chosen at random, and the return and risk values of the portfolio were computed by assigning a minimum of 25 dollars to each loan. The return and risk values derived from the genetic algorithm's optimization are presented in Table 10. The following describes the genetic algorithm specifications for portfolio asset allocation optimization, which seeks to maximize returns while minimizing risk:

- *Fitness Function*: The ratio of portfolio return to portfolio risk is the GA's objective function.
- *Population Generation*: Random weight distributions across the assets in the portfolio are used to create an initial population. For this study, the population size is 100.
- *Selection Process*: Based on fitness ratings, the selection process keeps the best performers and eliminates the others to maintain a consistent population size.
- *Crossover*: This stage creates new asset allocations by combining chosen parent pairs to produce

Table 10. Return and risk by NSGA-II.

Portfolio return	Portfolio risk
1.9654	0.0990

Table 11. The effect of different amounts of investment.

M	Portfolio return	Portfolio risk
5000	1.5372	0.0994
10,000	2.1914	0.1029
15,000	2.2621	0.1002
20,000	**2.2845**	0.1022
25,000	2.2742	0.0970
30,000	1.8149	0.0969
35,000	2.1479	**0.0901**
40,000	2.2200	0.1027
45,000	2.1464	0.0997
50,000	2.1841	0.0987

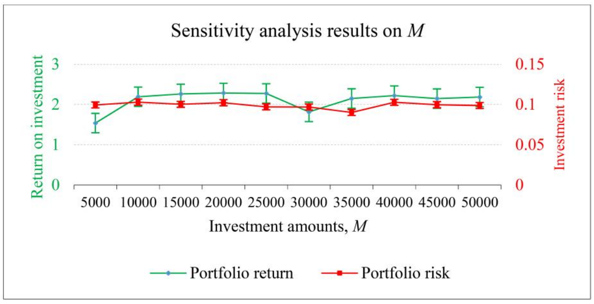

Figure 8. Portfolio return for different amounts of investment, *M*.

offspring, each of whom inherits characteristics from both parents.

- *Mutation*: By introducing changes in asset weights, a mutation rate of 0.05 is applied to the population to preserve diversity within it and avoid premature convergence on local optima.
- *GA Iteration*: To gradually optimize the population, the GA function iterates across generations while supervising the entire evolutionary process. After a predetermined number of generations, it returns the optimal asset allocation.

Table 11 displays the impact of varying investment amounts, i.e., *M*, on the risk and return of the suggested approach. Accordingly, the highest risk for the investment amount is $10,000, and the highest return for the investment amount is $20,000. Since the level of risk tolerance of each investor plays an important role in choosing the type of investment, the investment amount of $35,000 has the lowest risk.

Figure 8 shows the dynamics in portfolio return and risk for different amounts of investment, *M*. This graph displays how changes in *M* affect return and risk. The figure makes it evident that changes in the rate of return outweigh variations in risk. This indicates that altering investment amounts affects return more than it does risk.

Within the suggested AI-driven investment model, the sensitivity analysis demonstrates the

Table 12. Comparing the performance of two investment models.

Model	Portfolio return	Portfolio risk
The proposed model	1.9654	0.1
Traditional single-objective model	1.8325	0.1

complex relationships between investment levels and their associated impacts on risk and portfolio return. The investigation highlights how different investment amounts have a major impact on projected returns by choosing 100 loans at random and calculating their performance characteristics. Table 10 shows that the portfolio attains a noteworthy return of 1.9654 with a risk level of 0.0990 through optimization using the NSGA-II. This provides a starting point for understanding how to properly manage risk and maximize rewards. The subsequent analysis in Table 12 shows that larger investment amounts typically result in higher returns; for instance, an investment of $20,000 yields the greatest return of 2.2845. Conversely, a $35,000 investment is shown to carry the least amount of risk, demonstrating the model's flexibility in accommodating investors with different levels of risk tolerance.

The findings further illustrate that changes in investment quantity have a greater impact on return than risk, as depicted in the analysis. For example, the risk just slightly increases from 0.0994 to 0.1029, but the return at $10,000 highly surpasses the return at $5000, which is 1.5372. These findings imply that without correspondingly raising their risk exposure, investors can strategically manage their portfolios to generate the best returns. By graphically illustrating how changes in investment amounts might result in disproportionately higher returns in comparison to any fluctuations in related risks, the graphical representation supports this idea. The relationship demonstrates the model's effectiveness in directing investment choices and implies that, in the context of P2P lending, strategic allocation can result in notable improvements in portfolio performance.

The sensitivity analysis serves the different profiles of risk-taking, normal, and risk-averse investors by clearly defining how different investment quantities affect returns and risks. As evidenced by the peak return of 2.2845 at a $20,000 investment, the data shows that larger investment amounts can provide much higher returns for risk-taking investors who desire high returns and are ready to endure increased volatility. Despite the associated increase in risk, these investors would probably gravitate toward larger quantities to take advantage of the possibility of outsized gains. On the other hand, typical investors who look for a balanced strategy can use the results to maximize their investment in the $15,000–$20,000 range, where returns start to level out without unduly increasing risk. The model's success at smaller investment levels, including the $35,000 investment, which produced the lowest risk (0.0901) despite a lower return (2.1479), might reassure risk-averse investors. This tiered strategy helps investors of all risk levels make well-informed decisions according to their personal risk tolerance, which in turn helps them build a diversified portfolio that fits their risk appetite and financial goals. The model's versatility to different investor profiles is highlighted by the detailed knowledge of these processes, which increases its usefulness in the context of P2P lending.

5.1. Validation

In this section, the proposed model is compared with a single-objective model with the aim of reducing risk. In the single-objective model, the expected return is calculated without using deep learning, and resampling is not used.

Table 12 displays the findings of comparing the risk and return of the two models and demonstrates that the model put forth in this study has a higher return (more than 7%) with about the same risk. Because the suggested model is more appropriate for investors, it can therefore be validated.

Compared to conventional single-objective models, the suggested model greatly improves investment decision-making in P2P lending, mainly because of its multi-objective framework and incorporation of cutting-edge AI techniques. Conventional models usually ignore the intricacies and trade-offs present in actual financial decision-making and concentrate on achieving a single objective, such as maximizing return or minimizing risk. On the other hand, the suggested model uses multi-objective optimization to aim for both return maximization and risk minimization at the same time. In doing so, it offers a more comprehensive strategy that fits the diverse inclinations of risk-averse, risk-neutral, and risk-seeking investor profiles. Because of this flexibility, investors may more accurately determine their risk tolerance while aiming for larger returns, which eventually results in more thoughtful and calculated investing decisions.

Furthermore, the predictive power of the suggested model is improved using deep learning techniques like CNNs. The suggested model makes use of sophisticated data analytics to uncover intricate linkages in the data, in contrast to conventional models that frequently depend on linear assumptions and less sophisticated statistical techniques. This enables improved risk assessment and more precise NPV projection.

On the other hand, the problem of data imbalance, which has a major influence on risk prediction and portfolio optimization, is frequently ignored by traditional models for P2P lending investment decision-making. When used on highly imbalanced datasets, many traditional methods—such as logistic regression and machine learning classifiers—fail to capture intricate, nonlinear borrower-risk relationships. Because there are considerably more successful loans than defaulted loans in P2P lending, models tend to favor the majority class and generate skewed risk evaluations. Due to this mismatch, high-risk loans may be mistakenly classed as low-risk, which leads to poor investment decisions and, ultimately, an unfeasible portfolio optimization solution. Our study illustrated the shortcomings of conventional approaches in actual P2P lending contexts by showing that the optimization process was unable to produce feasible investment allocations when employing imbalanced data without resampling.

To improve classification accuracy and guarantee a more balanced dataset for model training, our work incorporates data resampling approaches, such as oversampling the minority class and undersampling the majority class. To increase model robustness and risk differentiation, we used SMOTE to provide realistic but synthetic examples of underrepresented categories. Through the use of these techniques, our model effectively generates feasible portfolio solutions while balancing risk and return. This result highlights our approach's superiority and uniqueness since it successfully overcomes the drawbacks of conventional models by facilitating accurate risk assessments and useful portfolio optimization. The power of our approach in digital lending markets is further validated by the capacity to turn an initially unfeasible problem into a possible investment strategy.

5.2. Financial implications

The experimental results of the proposed model indicate the portfolio return exceeding 7%, higher than the conventional methods with a comparable risk level. Understanding how this performance typically compares to more traditional investing options is essential to understanding the importance of this result. This section attempts to contextualize the financial performance of the suggested model by offering the general comparison, making it understandable to a wider audience while simultaneously emphasizing its potential significance.

When compared to more conventional asset classes, the model's annual return points to a potentially attractive investment opportunity. Because of their lower risk profile, government bonds often yield lower returns. Despite varying levels of risk, corporate bonds sometimes offer a marginally greater yield than government bonds. Although equities, or stocks, have the potential to yield larger profits in the long term, they are also much riskier and more volatile than bonds. As a result, the model's prediction of the return on our P2P lending portfolio places it in a position to potentially offer a competitive return.

The research points out that it is comparable to traditional methods when taking the risk into account. This implies that compared to other common investment types, the level of volatility or loss potential associated with this P2P lending portfolio is neither significantly higher nor lower. The model might be viewed as offering higher returns for a similar level of risk if the risk is closer to that of bonds. Investors might see the return as somewhat lower but possibly gain from diversification if the risk is closer to equities. Direct exposure to borrower default, which may behave differently from broader market risks affecting stocks and bonds, is the primary risk differential for P2P lending.

These findings have significant implications for investors and the online lending industry. A P2P lending strategy based on this approach might be a useful addition for investors wishing to diversify their portfolios beyond traditional assets. Particularly alluring may be the possibility of a high return, as well as a risk profile that isn't too high. Additionally, the model's ability to handle the intricacies of P2P lending suggests a more advanced approach for investing in this market. By offering tools for better informed and possibly profitable investment decisions, this kind of study can help the digital lending sector increase investor confidence.

6. Discussion

6.1. Managerial insights

Managerial insights derived from this research emphasize the significance of incorporating AI-powered technologies into investment decision-making procedures, particularly in dynamic and swiftly changing settings like P2P lending marketplaces and digital platforms. Using the information that this study generates, managers and decision-makers may create effective risk management plans and pinpoint areas in their investment portfolios that may be optimized. Additionally, the model's focus on multi-objective optimization emphasizes the necessity of a balanced strategy for making investment decisions that considers both return maximization and risk reduction goals at the same time. By using these insights, managers may adjust their investment plans and make sure they are in line with the larger goals and objectives of their company.

6.2. Practical implications

The practical implications of the AI-driven investment decision model extend to both institutional and individual investors by providing a data-driven approach for optimizing risk and return in P2P lending market. Conventional investment models frequently find it difficult to handle the intricacies of imbalanced data and fail to capture the complicated risk-return trade-offs inherent in digital lending platforms. The suggested approach improves decision-making by offering more accurate risk assessments and return projections by combining optimization techniques like NSGA-II with AI and machine learning techniques like logistic regression, and deep learning algorithms. Financial institutions and portfolio managers especially benefit from this capability since it enables them to create investment strategies that align with varying degrees of risk tolerance. Additionally, this model can be used by digital lending platforms to improve their risk assessment processes, which will raise the overall quality of loan recommendations and lower default rates.

For individual investors, the model provides a structured and intelligent framework to navigate the intricacies of P2P lending, which is frequently marked by a diverse borrower profile. The findings show that customized investment strategies can be advantageous for a variety of investor characteristics, from risk-averse to risk-tolerant. As demonstrated in the analysis, where returns increased significantly with larger investments, risk-seeking investors may prefer, for instance, larger investment amounts. On the other hand, risk-averse investors can give preference to portfolio configurations with the lowest risk, like the $35,000 investment that reduced risk. Because of the model's flexibility, investors can tailor their portfolios to their own risk tolerance and financial objectives. This personalized approach is especially helpful in democratizing investment opportunities by enabling a greater number of investors to engage in P2P lending with confidence and strategic precision.

6.3. Limitations of the proposed method

Even though this study offers a fresh combination of AI and OR for P2P lending investment decision-making, there are some limitations:

i. *Computational Complexity*: Because financial data is high-dimensional, it necessitates complex architectures and significant processing capacity to extract meaningful patterns. Scalability issues for real-time decision-making arise from the resource-intensive nature of NSGA-II portfolio optimization and deep learning model training.

ii. *Deep Learning Model Interpretability*: Despite increasing the accuracy of return estimation, deep learning's complex nature makes it challenging to communicate investment choices to stakeholders. Improving model interpretability is still a difficult task, especially in financial applications where regulatory compliance and transparency are crucial.

6.4. Future research perspectives

From a research perspective, this work creates opportunities for more investigation and improvement of AI-powered investment decision models concerning digital platforms and P2P lending. Subsequent investigations may concentrate on augmenting the forecast precision and resilience of the suggested model through the integration of supplementary data sources and sophisticated machine learning methodologies. Furthermore, there is room to investigate whether comparable concepts may be applied in businesses other than P2P lending, like alternative investment vehicles, equity markets, and crowdfunding platforms. Further research could be conducted to better understand the complex dynamics of risk and return in digital economies, considering variables including market volatility, regulatory settings, and technical improvements. Overall, this work lays the groundwork for future investigations into the use of AI and OR to improve investment decision-making in the digital world.

7. Conclusions

Digital platforms are at the forefront of utilizing cutting-edge technology, taking advantage of sophisticated operational research methods and artificial intelligence's technological breakthroughs to transform user experiences and spur innovation. As the global economy continues to evolve, emerging lending platforms are growing more and more appealing to investors looking for new opportunities for growth and diversification. P2P lending platforms are a global loan market that connects borrowers and lenders using internet applications, open to all individuals regardless of location or financial situation. One of the key factors that has contributed to the strengthening of the economic system is P2P lending through online trading platforms. Because P2P lending relies on estimating loan profitability and evaluating risk, it can be helpful in rating investments. When it comes to artificial intelligence and machine learning, the problem of imbalanced data that results in inefficient classification should be handled with extreme caution because it affects this kind of lending and causes credit risk assessment models to be less effective. This research has used the Lending Club platform dataset to offer a multi-objective recommender system for P2P lending investment decision. Based on a list of fully paid and defaulted loans, the data have been analyzed and interpreted. The performance results of the LGBM, random forest, decision tree, and logistic regression classification models have been compared in random oversampling and SMOTE mode to address data imbalance, and SMOTE sampling has been used at the end. The net present value was used to calculate the return on investment, and a convolutional neural network was utilized for prediction. The effectiveness of logistic regression was assessed by estimating the default probability of each loan using 10-fold cross-validation. Ultimately, the model was refined using the genetic algorithm, and the model's sensitivity was examined for varying investment amounts. The sensitivity analysis that was done emphasizes how important it is to take different investment levels into account while managing a portfolio. According to the findings, portfolio returns show a higher degree of responsiveness to changes in investment levels, even while risks remain a critical factor.

The research argues that predictive models not only enhance investment outcomes but also play an important role in aligning portfolio management with investor risk tolerance. Also, AI-driven strategies and advanced operational research techniques can be used to navigate the complexities of P2P lending businesses to provide clearer information for investors seeking to optimize returns and navigate emerging lending opportunities.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

None.

ORCID

Shahrooz Bamdad http://orcid.org/0000-0002-3217-1109

Data Availability statement

Not applicable.

References

Abba, S. (2022). AI in fintech: Personalized payment recommendations for enhanced user engagement. International Journal of Research In Computer *Applications and Information Technology*, *5*(01), 13–20.

Ahmadian, S., Ahmadian, M., & Jalili, M. (2022). A deep learning based trust- and tag-aware recommender system. *Neurocomputing*, *488*, 557–571. https://doi.org/10.1016/j.neucom.2021.11.064

Albehery, N., Helal, M. A., & Ghania, A. F. (2025). Inverse multi-objective optimization for portfolio allocation in commercial banks. *Statistics, Optimization & Information Computing*, *13*(3), 977–992. https://doi.org/10.19139/soic-2310-5070-2188

Alfonso-Sánchez, S., Solano, J., Correa-Bahnsen, A., Sendova, K. P., & Bravo, C. (2024). Optimizing credit limit adjustments under adversarial goals using reinforcement learning. *European Journal of Operational Research*, *315*(2), 802–817. https://doi.org/10.1016/j.ejor.2023.12.025

Asencios, R., Asencios, C. H., & Ramos, E. (2023). Profit scoring for credit unions using the multilayer perceptron, XGBoost and TabNet algorithms: Evidence from Peru. *Expert Systems with Applications*, *213*, 119201. https://doi.org/10.1016/j.eswa.2022.119201

Awad, M., Abouhawwash, M., & Agiza, H. N. (2022). On NSGA-II and NSGA-III in Portfolio Management. *Intelligent Automation & Soft Computing*, *32*(3), 1893–1904. https://doi.org/10.32604/iasc.2022.023510

Babaei, G., & Bamdad, S. (2020). A neural-network-based decisionmaking model in the peer-to-peer lending market. *Intelligent Systems in Accounting Finance and Management*, *7*, 142–150. https://doi.org/10.1002/isaf.1480

Babo, S. B., & Beyene, A. M. (2023). Bank loan classification of imbalanced dataset using machine learning approach. *Research Square*. https://doi.org/10.21203/rs.3.rs-2667057/v1

Bağcı, M., & Soylu, P. K. (2024). Optimal portfolio selection with volatility information for a high frequency rebalancing algorithm. *Financial Innovation*, *10*(1), 107. https://doi.org/10.1186/s40854-023-00590-3

Bhat, A. K. (2024). Application and impact of artificial intelligence in financial decision making. *International Journal of Scientific Research in Science, Engineering and Technology*, *11*(5), 57–63. https://doi.org/10.32628/IJSRSET2411417

Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2022). Deep neural networks and tabular data: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(6), 7499–7519. https://doi.org/10.1109/TNNLS.2022.3229161

Castro, R. (2022). Economic assessment of renewable energy projects. In *Electricity production from renewables* (pp. 105–123). Springer. https://doi.org/10.1007/978-3-030-82416-7

Cefis, M., & Carpita, M. (2025). A new xG model for football analytics. *Journal of the Operational Research Society*, *76*(1), 1–13. https://doi.org/10.1080/01605682.2024.2323669

Chawla, N. V., Japkowicz, N., & Kotcz, A. (2004). Editorial: Special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, *6*(1), 1–6. https://doi.org/10.1145/1007730.1007733

Chintala, S., & Thiyagarajan, V. (2023). AI-driven business intelligence: Unlocking the future of decision-making. *ESP International Journal of Advancements in Computational Technology*, *1*, 73–84. https://doi.org/10.56472/25838628/IJACT-V1I2P108

Dong, B., Zhou, Y., Sun, X., & Zhu, J. (2024). Textual analysis and credit scoring: A new matrix factorization approach. *Journal of the Operational Research Society*, 1–15. https://doi.org/10.1080/01605682.2024.2416908

Flores-Fernandez, G. A., Jimenez-Carrion, M., Gutierrez, F., & R. A., Sanchez-Ancajima (2022). Genetic algorithm and LSTM artificial neural network for investment portfolio optimization. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *15*(2), 27–46. https://doi.org/10.58346/JOWUA.2024.I2.003

García, F., González-Bueno, J., Guijarro, F., & Oliver, J. (2020). A multiobjective credibilistic portfolio selection model. Empirical study in the Latin American integrated market. *Entrepreneurship and Sustainability Issues*, *8*(2), 1027–1046. https://doi.org/10.9770/jesi.2020.8.2(62)

Gotardelo, D., & Goliatt, L. (2024). Multi-objective optimization of portfolio selection involving non-convex attributes in an anti-fragile perspective. *Evolving Systems*, *15*(3), 703–715. https://doi.org/10.1007/s12530-023-09497-x

Gozman, D., & Willcocks, L. (2019). The emerging Cloud Dilemma: Balancing innovation with cross-border privacy and outsourcing regulations. *Journal of Business Research*, *97*, 235–256. https://doi.org/10.1016/j.jbusres.2018.06.006

Grina, F., Elouedi, Z., & Lefevre, E. (2023). Re-sampling of multi-class imbalanced data using belief function theory and ensemble learning. *International Journal of Approximate Reasoning*, *156*, 1–15. https://doi.org/10.1016/j.ijar.2023.02.006

Ha, V. S., Lu, D. N., Choi, G. S., Nguyen, H. N., & Byeongnam Yoon, B. (2019). Improving credit risk prediction in online peer to peer (P2P) lending using feature selection with deep learning. In *2019 21st International Conference on Advanced Communication Technology (ICACT)* (pp. 511–515). IEEE.

Hamad, R. A., Kimura, M., & Lundström, J. (2020). Efficacy of imbalanced data handling methods on deep learning for smart homes environments. *SN Computer Science*, *1*(4), 204. https://doi.org/10.1007/s42979-020-00211-1

Hoogstra, B., Velichety, S., & Zhang, C. (2024). Developing a contextual model of poverty prediction using data science and analytics–The case of Shelby County. *Decision Support Systems*, *177*, 114080. https://doi.org/10.1016/j.dss.2023.114080

Hu, X., Chen, Y., Ren, L., & Xu, Z. (2023). Investor preference analysis: An online optimization approach with missing information. *Information Sciences*, *633*, 27–40. https://doi.org/10.1016/j.ins.2023.03.066

Huang, X., & Ma, D. (2023). Uncertain mean-chance model for portfolio selection with multiplicative background risk. *International Journal of Systems Science: Operations & Logistics*, *10*(1), 2158443. https://doi.org/10.1080/23302674.2022.2158443

Jones, T. W., & David Smith, J. (1982). An historical perspective of net present value and equivalent annual cost. *Accounting Historians Journal*, *9*(1), 103–110. https://doi.org/10.2308/0148-4184.9.1.103

Kaya, A., Pamucar, D., Gürler, H. E., & Ozcalici, M. (2024). Determining the financial performance of the firms in the Borsa Istanbul sustainability index: Integrating multi criteria decision making methods with simulation. *Financial Innovation*, *10*(1), 21. https://doi.org/10.1186/s40854-023-00512-3

Kely de Melo, M., R. T. N., Cardoso, & T. A., Jesus (2022). Multiobjective dynamic optimization of investment portfolio based on model predictive control. *SIAM Journal on Control and Optimization*, *60*(1), 104–123. https://doi.org/10.1137/20M1346420

Kim, J. Y., & Cho, S. B. (2019). *Predicting repayment of borrows in peer-to-peer social lending with deep dense convolutional network*. Wiley Online Library. https://doi.org/10.1111/exsy.12403

Ko, P. C., Lin, P. C., Do, H. T., & Huang, Y. F. (2022). P2P lending default prediction based on AI and statistical models. *Entropy (Basel, Switzerland)*, *24*(6), 801. https://doi.org/10.3390/e24060801

Koyyada, S. P., & Singh, T. P. (2023). A multi stage approach to handle class imbalance: An ensemble method. *Procedia Computer Science*, *218*, 2666–2674. https://doi.org/10.1016/j.procs.2023.01.239

Lee, J., & Sohn, S. Y. (2021). Recommendation system for technology convergence opportunities based on self-supervised representation learning. *Scientometrics*, *126*(1), 1–25. https://doi.org/10.1007/s11192-020-03731-y

Liu, J., & Wang, Z. (2025). Financial distress prediction using an improved particle swarm optimization wrapper feature selection method and tree boosting ensemble. *Journal of the Operational Research Society*, *76*(4), 617–640. https://doi.org/10.1080/01605682.2024.2385467

Liu, Y., Ma, H., Jiang, Y., & Li, Z. H. (2021). Learning to recommend via random walk with profile of loan and lender in P2P lending. *Expert Systems with Applications*, *174*, 114763. https://doi.org/10.1016/j.eswa.2021.114763

Lopez-Arevalo, I., Aldana-Bobadilla, E., Molina-Villegas, A., Galeana-Zapién, H., Muñiz-Sanchez, V., & Gausin-Valle, S. (2020). A memory-efficient encoding method for processing mixed-type data on machine learning. *Entropy (Basel, Switzerland)*, *22*(12), 1391. https://doi.org/10.3390/e22121391

Lu, Z., Zhitao, W., Minghuan, L., Yajun, D., & Runkai, N. (2024). How digital finance affects the financial asset allocation of brick-and-mortar businesses. *Economic Research - Ekonomska Istraživanja*, *37*(1), 2341219. https://doi.org/10.1080/1331677X.2024.2341219

Machado, M. R., & Karray, S. (2022). Assessing credit risk of commercial customers using hybrid machine learning algorithms. *Expert Systems with Applications*, *200*, 116889. https://doi.org/10.1016/j.eswa.2022.116889

Makridakis, S., Spiliotis, E., Assimakopoulos, V., Semenoglou, A. A., Mulder, G., & Nikolopoulos, K. (2023). Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. *Journal of the Operational Research Society*, *74*(3), 840–859. https://doi.org/10.1080/01605682.2022.2118629

Markowitz, H. (1959). *Portfolio selection: Efficient diversification of investments. Cowles Foundation Monograph 16, New Haven and London*. Yale University Press.

Mba, J. C. (2024). Assessing portfolio vulnerability to systemic risk: A vine copula and APARCH-DCC approach. *Financial Innovation*, *10*, 20. https://doi.org/10.1186/s40854-023-00559-2

Mishra, S. (2017). Handling imbalanced data: SMOTE vs. random undersampling. *International Research Journal of Engineering and Technology*, *4*(8), 317–320.

Muslim, Much Aziz, Nikmah, Tiara Lailatul, Pertiwi, Dwika Ananda Agustina, Dasril, Yosza, Subhan , Jumanto , Iswanto , (2023). New model combination meta-learner to improve accuracy prediction P2P lending with stacking ensemble learning. *Intelligent Systems with Applications*, *18*, 200204. https://doi.org/10.1016/j.iswa.2023.200204

Porfírio, J. A., Felício, J. A., & Carrilho, T. (2024). Factors affecting digital transformation in banking. *Journal of Business Research*, *171*, 114393. https://doi.org/10.1016/j.jbusres.2023.114393

Qin, J., Zeng, M., Wei, X., & Pedrycz, W. (2024). Ranking products through online reviews: A novel data-driven method based on interval type-2 fuzzy sets and sentiment analysis. *Journal of the Operational Research Society*, *75*(5), 860–873. https://doi.org/10.1080/01605682.2023.2215823

Rahmati, R., Neghabi, H., Bashiri, M., & Salari, M. (2024). Stochastic green profit-maximizing hub location problem. *Journal of the Operational Research Society*, *75*(1), 99–121. https://doi.org/10.1080/01605682.2023.2175734

Ren, J., Wang, Y., Cheung, Y. M., Gao, X., & Guo, X. (2023). Grouping-based oversampling in kernel space for imbalanced data classification. *Pattern Recognition*, *133*, 108992. https://doi.org/10.1016/j.patcog.2022.108992

Ren, T., Liu, S., G., Yan, & Mu, H. (2016). Temperature prediction of the molten salt collector tube using BP neural network. *IET Renewable Power Generation*, *10*(2), 212–220. https://doi.org/10.1049/iet-rpg.2015.0065

Shi, X., Wu, J., & Hollingsworth, J. (2019). How does P2P lending platform reputation affect lenders' decision in China? *International Journal of Bank Marketing*, *37*(7), 1566–1589. https://doi.org/10.1108/IJBM-11-2018-0309

Shih, W. L., Wei, C. C., Lin, H. H., & Chang, P. H. (2022). RETRACTED: Selecting and balancing market portfolio using artificial intelligence and fuzzy multiobjective decision-making model. *Concurrent Engineering*, *30*(4), 382–398. https://doi.org/10.1177/1063293X221115768

Sifrain, R. (2023). Predictive analysis of default risk in peer-to-peer lending platforms: Empirical evidence from lendingclub. *Journal of Financial Risk Management*, *12*(01), 28–49. https://doi.org/10.4236/jfrm.2023.121003

Sunardi, R., Hamidah, H., Buchdadi, A. D., & Purwana, D. (2022). Factors determining adoption of fintech peer-to-peer lending platform: An empirical study in Indonesia. *Journal of Asian Finance. Economics and Business*, *9*(1), 0043–0051. https://doi.org/10.13106/jafeb.2022.vol9.no1.0043

Syzdykov, M. (2020). *Peer-to-peer lending loan default prediction: Machine learning classification algorithms applied to Lending Club data, investors' perspective*. Digital repository of KAZGUU University. http://repository.kazguu.kz/handle/123456789/1004

Thejas, G. S., Hariprasad, Y., Iyengar, S. S., Sunitha, N. R., Badrinath, P., & Chennupati, S. (2022). An extension of synthetic minority oversampling technique based on Kalman filter for imbalanced datasets. *Machine Learning with Applications*, *8*, 100267. https://doi.org/10.1016/j.mlwa.2022.100267

Wang, H., Chen, K., Zhu, W., & Song, Z. (2015). A process model on P2P lending. *Financial Innovation*, *1*(1), 1–8. https://doi.org/10.1186/s40854-015-0002-9

Wang, H., Kou, G., & Peng, Y. (2021). Multi-class misclassification cost matrix for credit ratings in peer-to-peer lending. *Journal of the Operational Research Society*, *72*(4), 923–934. https://doi.org/10.1080/01605682.2019.1705193

Wang, X., Wang, B., Li, T., Li, H., & Watada, J. (2023). Multi-criteria fuzzy portfolio selection based on three-way decisions and cumulative prospect theory. *Applied Soft Computing*, *134*, 110033. https://doi.org/10.1016/j.asoc.2023.110033

Wei, Y., Kirkulak-Uludag, B., Zhu, D., Luo, X. (2023). *Stacking ensemble method for personal credit risk assessment in P2P lending*. SSRN. https://ssrn.com/abstract=4318348

Wu, S., Gao, X., & Zhou, W. (2022). COSLE: Cost sensitive loan evaluation for P2P lending. *Information Sciences*, *586*, 74–98. https://doi.org/10.1016/j.ins.2021.11.055

Yap, S. T. L. (2023). *Fintech market development in China. Contemporary strategic Chinese American business negotiations and market entry*. Palgrave Macmillan. https://doi.org/10.1007/978-981-19-6986-7_15

Yan, Y., Chen, M., Shyu, M. L., Chen, S. C. (2015). Deep learning for imbalanced multimedia data classification. In 2015 IEEE International Symposium on Multimedia (ISM) (pp. 483–488). IEEE. https://doi.org/10.1109/ISM.2015.126

Yoomak, S., Patcharoen, T., & Ngaopitakkul, A. (2019). Performance and economic evaluation of solar rooftop systems in different regions of Thailand. *Sustainability*, *11*(23), 6647. https://doi.org/10.3390/su11236647

Yu, W., Wen, X., Huberts, N. F., & Kort, P. M. (2024). Strategic investment under uncertainty: Why multi-option firms lose the preemption run. *Journal of the Operational Research Society*, *75*(9), 1855–1872. https://doi.org/10.1080/01605682.2023.2281535

Zao, Y., Zou, Z., & Shi, Z. (2024). Road graph extraction via transformer and topological representation. *IEEE Geoscience and Remote Sensing Letters*, *21*, 1–5. https://doi.org/10.1109/LGRS.2024.3380593

Zhang, H. (2024). Application of multi-objective optimization algorithm based on artificial fish school algorithm in financial investment portfolio problems. *Scalable Computing: Practice and Experience*, *25*(5), 3540–3546. https://doi.org/10.12694/scpe.v25i5.3122

Zhang, H., Kou, G., & Peng, Y. (2022). Large-scale peer-to-peer loan consensus based on minimum cost consensus. *Journal of the Operational Research Society*, *73*(10), 2326–2337. https://doi.org/10.1080/01605682.2021.1981782

Zhang, J., Calabrese, R., & Dong, Y. (2024). A novel generalised extreme value gradient boosting decision tree for the class imbalanced problem in credit scoring. *Journal of the Operational Research Society*, 1–18. https://doi.org/10.1080/01605682.2024.2418882

Zhao, H., Ge, Y., Liu, Q., Wang, G., Chen, E., & Zhang, H. (2017). P2P lending survey: Platforms, recent advances and prospects. *ACM Transactions on Intelligent Systems and Technology*, *8*(6), 1–28. https://doi.org/10.1145/3078848

Zhu, Y., & Wu, D. (2025). P2P credit risk management with KG-GNN: A knowledge graph and graph neural network-based approach. *Journal of the Operational Research Society*, *76*(5), 866–880. https://doi.org/10.1080/01605682.2024.2398762

Ziane, M., Sara, C., Fatima, B., Abdelhakim, C., & EL Moutaouakil, K. (2024). Portfolio selection problem: Main knowledge and models (A systematic review). *Statistics, Optimization & Information Computing*, *12*(3), 799–816. https://doi.org/10.19139/soic-2310-5070-1961