Knowledge-Based Systems 295 (2024) 111761

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

An explainable data-driven decision support framework for strategic customer development

Mohsen Abbaspour Onari [a,b,*], Mustafa Jahangoshai Rezaee [c], Morteza Saberi [d], Marco S. Nobile [e]

[a] Information Systems Group, Eindhoven University of Technology, Eindhoven, The Netherlands
[b] Eindhoven Artificial Intelligence Systems Institute, Eindhoven, The Netherlands
[c] Faculty of Industrial Engineering Department, Urmia University of Technology, Urmia, Iran
[d] School of Information, Systems, and Modelling, University of Technology Sydney, Sydney, Australia
[e] Environmental Sciences, Informatics, and Statistics, Ca' Foscari University of Venice, Venice, Italy

ARTICLE INFO

Keywords:
Credit scoring
Customer development
Decision support framework
Explainable AI
Prisoner's dilemma
Target setting

ABSTRACT

Financial institutions benefit from the advanced predictive performance of machine learning algorithms in automatic decision-making for credit scoring. However, two main challenges hamper machine learning algorithms' applicability in practice: the complex and black-box nature of algorithms that hinder their understandability and the inability to guide rejected customers to have a successful application. Regarding customer relationship management is one of the main responsibilities of financial institutions; they must clarify the decision-making process to guide them. However, financial institutions are not willing to disclose their decision-making procedure to prevent potential risks from customers or competitors side. Hence, in this study, a decision support framework is proposed to clarify the decision-making process and model strategic decision-making to guide rejected customers simultaneously. To do so, after classifying customers in their corresponding groups, the capability of Shapley additive exPlanations method is exploited to extract the most impactful features to the prediction's outcome globally and locally. Then, based on the benchmarking approach, the equivalent approved peer is found for the rejected customer for target setting to modify the application. To find the optimal modified values for a counterfactual prediction, a multi-objective gamed-based counterfactual explanation model is developed using the prisoner's dilemma game as the constraint to simulate strategic decision-making. After optimization, the decision is reported to the customers concerning the credential background. A public data set is used to elaborate on the proposed framework. This framework can generate counterfactual predictions successfully by modifying perspective features.

1. Introduction

1.1. Problem description

Credit lending products, such as credit cards, personal loans, mortgages, and corporate loans, are the primary business for banks and other financial institutions. Accordingly, good lending practices lead to high profits [1]. The credit scoring models have been developed for classifying loan customers as either a good credit group (approved) or a bad credit group (rejected). These models are based on customers' related characteristics such as age, income, and marital status or the data of the previously approved and rejected applicants [2]. It is possible to describe the advantages of using credit scoring models as reducing the cost of credit analysis, enabling faster credit decisions, ensuring credit collections, and diminishing possible risk [3].

Credit scoring can be regarded as a binary classification problem of instances into one of the two pre-defined groups. Thus, Machine Learning (ML) models can showcase their superior performance compared to traditional (statistical) methods in dealing with credit scoring problems, especially in nonlinear pattern classification [2]. However, there are two serious concerns about using ML models [4]:

- The predictive model's objective is to provide accurate predictions of the outcomes from a set of observable features, while the Decision Makers (DMs) seek to make decisions that maximize a given utility function.
- Although there is increasing excitement about using data-driven predictive models to improve decision-making in high-stakes

[*] Corresponding author.
E-mail addresses: m.abbaspour.onari@tue.nl (M.A. Onari), m.jahangoshai@uut.ac.ir (M.J. Rezaee), Morteza.Saberi@uts.edu.au (M. Saberi), marco.nobile@unive.it (M.S. Nobile).

https://doi.org/10.1016/j.knosys.2024.111761
Received 19 July 2023; Received in revised form 6 December 2023; Accepted 3 April 2024
Available online 16 April 2024

applications, there is also a heated debate about their lack of transparency and explainability.

Hence, the presence of expert domain knowledge and explainable tools is a way forward in fair decision-making and monitoring proper ML performance, both of which have attracted limited attention in the literature.

Increasing attention in the literature to improve ML models' predictive power causes ignoring Customer Relationship Management (CRM), which can change a customer's relationship with a company and increase revenues in the bargain [5]. An increasing number of companies are embracing strategies, programs, tools, and technology that prioritize customers for efficient and effective CRM. Recognizing the importance of comprehensive and integrated customer insights, these companies aim to foster close, cooperative, and partnership-based customer relationships [6]. CRM involves strategically choosing customers that a company can serve most profitably and managing the interactions between the company and these customers. The ultimate objective is to optimize the present and future value of customers for the company [7]. A crucial aspect of CRM involves identifying distinct customer types, followed by formulating targeted strategies for engaging with each category. These strategies may encompass enhancing connections with lucrative customers, attracting new and profitable clientele, and devising suitable approaches for unprofitable customers, which could involve discontinuing relationships that result in financial losses for the company [7]. Thus, it can be stated that credit scoring is not the only task of financial institutions, and they should guide rejected customers to have a successful application in the future as well. Presenting findings from a survey targeting financial institutions that had implemented and were actively utilizing CRM, Ryals [8] asserts that every participant interviewed underscored the pivotal role of CRM in their business management. Moreover, they foresaw its growing influence in the future. Numerous respondents stressed that enhancing the management of customer relationships was not merely a choice but a necessity. There were concerns expressed that sustaining their businesses would be challenging without discovering improved methods of managing customer relationships. Consequently, they anticipate gaining a substantial competitive advantage through this capability in the financial services marketplace. To accomplish this objective, financial institutions aim to identify their most valuable customers, retain their loyalty, and enhance their "share of wallet" by determining additional services or products that may appeal to them. The goal is to foster customer-centric or one-to-one relationships and boost shareholder value. However, at the core of these aspirations lies the imperative of managing existing and potential customers more effectively [9]. This necessitates access to information that aids in making optimal decisions for cultivating and overseeing appropriate relationships, mitigating risks, controlling costs, and navigating markets. Understanding customer behavior and preferences empowers financial institutions to reconfigure core product offerings and develop suitable channel strategies. In essence, recognizing current customers' value and prospective customers' potential long-term value is crucial for financial institutions. Neglecting customer needs and treating all customers uniformly can result in costly investment mistakes [9]. Hence, CRM in financial institutions is a critical strategy twofold. First, customers are a valuable asset, and acquiring new customers is expensive and time-consuming. Second, it can increase loyalty, customer satisfaction, and revenue, justifying "customer loyalty at any cost—even if we don't see a return on investment [5]". Hence, if DMs advocate for their customers through rational behavior, they will reciprocate with their trust and loyalty, either now or in the future [10].

1.2. Research questions

The first important matter for customers in the case of application rejection is finding the main reasons for rejection. It has been proven that responsiveness has a direct relationship with customer satisfaction, revisits intentions, and referral behavior [11]. Regarding one of the pursued strategic goals of CRM programs is enhancing customer loyalty/satisfaction/engagement [12], transparency in the decision-making [13], and ethical perceptions [14] of institutions can improve customer trust [15], satisfaction, and loyalty. It becomes significantly more crucial due to the legal requirement outlined in the European Union General Data Protection Regulation (GDPR) that mandates a "right to explanation" for decisions produced by automated and AI algorithmic systems [16]. Here, the presence of DMs is necessary to explain the main shortcomings in the application to the customer because developed data-driven credit scoring models only predict the result of the application. However, these models have a black-box nature without providing inherent interpretability because of training with big heterogeneous data sets. Conversely, these models are not decision-support systems for post-decision making. Hence, it is possible to define the first research question:

- RQ1: How is it possible to find reasons causing the application's rejection?

The decision is normally reflected in verbal terms by DMs to customers, which can cause another question of what changes should be applied to have a successful application or how much modification is sufficient. Although it would initially be construed that it is easy to answer these questions to guide customers, it is very important to have strategic behavior. Allenspach [17] demonstrated that revealing clear information about the interim condition of a financial institution could have adverse effects. Consequently, surpassing a specific threshold of transparency might result in the ineffective liquidation of a bank. If the answer is not informative enough, it might cause customer churn because they would never trust the institution anymore. Conversely, if more information is presented to customers, they might abuse information against the institution. The response is highly dependent on the customers' credential background, and the credit manager should not reveal the decision-making procedure to the customer due to the possibility of fraud. Hence, three conditions must be met by credit managers: first, disclosing a safe amount of information to avoid fraud. Second, reflecting in a way that embeds enough information for customers without strict bias about their background. Third, the modification should be actionable and feasible for the customers. Hence, the second research question can be defined as:

- RQ2: How is it possible to model strategic behavior to help customers?

Buttle [12] also claims that answering the aforementioned question supports strategic and tactical decision-making in analytical CRM by AI and ML models. After finding the answer for RQ2, credit managers can rest assured that their answer will not cause fraud in the institution and can help customers modify their application to have loyal and trustworthy customers.

1.3. Solution

A decision support framework is proposed in this study to help rejected customers to have a successful application based on the credit manager's strategic decision-making and their credential background while avoiding the risk of fraud. To do so, the capability of the SHapley Additive exPlanations (SHAP) method is exploited to extract the impact of features on the prediction of black-box ML models. SHAP can be used for all ML models, and it provides local explanations that are very useful to have explanations for every single instance independently. In order to answer RQ1, the benchmarking approach is followed by finding the approved peer of the rejected customer by calculating their Euclidean distance, which we showed its efficiency in a similar application before [18]. Thus, it is possible to identify the shortcomings of the application by comparing its features pairwise with those of the approved peer.

To address RQ2, the decision-making process of a credit manager is modeled as a Multi-Objective Game-based Counterfactual Explanation (MOGCE) to embed triple predefined conditions. The objective function is defined as a multi-objective problem by setting targets to reduce the distance of rejected customer's actionable features' to the approved peer. The idea is to find the least amount of modification helpful for a rejected application to turn into an approved one. The point is to inform this information to customers strategically to avoid disclosing the decision-making procedure. To do so, the most famous two-person mixed-motive game, Prisoner's Dilemma (PD), is used as the constraint of the MOGCE to satisfy the conditions of RQ2. The institution has two strategies: trust in the customer to cooperate or distrust them. Conversely, the customer might show trustworthy behavior in cooperation to follow the institute's rules precisely or defect them. PD constraints on MOGCE guarantee that if the customer decides to deceive the institution, the loss payoff for the institution will be the minimum possible amount, and if she decides to be trustworthy, the payoff for both of them will be optimal. Then, the developed MOGCE model is optimized by Multi-Objective Particle Swarm Optimization (MOPSO) to obtain the optimal solutions. Finally, the decision is reported to the customer based on their credential background. If the credit manager trusts the customer, they can report the exact amount of modification to the customer. Otherwise, fuzzy linguistic terms can be used to hide the exact modification.

1.4. Contributions

The technical aspect of developing the ML model for credit scoring is out of this paper's scope because there are already many powerful models in the literature. This study has focused on developing a decision support framework to help rejected customers to have a successful application based on the credit manager's strategic decision-making. The contributions of this paper can be listed as follows:

- C1: To the best of our knowledge, this work is the first attempt in the literature to propose a framework for helping rejected customers to approve their applications.
- C2: The impact of credit manager strategic behavior has been missed in the literature due to focusing on developing predictive data-driven credit scoring models, which is not decision-making.
- C3: There is less attention to developing explainable credit scoring models in the literature. This study contributes by highlighting the impact of XAI methods for explaining black-box models and helping credit managers in decision-making. This approach uses two model-agnostic and post-hoc explainable models suitable for all ML models.
- C4: In order to answer RQ1, a benchmarking approach has been implemented based on Euclidean distance to obtain the required information about the target setting for the rejected customers.
- C5: Informing optimal decisions to customers is modeled as a two-person PD game for a better simulation of the real-world strategic behavior based on Counterfactual Explanation (CE). This model is multi-objective and NP-hard, but the results of this paper show that its generalizability outperforms its hard solution.

The remainder of this paper is structured as follows. In the next section, the related literature is reviewed. Then, in Section 3, the implemented methods are introduced. In Section 4, the proposed decision support framework is presented. Subsequently, in Section 5, the proposed approach is elaborated based on a public data set. Finally, in Section 6, conclusions and future research opportunities are presented.

2. Literature review

In this section, the literature on credit scoring is covered. In Section 2.1, the implemented supervised ML algorithms in the credit scoring domain are reviewed. Then, in Section 2.2, the recent state-of-the-art developed methods of credit scoring are covered.

2.1. Supervised ML algorithms in credit scoring

In the literature, various ML models have been developed for credit scoring. Logistic Regression (LR) is the traditional method for financial institutions to score borrowers due to its simplicity and transparency [19]. However, it does not have conspicuous performance in dealing with complex data sets with nonlinear behavior. Hence, it is necessary to use ML models with high predictive performance capability. After comparing multiple feature selection methods to examine their prediction performance, Tsai [20] used Multi-Layer Perceptron (MLP) as the prediction model. Zhang et al. [21] proposed a Vertical Bagging Decision Trees Model (VBDTM) with a new bagging method different from traditional bagging. To optimize feature space, Chen and Li [22] proposed a hybrid credit scoring approach based on Support Vector Machine (SVM), conventional statistical LDA, DT, rough sets, and F-score approaches as feature preprocessing steps. Wang et al. [23] compared the performance of three popular ensemble methods, Bagging, Boosting, and Stacking, based on four base learners, LR, DT, Artificial Neural Network (ANN), and SVM, to find the best predictor. Two dual strategy ensemble trees: RS-Bagging DT and Bagging-RS DT, were proposed by Wang et al. [24] to decrease the noisy data effect and the redundant attributes of data to get a relatively higher classification accuracy. A Bayesian latent variable model with a classification and regression tree approach was built by Kao et al. [25], which successfully could outperform conventional credit scoring methods. Han et al. [26] introduced a new way to address LR, and SVM suffers from the curse of dimension, defined as orthogonal dimension reduction. In order to produce accurate and comprehensible models, Florez-Lopez and Ramon-Jeronimo [27] proposed an ensemble approach based on merged DT and Correlated-Adjusted Decision Forest (CADF). Ala'raj and Abbod [28] implemented five well-known base classifiers, namely, ANN, SVM, Random Forests (RF), DT, and naïve Bayes (NB), to present a new combination approach based on classifier consensus to combine multiple classifier systems of different classification algorithms. Xia et al. [29] integrated the bagging algorithm with the stacking method to propose a novel heterogeneous ensemble credit model, which differs from the extant ensemble credit models in three aspects: pool generation, selection of base learners, and trainable fuser.

2.2. Recent novel approaches for credit scoring

Reviewing the credit scoring literature shows that most credit lending decisions are made based on two levels of abstractions, i.e., either to lend credit or not to lend credit [30]. Accordingly, developing a myriad of ML classification models causes ignoring the importance of designing decision supports for DMs to guide rejected customers to have a successful application. Reviewing the state-of-the-art recent studies reveals that this aspect of the problem still has been missed. Herasymovych et al. [31] evaluated the capability of reinforcement learning to optimize the acceptance threshold of a credit score leads to higher profits for the lender. Pławiak et al. [32] proposed a novel hybrid approach based on a deep genetic cascade ensemble of SVM classifiers, which merges the benefits of evolutionary computation, ensemble learning, and deep learning. In order to fill the existing gap of utilizing advanced tree-based classifiers as components of ensemble models and considering dynamic ensemble selection, Xia et al. [33] developed a novel tree-based overfitting-cautious heterogeneous ensemble model for credit scoring. Tripathi et al. [34] proposed a novel activation function and an evolutionary approach to get optimized weights and biases by utilizing the Bat optimization algorithm for the extreme learning machine used for the credit risk evaluation model. In order to address reject inference, Shen et al. [35] proposed a novel three-stage learning framework implementing unsupervised transfer learning and three-way decision theory and to learn higher-level representations for the credit risk classification task employing a self-taught learning technique. Wu

Table 1
Comparison of implemented approaches in credit scoring literature.

Article	Solution approach	C1	C2	C3	C4	C5
Gorzałczany and Rudziński [40]	Fuzzy Rule-based Classifiers			✓		
Xia et al. [41]	Bayesian Hyperparameter Optimization			✓		
Lee et al. [42]	Rule-based ML		✓	✓		
Lan et al. [43]	Bayesian Network		✓			
Tezerjan et al. [44]	Fuzzy Rule-based Model		✓			
Moscato et al. [45]	Benchmarking			✓	✓	
Lappas et al. [46]	Clustering and Genetic Algorithm		✓	✓		
Visani et al. [47]	LIME			✓		
Dastile et al. [48]	Counterfactual Explanation		✓	✓		
Bueff et al. [49]	Counterfactual Constraints			✓		
Dumitrescu et al. [50]	Penalised Logistic Tree Regression			✓		
Bücker et al. [51]	Multiple XAI methods			✓		
Current paper	SHAP and Counterfactual Explanation	✓	✓	✓	✓	✓

et al. [36] applied a deep multiple kernel classifier as a state-of-the-art technique, which is proficient in dealing with deep structure and complex data that outperforms conventional and ensemble models. A Graph Convolutional Network (GCN)-based credit default prediction model was proposed by Lee et al. [37] to reflect nonlinear relationships between borrower's attributes and default risk and high-order relationships between the borrowers. Maldonado et al. [38] proposed a novel Fuzzy SVM strategy in which the traditional hinge loss function is redefined to account for data set shift. Djeundje et al. [39] demonstrated that a model containing email usage, psychometric variables, and demographic variables could successfully give higher predictive accuracy than a model that implements demographic data in credit scoring.

The remaining approaches have been presented in Table 1. These papers have been monitored based on the contributions of the current paper. This table demonstrates that although focusing on the explainability of ML predictors has increased recently, other contributions have not been addressed yet in the literature.

The developed decision support framework holds a distinctive advantage by addressing critical gaps in the existing literature related to helping rejected customers in a credit scoring application. Through the integration of XAI techniques, our framework illuminates the primary deficiencies in applications leading to rejection—a contribution not previously explored in the literature. Additionally, we extend the literature by introducing strategic decision-making through the identification and communication of optimal modifications for rejected applicants. This strategic behavior, facilitated by optimizing the MOGCE model, not only aids credit managers in identifying suboptimal customers among rejections but also guides them in proposing actionable solutions to enhance the application. The framework's unique proposition lies in its ability to empower rejected customers to modify their applications effectively, ensuring compliance with the MOGCE model's conditions and the background credentials of the rejected customer. This approach not only mitigates the risk of compromising the institution's decision-making process but also provides a secure and informative avenue for rejected applicants to enhance their application and achieve successful outcomes.

3. Methodology

In this section, the implemented methods in this study are introduced. In Section 3.1, the concept of SHAP post-hoc explainable model is presented. In Section 3.2, the multi-objective CE model is introduced. Finally, in Section 3.3, the concept of the two-person PD game is covered.

3.1. SHapley additive exPlanations

Lundberg and Lee [52] borrowed the concept of Shapley value from coalitional game theory to develop a post-hoc model-agnostic explainable method to open the black-box of ML algorithms called SHAP. The coalitional game is a cooperative game, and Shapely value as a measurement can calculate power distribution among factions in coalitions [53]. If we intend to map this concept as an XAI approach, we can consider a game as a prediction model for a single instance and players as feature values of the instance collaborating to receive gain. This gain is the difference between the prediction's Shapley value and the average of the Shapley values of the predictions among the feature values of the instance to be explained [54].

SHAP for calculating the importance score of any given instance i of feature X considers all feature subsets except X itself and then computes the effect on predictions of adding $X(i)$ to all those subsets [55]. Although this approach makes SHAP an inherently local explainable approach, it can provide global explanations considering all sets of instances on which the model has been trained as well [56].

3.2. Multi-objective counterfactual explanation

CE is a post-hoc method that provides information to users on what they might do to change the outcome of an automated decision [57]. Wachter et al. [58] introduced CE for the first time as an optimization problem. The easiest way to grab the CE idea is by recalling the classic example of a customer who seeks a home mortgage loan in a bank. An ML classifier predicts outcomes as final decisions by considering the customer's feature vector of $\{Income, Credit, Sex, Age, Marital, Education\}$. Whenever the customer is denied the loan that it seeks, the following questions arise: (i) why was the loan denied? and (ii) what actions he/she can take differently in the future to approve the loan? To address the first question, it is possible to provide the following explanation: "Income was not satisfying". The latter question forms the basis of a CE: what small modifications could be feasible for the customer to acquire validation to obtain the loan? For example, the customer can increase its credit [59]. The formalized objective is optimized by minimizing the distance between the counterfactual (x') and the original data point (x), subject to the constraint that the output of the classifier on the counterfactual is the desired label ($y' \in Y$). Dandl et al. [60] proposed the concept of Multi-objective Counterfactual Explanations (MOCE) to formalize the counterfactual search as a multi-objective optimization problem. MOCE returns a Pareto set of counterfactuals representing different trade-offs between proposed objectives, which are diverse in feature space. Changing to different features can lead to a desired counterfactual prediction, which seems preferable, and it is more likely that some counterfactuals meet a user's (hidden) preferences. Moreover, if multiple otherwise quite different counterfactuals propose changes to the same feature, the user rests assured that the feature is a significant lever to attain the desired outcome [60].

3.3. Prisoner's dilemma

As a discipline focused on strategic decision-making, game theory serves as a powerful tool for understanding the dynamics of

Table 2
The general form of the PD payoff matrix.

		Player B	
		Cooperation	Defection
Player A	Cooperation	$R = 3$	$S = 0$
	Defection	$T = 5$	$P = 1$

Table 3
Payoff matrix for each move of the PD.

		Player B	
		Cooperation	Defection
Player A	Cooperation	3,3	0,5
	Defection	5,0	1,1

relationships formed and dissolved in the realms of competition and cooperation [61,62]. Players attempt to maximize their individual benefit in the game, knowing that the outcome is the product of all the decisions made. The Prisoner's Dilemma (PD) is the most famous 2 × 2 game of all, which, as a minimal game, demonstrates that individually beneficial actions are socially harmful [63]. In the PD game, two players can each either cooperate or defect. The selfish choice of defection yields a higher payoff than cooperation, no matter the other player's action. However, if both defects, both do worse than if both had cooperated [64]. In any given round of PD, the two players receive R points if both cooperate (the reward for mutual cooperation) and only P points if both defect (the punishment for mutual defection). On the other side, a defector exploiting a cooperator gets T points (the temptation to defect), while the cooperator receives S (the sucker's payoff) (See Table 2) [64,65]. A payoff matrix for the PD is subject to two restrictions:

1. $T > R > P > S$
2. $R > (T + S)/2$

The PD game is a non-cooperative game in which communication among players is forbidden, impossible, or irrelevant. When both players select their dominant strategy, given these assumptions, they produce an equilibrium that is the third-best result for both (P). Here, none of the players has the incentive to change that is independent of the strategy choice of the other [66]. Thus, in a single round, the best strategy is always defection (See Table 3).

However, neither side can benefit itself with a selfish choice sufficiently in the short run to compensate for the harm committed to it from a selfish choice by the other player. This characteristic makes the PD distinctive and gives both sides an incentive to defect. The main point is that if both do defection, both do poorly. Therefore, the PD embodies the tension between individual rationality to be selfish and group rationality to both sides for mutual cooperation over mutual defection [67]. A Pareto-optimal outcome occurs whenever no different outcome is strictly favorable for at least one player that is at least as good for the others. In the two-person PD game, the (Cooperation, Cooperation) outcome is more favorable for both players than the (Defection, Defection) outcome. Accordingly, the equilibrium outcome is Pareto-inferior [66].

4. Proposed approach

The proposed decision support framework in this study has been developed in six steps. In Section 4.1, the data set is prepared to feed the ML models by applying different preprocessing techniques, including handling missing values, feature engineering, and data normalization. Then, among different exploited ML models, the one with the highest predictive performance is selected. Next, in Section 4.2, the benchmarking approach is used to select a rejected reference customer and its approved peer using Euclidean distance. In Section 4.3, SHAP technique is used to determine the features' contribution to classifying selected customers in their corresponding groups. Afterward, in Section 4.4, MOGCE is developed based on actionable features extracted by SHAP. Then, in Section 4.5, MOPSO algorithm is implemented to optimize the MOGCE and generate optimal solutions for modifying the suboptimal rejected customer's application. Finally, in Section 4.6, the decision is reported to the customer based on the optimized model and credential background. In the following, we are going to elaborate on each step.

4.1. Classifying customers based on ML model

For the first step, applying the preprocessing phase to the data set is important. The preprocessing phase can include tackling missing values, feature engineering, and data normalization. Then, the ML models are trained to classify customers into approved/rejected classes. This study does not seek to develop advanced ML models with high predictive performance. Hence, different ML models are exploited in the data set, and the best model in terms of predictive performance is selected to continue the framework.

4.2. Distance measure to find the equivalent approved customer

After classifying customers with the best ML model into their corresponding groups, a customer among the rejected group that we would like to discover its application shortcomings is selected for benchmarking as the reference customer. Then, the equivalent peer among approved customers is found using Euclidean distance. The approved peer customer has the least Euclidean distance with the reference rejected customer. Here, features are considered as elements of Euclidean distance. In benchmarking, the main goal of target setting is to try to reduce the distance between the rejected customer's elements and the approved ones. This condition can be met by only numeric features because of two reasons. First, Euclidean distance is only suitable for numeric features. Second, according to the counterfactual logic, it does not make sense to modify categorical features because it might be impossible or illogical.

4.3. Feature importance based on post-hoc SHAP method

After selecting the reference rejected customer and finding its approved peer, the most important features that contribute to classifying them in their corresponding groups are extracted by SHAP. SHAP represents explanations of the classifier's predictions, which can be understandable for DMs. SHAP determines the contribution of each feature to the classification's outcome. Here, by analyzing the features of both customers, we can figure out the shortcomings of the rejected customer's application and the strengths of the approved one. Then, it would be possible to adjust the weak features of the rejected customer by setting targets with respect to the approved peer to improve them. After adjustment of the rejected customer's weak features, it would be possible to turn its application into an approved one.

4.4. Multi-objective game-based counterfactual explanation model

After implementing SHAP to extract features' contribution to classify customers in their corresponding groups, it is crucial to determine actionable and feasible features for target setting. The main requirement to do so is that the target setting should make sense and be actionable in practice based on the customer's application. To elucidate, consider asking the rejected customer to change their gender or divorce their partner to approve the application. This adjustment is impossible and has a bias against the customer, which is unethical. Hence, some features (especially categorical features) are dropped from the target setting. After determining the actionable features, the institution must make strategic decision-making for the customers, especially those with credential background issues, to report the required adjustment of the application. The logic behind it is they might threaten decision-making procedures open to leaks or the possibility of fraud. The MOGCE model is developed based on a two-person PD game to address these two prerequisites simultaneously. The institution has two strategies: either trust the customer or distrust them when it wants to disclose information regarding modifying applications. On the other hand, it should be considered that customers might be trustworthy with the institution or try to deceive it by grasping more information. However, it is not easy for credit managers to confidently estimate the customers' intentions because it depends on various criteria that are impossible to model. It is possible to have a strategy based on the credential background of the customers, but it has not been guaranteed that customers act exactly like their historical behavior. For example, suppose a customer has a credit issue based on their historical collaboration with the institution. In that case, they might change their strategy, be honest, and follow the rules to have a successful application. Here, the institution should have its own strategy to minimize the possibility of customer fraud. On the other side, the institution must consider that a wrong strategy with an honest customer may cause customer churn, which is not beneficial in the long term. Here, the PD suits our expected assumptions to model the MOGCE. The obtained payoff values for the PD by MOGCE guarantee that the loss payoff for the institution is the minimum possible amount in the case of fraudulent behavior from the customer's side. On the other side, it guarantees that the trust payoff in the case of trustworthy behavior by the customer is optimal for both parties. The payoffs in this are normalized values in [0, 1] intervals because features might have different scales. Hence, they are normalized during the optimization to have a standard scale for comparison. The developed MOGCE model is presented in Eq. (1):

$$\min_{\forall i,r}(\|x_{io} - \Delta x_{io} - x^*_{ip}\|, \|y_{ro} + \Delta y_{ro} - y^*_{rp}\|)$$

subject to:

$$R = \alpha_1 \left(\sum_{i=1}^{n} \beta_i (x_{io} - \Delta x_{io}) + \sum_{r=1}^{n} \beta_r (y_{ro} + \Delta y_{ro}) \right)$$

$$S = \sum_{i=1}^{n} \beta_i x^*_{ip} + \sum_{r=1}^{n} \beta_r y_{ro}$$

$$T = \sum_{i=1}^{n} \beta_i x_{io} + \sum_{r=1}^{n} \beta_r y^*_{rp}$$

$$P = \alpha_2 \left(\sum_{i=1}^{n} \beta_i (x_{io} - \Delta x_{io}) + \sum_{r=1}^{n} \beta_r (y_{ro} + \Delta y_{ro}) \right) \tag{1}$$

First PD constraint:

$$T - R > 0$$

$$R - P > 0$$

$$P - S > 0$$

Second PD constraint:

$$2R - T - S > 0$$

$$x_{io} - \Delta x_{io} \geq x^*_{ip}$$

$$y_{ro} + \Delta y_{ro} \leq y^*_{rp}$$

$$\Delta x_{io}, \Delta y_{ro} \geq 0$$

In this model, x_{io} and y_{ro} indicate the actionable features for the rejected customer that should be decreased and increased in the target setting process, respectively. In the same way, x^*_{io} and y^*_{rp} refer to desired values for the corresponding actionable features based on the approved peer in the target setting process. Hence, the objective function would minimize the distance between the rejected and approved applicants based on their actionable features by optimizing the values of Δx_{io} and Δy_{rp}. This model should satisfy the dual constraints of the PD game. To do so, β, which is the impact of each actionable feature extracted by SHAP, is used to model them. Besides, α_1 and α_2 ($\alpha_1 > \alpha_2$) are balance criteria that are automatically tuned during the optimization process to satisfy $R > P$ condition in the PD game.

4.5. Optimizing the model by multi-objective particle swarm optimization

After developing MOGCE, MOPSO introduced by Coello et al. [68] is implemented to optimize the developed mathematical model. MOPSO is the extension of the PSO algorithm, a population-based metaheuristic algorithm introduced by Kennedy and Eberhart [69]. The experimental results show that the PSO can converge fast because it does not involve selection operation or mutation calculation, so the search can be performed by repeatedly varying the particle's speed. Also, the performance of PSO is not susceptible to the population size, and PSO scales well [70]. PSO generates new solutions based on two equations:

$$v_i(t+1) = w * v_i(t) + c_1 * rnd() * (pbest_i - x_i(t)) + c_2 * rnd() * (gbest(t) - x_i(t)) \tag{2}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{3}$$

where c_1 and c_2 denote the acceleration constant for weighting the stochastic acceleration terms that pull a single particle toward personal best ($pbest$) global best ($gbest$) positions. $rand()$ indicates a random variable that is generated by uniform distribution between [0, 1]. w, x, v refer to inertia weight, the position vector, and velocity vector, respectively [61].

The multi-objective procedures should supply two main properties. First, generating high-quality nondominated solutions on the Pareto frontier of the Multi-Objective Decision-Making (MODM) problem [71]. Second, concerning a proper diversity for the generated solutions on the Pareto frontier of MODM problem [71]. For selecting the individual best of each particle, a single position $pbest$ is maintained and only is replaced if x_i is better than p_i. In the meantime, selecting the best position group can be performed randomly [72]. MOPSO algorithm has been selected for this study because the proposed model is non-linear, and multiple solutions are required. Besides, it is a popular algorithm for non-linear programming problems [73]. Finally, among multiple Pareto optimal solutions, only one solution with the least cost function is selected as the best solution.

4.6. Reporting the decision by credit manager with respect to the credential background

After optimizing the model and generating optimal solutions, the disclosing information based on the customer's credential background must be evaluated before informing the decision. To do so, there are two strategies for customers with positive credential backgrounds and negative ones. As we mentioned earlier, even though the (Defection, Defection) is considered the Pareto-inferior equilibrium point for the PD game, there are many motivations for both sides to think about a better solution in this case because the collaboration between the customer and institution could be long-term. Here, the credential background of customers plays an important role in the institution to choose its strategy. For a customer with a positive credential background, the exact solution extracted by MOGCE is presented to the customer. The PD constraints have already guaranteed the safety of disclosed information to help customers. On the other side, for customers with credential background issues, there would be less motivation for the institution to trust them. However, based on the nature of PD, both sides can change their strategies over time. The institution uses fuzzy linguistic terms to help the customer avoid fraudulent activity from her side. If the customer modifies the application, the credit manager approves the credit background. Otherwise, the informed information is protected by fuzzy linguistic terms not to be abused by the untrustable customer. Table 4 represents the fuzzy linguistic terms and their corresponding fuzzy numbers to do so. For every single actionable feature, the proportion of increment and decrement to modify the application is calculated, and the corresponding linguistic term based on the proportion is reported to the customer. It should be mentioned that if the modified proportion

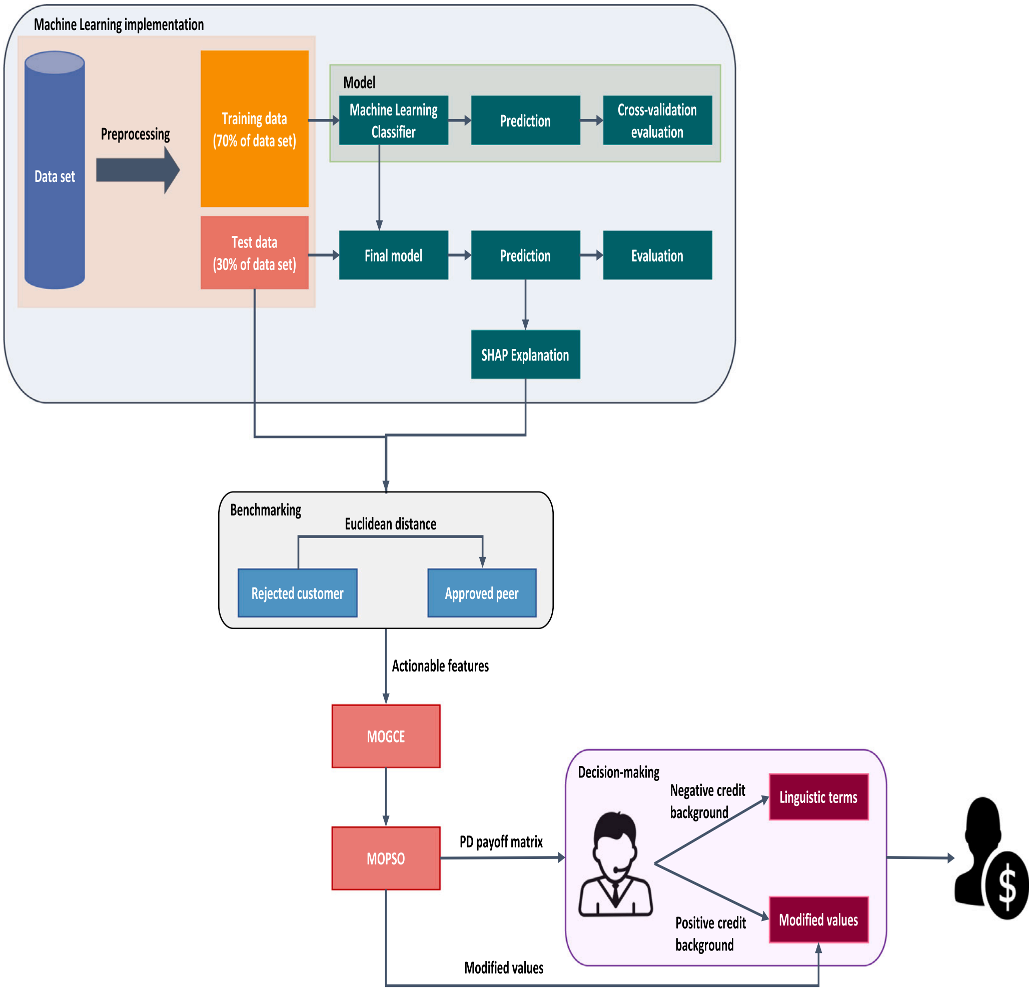

Fig. 1. The infographic of the proposed decision support framework.

exceeds the scale of the fuzzy number, it will behave the same way as the largest fuzzy number.

The infographic of the proposed decision support framework has been presented in Fig. 1.

5. Case study and result analysis

In this section, the proposed decision support framework is implemented on an open-access data set to elaborate its steps. In Section 5.1, the open-access data set is introduced, and the preprocessing phase is applied. In Section 5.2, different ML algorithms are exploited to classify customers into approved or rejected classes. In Section 5.3, the benchmarking process is implemented according to the XAI methods, and the decision-making process is conducted.

5.1. Data preprocessing

The Dream Housing Finance company data set [74] is used in this study to implement the proposed decision support framework. This company deals in all kinds of home loans and covers all urban, semi-urban, and rural areas. First, the customer applies for a home loan through an online form, and then the company validates the customer's eligibility for the loan. This data set has been used in many online projects and competitions to automate the loan eligibility process based on provided details by customers. The data set is literally small and consists of 614 instances, eleven features, and one label, which represents the status of the loan (approved or rejected). Table 5 demonstrates the data set information.

Table 4
Fuzzy linguistic terms and their corresponding fuzzy numbers.

Linguistic terms	Membership function
A little bit	(0, 0, 0.1)
Slightly	(0.1, 0.2, 0.3)
Moderately	(0.3, 0.4, 0.5)
Highly	(0.5, 0.6, 0.7)
Extremely	(0.7, 0.8, 0.9)
Absolutely	(0.9, 0.9, 1)

According to Table 5, some features have missing values and should tackle this problem. Regarding that data set is small it is not a smart approach to drop observations with missing values. Hence, data imputation approach is taken to fill missing values with appropriate values. Two different approaches are taken to impute categorical and numeric features: for categorical features, the most frequent value is used to impute missing values. For numeric features, K-Nearest Neighbors (KNN) algorithm [75] with K = 3 is used to handle missing values. After tackling with missing values, the behavior of the data set is studied. There is no missing value in the label, but a brief analysis of it shows that there are much more approved cases than rejected cases. There are 422 approved cases (68.72% entire instances) and 192 rejected cases (31.28% entire cases), showing the data set is imbalanced. In order to solve this problem, the Synthetic Minority Over-sampling Technique (SMOTE) [76] is applied to the training data set to have the same size for rejected customers with approved customers. The data set

Table 5
Data set description summary.

Feature	Missing values	Description	Type of data	Label
Loan_ID	0	Unique Loan ID	String	ID
Gender	13	Gender of applicant	Categorical	Male/Female
Married	3	Applicant marriage status	Categorical	Yes/No
Dependents	15	Number of dependents	Categorical	0, 1, 2, +3
Education	0	Applicant education	Categorical	Undergraduate/Graduate
Self_Employed	32	Occupation status	Categorical	Yes/No
ApplicantIncome	0	Applicant income	Numeric	[150, 81000]
CoapplicantIncome	0	Coapplicant income	Numeric	[0, 41667]
LoanAmount	22	Loan amount in thousands	Numeric	[9, 700]
Loan_Amount_Term	14	Term of the loan in months	Numeric	[12, 480]
Credit_History	50	Credit history meets guidelines	Categorical	0, 1
Property_Area	0	Location of accommodation	Categorical	Urban/Semi Urban/Rural
Loan_Status	0	Loan approval status	Categorical	Yes/No

Table 6
The predictive performance of the exploited ML models.

ML model	Accuracy	Precision	Recall	F1 score
LR	0.7189	0.7787	0.7917	0.7851
DT	0.7568	0.**7820**	0.8667	0.8221
MLP	0.7081	0.7578	0.8083	0.7823
RF	**0.7838**	0.7632	0.9667	0.8529
XGBoost	**0.7838**	0.7597	**0.9750**	**0.8540**

consists of seven categorical features, so one-hot encoding is applied to encode the data set in a proper format. Finally, the min–max data normalization is applied to the data set to be used for ML algorithms that need normalization. It should be mentioned that for ensemble models, it is feasible to use the original data without normalization.

5.2. Classifying customers into rejected and approved classes

In order to find the best-fitted ML algorithm for the data set, five different ML algorithms are examined to choose the best algorithm in terms of predictive performance: LR [77], DT [78], MLP [79], RF [80], and eXtreme Gradient Boosting (XGBoost) [81]. The data set is split into 70%–30% partitions representing training and test data sets, respectively. A 5-fold cross-validation approach embedding SMOTE oversampling to alleviate the imbalanced data set problem is used to validate ML algorithms on training data set. In order to obtain the best performance for all ML algorithms grid search algorithm is used to optimize hyperparameters.

After exploiting ML models, XGBoost obtained the best predictive performance among other ML models on the test data set (See Table 6). F1 score is considered the evaluation metric because the data set is imbalanced and reflects a better overview of the model's predictive performance than other metrics. The accuracy of RF and XGBoost are the same however, the F1 score for XGBoost is higher. The point is the very poor performance of MLP, which is not surprising from one side because the size of the data set is very small, and it cannot reflect its potential in data sets with small size because it is a data-hungry algorithm. Hence, the XGBoost is selected as the best model in terms of predictive performance, and the rest of the framework is continued based on its predictions.

Afterward, the SHAP method is implemented to extract the global feature importance for XGBoost. The result of SHAP in Fig. 2 represents the high importance of Credit_History in the model's prediction. This exciting finding overlaps real-world expectations from a customer's behavior and the institution's decision-making procedure. Hence, to have a successful application, the customer with a credit issue needs to reflect trustworthy behavior so that the institution approves their credit and then the application. The other interesting finding is the high impact of categorical features that are hard to modify.

5.3. Benchmarking, finding optimal solutions, and reporting the decision

In this step, a rejected customer is selected, and its approved peer is found using Euclidean distance. Then, SHAP is used to find the contribution of every single feature to the prediction of both customers. Then, actionable features for customers are used, and MOGCE is optimized to obtain the optimal solutions. In order to clarify the process, two instances are presented in this section. The first instance belongs to a customer with the credential background issue, and the second one for a customer with a positive credential background.

Instance 1: rejected customer index = 78, approved peer index = 110

In this instance, customer 78 is selected for target setting to find the optimal modified value to be approved. The approved peer of this customer is customer 110. The presented explanation of SHAP and each feature's contribution to the final prediction for both customers is presented in Table 7.

For customer 78, Credit_History is the main barrier in approving their application based on this observation that its impact on the prediction is more than other features. In this situation, it is up to the credit manager to approve the customer's credit background. To do so, based on actionable features, it is possible to modify their application. For customer 78, the most positively impactful numeric features for rejection based on Table 7 are Loan_Amount_Term, ApplicantIncome, and LoanAmount, respectively. Here, the strategy for customer 78 for target setting based on the application of customer 110 is increasing its Loan_Amount_Term up to 60 units, increasing ApplicantIncome up to 2149 units, and decreasing LoanAmount up to 44 units. It should be mentioned that the CoapplicantIncome is not modifiable because customer 78 is married, but customer 110 is single, so it does not make sense to modify this feature based on marital status. The mentioned target setting strategies are restricted to satisfying PD constraints in order to guarantee the main assumption: disclosing a safe amount of information without having bias based on the customer's credential background. Finally, MOGCE is optimized to obtain optimal modification value for customer 78. The MOGCE is optimized based on this assumption that Credit_History is triggered as positive to find the optimal amount of modification for numeric features. However, in practice, the credit manager approves it after applying modifications by the customer on the application. In this study, 30 independent runs are executed for MOPSO with 500 iterations to generate various populations with higher diversity. Also, population and repository size are 100, and w, c_1, and c_2 are set on 1, 0.2, and 0.1, respectively. Table 8 demonstrates the results of 30 independent runs for target setting with their counterfactual prediction. In this table, only solutions that satisfy PD constraints are collected and represented.

According to Table 8 the optimal solution is obtained in Batch=29, having the lowest cost function, which is the sum of all three objective functions for modifying actionable features. Hereby, Loan_Amount_Term, ApplicantIncome, and LoanAmount should be modified as 318.83,

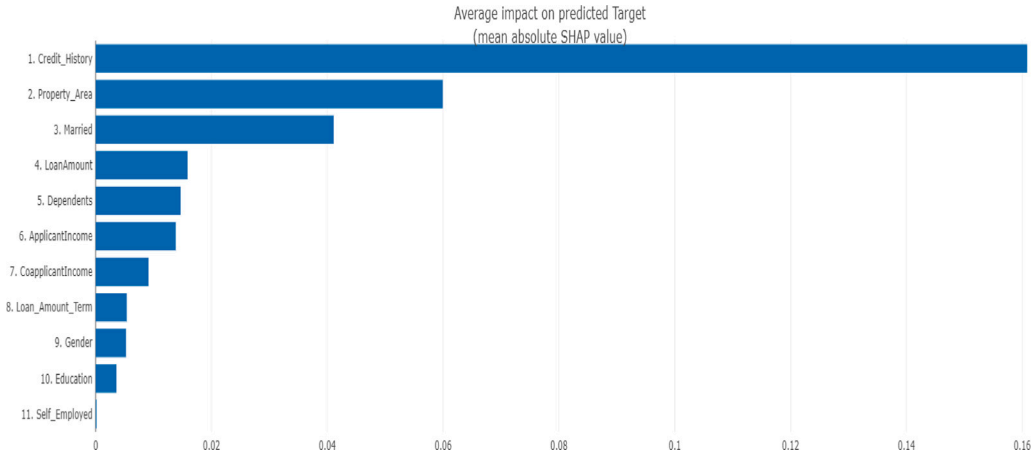

Fig. 2. Global feature importance of XGBoost extracted by SHAP.

Table 7
Feature importance of selected customers for target setting based on SHAP.

Features	Customer 78		Customer 110	
	Value	Impact	Value	Impact
Credit_History	0	+44.86%	1	+7.54%
Property_Area	1	-3.7%	2	-2.53%
Married	1	-1.37%	0	-8.36%
Loan_Amount_Term	300	+1.19%	360	+0.37%
Self_Employed	0	+1.02%	0	-0.27%
Dependents	3	-0.84%	0	-0.26%
CoapplicantIncome	4000	+0.68%	0	+0.46%
ApplicantIncome	3167	+0.28%	5316	0.0%
LoanAmount	180	+0.28%	136	+0.03%
Education	0	-0.24%	0	+0.59%
Gender	1	+0.22%	1	-0.39%
Prediction	Reject=84.27%		Approve=55.28%	

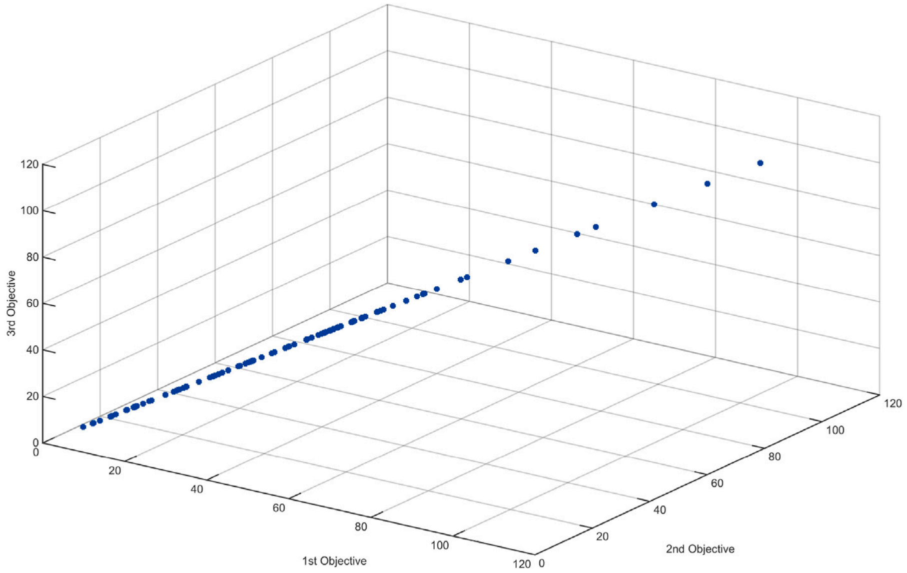

Fig. 3. The 3D Pareto frontier for optimal solution in Batch=29.

3564.53, and 168.75 for approving credit background. Fig. 3 represents the Pareto frontier for the optimal solution in Batch=29.

Regarding customer 78 having credential background issue, the response by the credit manager must be reported as fuzzy linguistic terms. To do so, the modification proportion is calculated based on the original values for actionable features and modified values obtained by MOGCE. The strategy has been demonstrated in Table 9. Thus, the decision by the credit manager is reported as increasing Loan_Amount_Term a little bit, increasing ApplicantIncome slightly, and decreasing LoanAmount a little bit. Then, it is the customer's responsibility to modify their application as much as possible to have a successful application. Moreover, the results of the counterfactual prediction demonstrate that

Table 8
Batch of optimal solutions for customer 78 obtained by MOGCE.

Batch	Cost function	Modified value			Pay-off matrix				Prediction
		Loan_Amount_Term	ApplicantIncome	LoanAmount	T	R	P	S	
1	464.232	352.46	3394.79	150.57	1	0.9610	0.1834	0.0613	Approve
2	416.839	303.23	4922.44	163.66	1	0.7892	0.1756	0.0613	Approve
3	336.974	319.80	3309.92	168.75	1	0.6587	0.1380	0.0613	Approve
5	440.897	346.61	3225.05	150.57	1	0.9113	0.1742	0.0613	Approve
6	416.679	335.89	3649.40	160.02	1	0.8887	0.1601	0.0613	Approve
7	413.045	324.68	4243.49	161.48	1	0.9625	0.1461	0.0613	Approve
8	386.536	335.89	3394.79	163.66	1	0.6948	0.1683	0.0613	Approve
9	404.829	351.00	4158.62	165.11	1	0.8971	0.1503	0.0613	Approve
10	373.334	349.05	3394.79	168.02	1	0.9562	0.1184	0.0613	Approve
11	469.864	347.10	3904.01	179.66	1	0.8904	0.1984	0.0613	Approve
12	403.251	362.21	4328.36	155.66	1	0.7976	0.1645	0.0613	Approve
13	376.067	343.20	4752.70	148.38	1	0.7251	0.1560	0.0613	Approve
14	401.473	360.75	5007.31	170.20	1	0.9204	0.1443	0.0613	Approve
15	498.163	359.29	5007.31	170.93	1	0.9819	0.2047	0.0613	Approve
17	349.518	305.66	3649.40	165.84	1	0.7272	0.1365	0.0613	Approve
18	391.967	330.04	3564.53	174.57	1	0.9120	0.1387	0.0613	Approve
19	364.354	350.03	5007.31	177.48	1	0.9757	0.1090	0.0613	Approve
20	370.727	326.63	4922.44	170.20	1	0.6516	0.1635	0.0613	Approve
21	436.393	332.48	3734.27	168.75	1	0.9741	0.1612	0.0613	Approve
22	375.574	345.15	4328.36	169.48	1	0.7481	0.1521	0.0613	Approve
23	411.132	347.59	3734.27	174.57	1	0.6443	0.1939	0.0613	Approve
24	345.073	320.29	3649.40	158.57	1	0.5645	0.1584	0.0613	Approve
25	350.227	329.55	4328.36	151.29	1	0.5776	0.1601	0.0613	Approve
26	376.080	349.05	3904.01	157.11	1	0.5308	0.1860	0.0613	Approve
27	371.997	310.54	4328.36	162.20	1	0.7013	0.1567	0.0613	Approve
28	422.851	350.03	3988.88	176.02	1	0.9516	0.1549	0.0613	Approve
29	308.684	318.83	3564.53	168.75	1	0.7162	0.1087	0.0613	Approve
30	324.018	326.14	4158.62	173.11	1	0.6741	0.1263	0.0613	Approve

Table 9
The modified actionable features and corresponding fuzzy linguistic terms.

	Loan_Amount_Term	ApplicantIncome	LoanAmount
Original value	300	3167	180
Modified value	318.83	3564.53	168.75
Modification proportion	0.0628	0.1256	0.0626
Response	A little bit	Slightly	A little bit

Credit_History has the highest impact on changing the prediction result. However, based on the strategic approach of this framework, approving Credit_History by the credit manager depends on the customer's sufficient modification of the application.

Instance 2: rejected customer index = 426, approved peer index = 361

In this example, the rejected customer and its approved peer are selected in the same way as explained before. Table 10 represents the original value of features for every customer and their impact on the final prediction. Both of the customers have a positive credit background; however, ApplicantIncome for customer 426 is not sufficient and has the highest impact on the application rejection. The negative value of Credit_History for customer 426 represents that her credit background positively contributed to approving her application, but the impact of other features, especially ApplicantIncome, is conspicuous. The next actionable feature is LoanAmount which seems the customer's request for a loan is less than the institution's minimum possible amount. Therefore, the customer needs to increase their loan amount as well. Finally, the strategy of customer 426 for target setting based on the customer 361 application is increasing ApplicantIncome up to 394 units and increasing LoanAmount up to 155 units.

Table 11 represents the optimal solutions for target setting in 30 independent runs. Again, it should be mentioned that only solutions that satisfy the PD constraint are represented in this table. Based on this table, only three solutions can change the prediction's outcome, and the remaining cannot have a counterfactual prediction to approve the application. The optimal solution has been obtained in Batch=14, which has the least cost function among other qualified solutions. In this solution, the modified values for ApplicantIncome and LoanAmount are 4916.67 and 150.15, respectively. An interesting finding for counterfactual prediction is that there should be a trade-off between modified values of ApplicantIncome and LoanAmount to change the prediction's result. For instance, in Batch=25, the sum of the modified values for both actionable features is greater than that in Batch=14. However, the counterfactual prediction is still rejected. This means that modifying values as much as possible does not guarantee a change in the default prediction's result. The modification should change the position of the instance on the decision boundary. The Pareto frontier for the optimal solution is demonstrated in Fig. 4

Finally, Table 12 represents the original value of the actionable features, their modified values, and modification proportion. Regarding the customer having a positive credential background, the credit manager can totally trust her and share the exact modified values with the customer.

The proposed explainable decision support framework for credit scoring represents a significant advancement, primarily benefiting DMs and rejected customers. Firstly, it assists DM in identifying suboptimal customers among those who have been rejected. This capability

Table 10
Feature importance of selected customers for target setting based on SHAP.

	Customer 426		Customer 361	
	Value	Impact	Value	Impact
ApplicantIncome	4606	+19.47%	5000	-0.87%
Property_Area	0	+11.18%	1	+10.14%
Married	0	+8.63%	1	+2.12%
Credit_History	1	-5.54%	1	+9.73%
LoanAmount	81	+5.08%	236	-1.92%
Dependents	1	+4.98%	2	+1.21%
Gender	0	-1.51%	1	-0.3%
Education	1	+0.95%	0	+0.2%
CoapplicantIncome	0	+0.64%	3667	0.0%
Loan_Amount_Term	360	-0.13%	360	+0.12%
Self_Employed	0	0.0%	0	0.0%
Prediction	Reject=78.44%		Approve=85.73%	

Table 11
Batch of optimal solutions for customer 426 obtained by MOGCE.

Batch	Cost function	Modified value		Pay-off matrix				Prediction
		ApplicantIncome	LoanAmount	T	R	P	S	
1	838.123	4750.00	229.89	1	0.9650	0.2444	0.1593	Reject
2	824.691	4916.67	232.77	1	0.8404	0.2665	0.1593	Reject
3	751.105	4750.00	206.19	1	0.6648	0.2663	0.1593	Reject
4	738.932	4666.67	209.06	1	0.7705	0.2341	0.1593	Reject
5	912.199	4916.67	234.20	1	0.8857	0.3058	0.1593	Reject
6	921.593	4750.00	216.96	1	0.9581	0.2938	0.1593	Reject
7	830.914	4916.67	224.15	1	0.7274	0.2970	0.1593	Reject
8	872.103	4750.00	221.27	1	0.9221	0.2741	0.1593	Reject
9	742.864	4916.67	199.00	1	0.7041	0.2521	0.1593	Reject
10	788.817	4666.67	178.17	1	0.8355	0.2470	0.1593	Reject
11	735.864	4750.00	170.26	1	0.7709	0.2321	0.1593	Approve
12	778.678	4916.67	232.05	1	0.7711	0.2567	0.1593	Reject
13	864.165	4916.67	187.51	1	0.8874	0.2778	0.1593	Approve
14	686.529	4916.67	150.15	1	0.7791	0.2019	0.1593	Approve
15	780.075	4666.67	153.74	1	0.8116	0.2477	0.1593	Reject
16	825.019	4916.67	194.69	1	0.8552	0.2632	0.1593	Reject
18	888.791	4833.33	224.15	1	0.9361	0.2803	0.1593	Reject
20	842.555	4916.67	219.12	1	0.7142	0.3069	0.1593	Reject
22	919.237	4750.00	229.18	1	0.9416	0.2965	0.1593	Reject
23	784.689	4583.33	205.47	1	0.8347	0.2449	0.1593	Reject
24	835.948	4666.67	186.07	1	0.8934	0.2602	0.1593	Reject
25	915.263	4916.67	233.49	1	0.8452	0.3172	0.1593	Reject
26	804.088	4916.67	210.50	1	0.7753	0.2702	0.1593	Reject
28	915.961	4833.33	209.78	1	0.9583	0.2906	0.1593	Reject
29	713.260	4666.67	178.89	1	0.8555	0.1990	0.1593	Reject

Table 12
The modified actionable features.

	ApplicantIncome	LoanAmount
Original value	4604	81
Modified value	4916.67	150.15
Modification proportion	0.0679	0.8538

addresses RQ1, contributing to the practical utility of the framework in financial and other institutions by shedding light on the deficiencies leading to rejection contributing to the understanding of reasons causing application rejection. Secondly, the framework enables DM to propose feasible modifications for rejected customers, offering strategic solutions that were previously unexplored in the literature using MOGCE. This aligns with the objectives of RQ2 and further underscores the practical benefits for institutions involved in credit scoring. Additionally, the framework encourages rejected customers to modify their applications, fostering a collaborative and informed approach to enhance their chances of success.

6. Conclusion and discussion

In this study, an explainable decision support framework was proposed for credit scoring with ML models, which help rejected customers modify their application to have a successful one. First, the XGBoost classifier was selected due to its superior predictive performance among other ML models on the data set in this research. Afterward, SHAP method was used to extract the importance of the features locally and globally to determine vital features for decision-making. According to

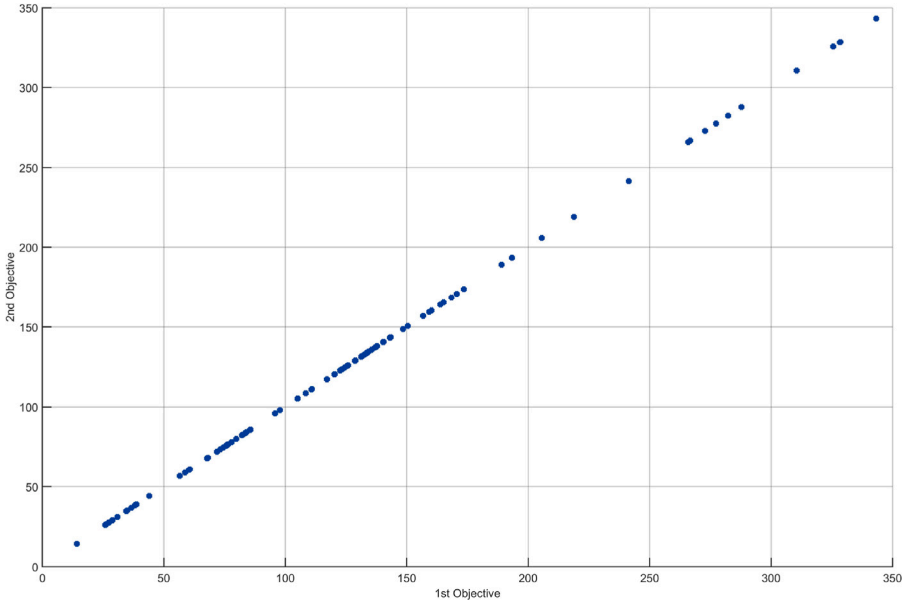

Fig. 4. The 2D Pareto frontier for optimal solution in Batch=14.

the global explanation by SHAP, Credit_History has the highest impact on the prediction's outcome. This finding supports the main expectation that Credit_History is the most important factor for a financial institution to provide financial aid for customers. To set targets for a rejected customer to modify their application, the benchmarking approach is followed. After finding the approved equivalent peer for the rejected customer, the local explanation by SHAP is used to set targets. Then, a MOGCE model is developed to find the optimal modified values for actionable features subject to satisfying PD's game constraints. Satisfying PD's constraints guarantees that the modified values are safe enough to report to the customers. In order to report the decision to the customers there are two scenarios. For customers with credit background issues, they are reported as fuzzy linguistic terms. However, for customers with a positive credit background, the exact decision can be reported. This is necessary because the institution needs to both manage relationships between customers as one of the main goals of CRM and have a strategic behavior to prevent fraudulent activities by customers. The proposed framework can successfully generate optimal solutions satisfying PD's constraints and making counterfactual predictions. For the first studied instance having a credit background issue, the modified value is found, and after modifying the application, the credit background can be approved by the credit manager. In the second instance, the customer has a positive credit background, but due to lack of other criteria, their application is rejected. In this case, the modified value is directly reported to the customer.

The proposed framework can successfully meet our two expectations: first, make explainable ML-based decision-making by finding the critical features that contributed to the model's prediction. Second, finding optimal solutions for customers to successfully modify their applications. The developed MOGCE model can successfully make counterfactual predictions based on the PD game's constraints, which formulates the strategic behavior of decision-making to disclose a safe amount of information to customers.

This study has some limitations that we will seek to address in future studies. First, we will focus on setting automatic parameters for MOPSO to adjust the parameters based on every single customer's important features. Second, post-hoc algorithms do not provide robust solutions, and we will try to develop an inherently interpretable framework to extract the impact of actionable features. Third, there should be a trade-off between generated optimal solutions and counterfactual prediction. We will seek how to modify actionable features to generate more solutions with counterfactual prediction. Fourth, in the interest of brevity, we have presented only two instances to illustrate the performance of the developed framework. In future endeavors, we plan to develop a software package that facilitates immediate access to necessary modifications and automatically generates comprehensive reports on decisions aligned with the credential background of customers. Fifth, we plan to enhance the robustness of our benchmarking technique by assessing the accuracy of identifying the best equivalent peer for a reference data point in future studies. Recognizing the limitation of evaluating a single data point, we acknowledge the necessity for a more comprehensive approach. Therefore, we will develop an improved benchmarking method that considers multiple data points, ensuring a thorough evaluation and selection of the best peer for the reference data point. Also, for future studies, we would like to evaluate the trust of experts in the proposed decision support framework based on a set of interviews and questionnaires.

CRediT authorship contribution statement

Mohsen Abbaspour Onari: Writing – original draft, Visualization, Project administration, Methodology, Conceptualization. **Mustafa Jahangoshai Rezaee:** Supervision, Methodology, Formal analysis, Conceptualization. **Morteza Saberi:** Validation, Supervision. **Marco S. Nobile:** Validation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] M. Ala'raj, M.F. Abbod, M. Majdalawieh, L. Jum'a, A deep learning model for behavioural credit scoring in banks, Neural Comput. Appl. 34 (8) (2022) 5839–5866.

[2] C.-F. Tsai, J.-W. Wu, Using neural network ensembles for bankruptcy prediction and credit scoring, Expert Syst. Appl. 34 (4) (2008) 2639–2649.

[3] C.-S. Ong, J.-J. Huang, G.-H. Tzeng, Building credit scoring models using genetic programming, Expert Syst. Appl. 29 (1) (2005) 41–47.

[4] S. Tsirtsis, M. Gomez Rodriguez, Decisions, counterfactual explanations and strategic behavior, Adv. Neural Inf. Process. Syst. 33 (2020) 16749–16760.

[5] J. Dyche, M.M. O'Brien, et al., The CRM handbook: A business guide to customer relationship management, Addison-Wesley Professional, 2002.

[6] A. Parvatiyar, J.N. Sheth, Customer relationship management: Emerging practice, process, and discipline., J. Econ. Soc. Res. 3 (2) (2001).

[7] V. Kumar, W. Reinartz, Customer relationship management, Springer, 2018.

[8] L. Ryals, A. Payne, Customer relationship management in financial services: towards information-enabled relationship marketing, J Strategic Market. 9 (1) (2001) 3–27.

[9] J. Peppard, Customer relationship management (CRM) in financial services, Eur. Manag. J. 18 (3) (2000) 312–327.

[10] G.L. Urban, The emerging era of customer advocacy, MIT Sloan Manag. Rev. 45 (2) (2004) 77.

[11] D. Sharma, J. Paul, S. Dhir, R. Taggar, Deciphering the impact of responsiveness on customer satisfaction, cross-buying behaviour, revisit intention and referral behaviour, Asia Pacific J. Market. Logist. 34 (10) (2022) 2052–2072.

[12] F. Buttle, S. Maklan, Customer relationship management: concepts and technologies, Routledge, 2019.

[13] I. Oino, Do disclosure and transparency affect bank's financial performance? Corpor. Governan. Int. J. Bus. Soc. 19 (6) (2019) 1344–1361.

[14] J.P. Mulki, F. Jaramillo, Ethical reputation and value received: customer perceptions, Int. J. Bank Market. 29 (5) (2011) 358–372.

[15] P.W. van Esterik-Plasmeijer, W.F. Van Raaij, Banking system trust, bank trust, and bank loyalty, Int. J. Bank Mark. 35 (1) (2017) 97–111.

[16] S. Wachter, B. Mittelstadt, L. Floridi, Why a right to explanation of automated decision-making does not exist in the general data protection regulation, Int. Data Privacy Law 7 (2) (2017) 76–99.

[17] N. Allenspach, Banking and transparency: is more information always better?, Swiss National Bank Working Papers 2009-11, 2009.

[18] M.J. Rezaee, M.A. Onari, M. Saberi, A data-driven decision support framework for DEA target setting: an explainable AI approach, Eng. Appl. Artif. Intell. 127 (2024) 107222.

[19] X. Dastile, T. Celik, M. Potsane, Statistical and machine learning models in credit scoring: A systematic literature survey, Appl. Soft Comput. 91 (2020) 106263.

[20] C.-F. Tsai, Feature selection in bankruptcy prediction, Knowl.-Based Syst. 22 (2) (2009) 120–127.

[21] D. Zhang, X. Zhou, S.C. Leung, J. Zheng, Vertical bagging decision trees model for credit scoring, Expert Syst. Appl. 37 (12) (2010) 7838–7843.

[22] F.-L. Chen, F.-C. Li, Combination of feature selection approaches with SVM in credit scoring, Expert Syst. Appl. 37 (7) (2010) 4902–4909.

[23] G. Wang, J. Hao, J. Ma, H. Jiang, A comparative assessment of ensemble learning for credit scoring, Expert Syst. Appl. 38 (1) (2011) 223–230.

[24] G. Wang, J. Ma, L. Huang, K. Xu, Two credit scoring models based on dual strategy ensemble trees, Knowl.-Based Syst. 26 (2012) 61–68.

[25] L.-J. Kao, C.-C. Chiu, F.-Y. Chiu, A Bayesian latent variable model with classification and regression tree approach for behavior and credit scoring, Knowl.-Based Syst. 36 (2012) 245–252.

[26] L. Han, L. Han, H. Zhao, Orthogonal support vector machine for credit scoring, Eng. Appl. Artif. Intell. 26 (2) (2013) 848–862.

[27] R. Florez-Lopez, J.M. Ramon-Jeronimo, Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment. A correlated-adjusted decision forest proposal, Expert Syst. Appl. 42 (13) (2015) 5737–5753.

[28] M. Ala'raj, M.F. Abbod, Classifiers consensus system approach for credit scoring, Knowl.-Based Syst. 104 (2016) 89–105.

[29] Y. Xia, C. Liu, B. Da, F. Xie, A novel heterogeneous ensemble credit scoring model based on bstacking approach, Expert Syst. Appl. 93 (2018) 182–199.

[30] M. Saberi, M.S. Mirtalaie, F.K. Hussain, A. Azadeh, O.K. Hussain, B. Ashjari, A granular computing-based approach to credit scoring modeling, Neurocomputing 122 (2013) 100–115.

[31] M. Herasymovych, K. Märka, O. Lukason, Using reinforcement learning to optimize the acceptance threshold of a credit scoring model, Appl. Soft Comput. 84 (2019) 105697.

[32] P. Pławiak, M. Abdar, U.R. Acharya, Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring, Appl. Soft Comput. 84 (2019) 105740.

[33] Y. Xia, J. Zhao, L. He, Y. Li, M. Niu, A novel tree-based dynamic heterogeneous ensemble method for credit scoring, Expert Syst. Appl. 159 (2020) 113615.

[34] D. Tripathi, D.R. Edla, V. Kuppili, A. Bablani, Evolutionary extreme learning machine with novel activation function for credit scoring, Eng. Appl. Artif. Intell. 96 (2020) 103980.

[35] F. Shen, X. Zhao, G. Kou, Three-stage reject inference learning framework for credit scoring using unsupervised transfer learning and three-way decision theory, Decis. Support Syst. 137 (2020) 113366.

[36] C.-F. Wu, S.-C. Huang, C.-C. Chiou, Y.-M. Wang, A predictive intelligence system of credit scoring based on deep multiple kernel learning, Appl. Soft Comput. 111 (2021) 107668.

[37] J.W. Lee, W.K. Lee, S.Y. Sohn, Graph convolutional network-based credit default prediction utilizing three types of virtual distances among borrowers, Expert Syst. Appl. 168 (2021) 114411.

[38] S. Maldonado, J. López, C. Vairetti, Time-weighted fuzzy support vector machines for classification in changing environments, Inform. Sci. 559 (2021) 97–110.

[39] V.B. Djeundje, J. Crook, R. Calabrese, M. Hamid, Enhancing credit scoring with alternative data, Expert Syst. Appl. 163 (2021) 113766.

[40] M.B. Gorzałczany, F. Rudziński, A multi-objective genetic optimization for fast, fuzzy rule-based credit classification with balanced accuracy and interpretability, Appl. Soft Comput. 40 (2016) 206–220.

[41] Y. Xia, C. Liu, Y. Li, N. Liu, A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring, Expert Syst. Appl. 78 (2017) 225–241.

[42] K. Lee, H. Lee, H. Lee, Y. Yoon, E. Lee, W. Rhee, Assuring explainability on demand response targeting via credit scoring, Energy 161 (2018) 670–679.

[43] Q. Lan, X. Xu, H. Ma, G. Li, Multivariable data imputation for the analysis of incomplete credit data, Expert Syst. Appl. 141 (2020) 112926.

[44] M.Y. Tezerjan, A.S. Samghabadi, A. Memariani, ARF: A hybrid model for credit scoring in complex systems, Expert Syst. Appl. 185 (2021) 115634.

[45] V. Moscato, A. Picariello, G. Sperlí, A benchmark of machine learning approaches for credit score prediction, Expert Syst. Appl. 165 (2021) 113986.

[46] P.Z. Lappas, A.N. Yannacopoulos, A machine learning approach combining expert knowledge with genetic algorithms in feature selection for credit risk assessment, Appl. Soft Comput. 107 (2021) 107391.

[47] G. Visani, E. Bagli, F. Chesani, A. Poluzzi, D. Capuzzo, Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models, J. Oper. Res. Soc. 73 (1) (2022) 91–101.

[48] X. Dastile, T. Celik, H. Vandierendonck, Model-agnostic counterfactual explanations in credit scoring, IEEE Access (2022).

[49] A.C. Bueff, M. Cytryński, R. Calabrese, M. Jones, J. Roberts, J. Moore, I. Brown, Machine learning interpretability for a stress scenario generation in credit scoring based on counterfactuals, Expert Syst. Appl. 202 (2022) 117271.

[50] E. Dumitrescu, S. Hué, C. Hurlin, S. Tokpavi, Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects, European J. Oper. Res. 297 (3) (2022) 1178–1192.

[51] M. Bücker, G. Szepannek, A. Gosiewska, P. Biecek, Transparency, auditability, and explainability of machine learning models in credit scoring, J. Oper. Res. Soc. 73 (1) (2022) 70–90.

[52] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).

[53] M.J. Rezaee, Using Shapley value in multi-objective data envelopment analysis: power plants evaluation with multiple frontiers, Int. J. Electr. Power Energy Syst. 69 (2015) 141–149.

[54] C. Moreira, Y.-L. Chou, M. Velmurugan, C. Ouyang, R. Sindhgatta, P. Bruza, LINDA-BN: An interpretable probabilistic approach for demystifying black-box predictive models, Decis. Support Syst. 150 (2021) 113561.

[55] B. Davazdahemami, H.M. Zolbanin, D. Delen, An explanatory machine learning framework for studying pandemics: The case of COVID-19 emergency department readmissions, Decis. Support Syst. 161 (2022) 113730.

[56] N. Gozzi, L. Malandri, F. Mercorio, A. Pedrocchi, XAI for myo-controlled prosthesis: Explaining EMG data for hand gesture classification, Knowl.-Based Syst. 240 (2022) 108053.

[57] M.T. Keane, E.M. Kenny, E. Delaney, B. Smyth, If only we had better counterfactual explanations: Five key deficits to rectify in the evaluation of counterfactual xai techniques, 2021, arXiv preprint arXiv:2103.01035.

[58] S. Wachter, B. Mittelstadt, C. Russell, Counterfactual explanations without opening the black box: Automated decisions and the GDPR, Harv. JL Tech. 31 (2017) 841.

[59] S. Verma, J. Dickerson, K. Hines, Counterfactual explanations for machine learning: A review, 2020, arXiv preprint arXiv:2010.10596.

[60] S. Dandl, C. Molnar, M. Binder, B. Bischl, Multi-objective counterfactual explanations, in: International Conference on Parallel Problem Solving from Nature, Springer, 2020, pp. 448–469.

[61] M. Abbaspour Onari, M. Jahangoshai Rezaee, A fuzzy cognitive map based on Nash bargaining game for supplier selection problem: a case study on auto parts industry, Oper. Res. (2020) 1–39.

[62] M. Abbaspour Onari, M. Jahangoshai Rezaee, Implementing bargaining game-based fuzzy cognitive map and mixed-motive games for group decisions in the healthcare supplier selection, Artif. Intell. Rev. (2023) 1–34.

[63] D. Robinson, D. Goforth, The topology of the 2x2 games: a new periodic table, vol. 3, Psychology Press, 2005.

[64] R. Axelrod, W.D. Hamilton, The evolution of cooperation, Science 211 (4489) (1981) 1390–1396.

[65] M. Nowak, K. Sigmund, A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game, Nature 364 (6432) (1993) 56–58.

[66] E. Ostrom, Governing the commons: The evolution of institutions for collective action, Cambridge University Press, 1990.

[67] R. Axelrod, Effective choice in the prisoner's dilemma, J. Conflict Resol. 24 (1) (1980) 3–25.

[68] C.A.C. Coello, G.T. Pulido, M.S. Lechuga, Handling multiple objectives with particle swarm optimization, IEEE Trans. Evolut. Comput. 8 (3) (2004) 256–279.

[69] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95-International Conference on Neural Networks, Vol. 4, IEEE, 1995, pp. 1942–1948.

[70] M. Abbaspour Onari, S. Yousefi, M. Jahangoshai Rezaee, Risk assessment in discrete production processes considering uncertainty and reliability: Z-number multi-stage fuzzy cognitive map with fuzzy learning algorithm, Artif. Intell. Rev. 54 (2) (2021) 1349–1383.

[71] K. Khalili-Damghani, A.-R. Abtahi, M. Tavana, A new multi-objective particle swarm optimization method for solving reliability redundancy allocation problems, Reliab. Eng. Syst. Saf. 111 (2013) 58–75.

[72] J. Meza, H. Espitia, C. Montenegro, E. Giménez, R. González-Crespo, MOVPSO: Vortex multi-objective particle swarm optimization, Appl. Soft Comput. 52 (2017) 1042–1057.

[73] M. Jabbari, S. Sheikh, M. Rabiee, A. Oztekin, A collaborative decision support system for multi-criteria automatic clustering, Decis. Support Syst. 153 (2022) 113671.

[74] The Dream Housing data set, https://datahack.analyticsvidhya.com/contest/practice-problem-loan-prediction-iii/.

[75] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1) (1967) 21–27.

[76] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.

[77] J. Berkson, Application of the logistic function to bio-assay, J. Am. Statist. Assoc. 39 (227) (1944) 357–365.

[78] J.R. Quinlan, Induction of decision trees, Mach. Learn. 1 (1) (1986) 81–106.

[79] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain., Psychol. Rev. 65 (6) (1958) 386.

[80] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[81] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.