Profit-Driven Fusion Framework based on Bagging and Boosting Classifiers for Potential Purchaser Prediction

Abstract

Accurately identifying potential purchasers (PPers) is pivotal for enhancing an enterprise's core competitiveness in a competitive market. Although existing research focused on individual classifiers for PPer prediction, there is a notable gap in the integration of the bagging and boosting algorithms, resulting in suboptimal performance. This study introduces a novel fusion framework for profit-oriented PPer prediction that combines the strengths of the bagging (specifically, random forest, RF) and boosting (utilizing categorical boosting, CatBoost) algorithms. CatBoost replaces the original base learner in RF, leveraging the advantages of both classifiers to reduce the variance and bias. To optimize the proposed RF-CatBoost-based fusion framework for profit maximization, we employ a grid search to fine-tune hyperparameters. This approach aligns with enterprises' profit-driven objectives. The experimental results, statistical tests, and Bayesian A/B tests collectively demonstrate that the proposed framework outperforms all comparative classifiers, yielding the highest profits. Furthermore, an interpretability analysis reveals the significant factors influencing the prediction results, providing valuable insights for decision makers in identifying PPers within customer groups.

Keywords: Decision support systems; Bagging and boosting classifiers; Fusion framework; Potential purchaser prediction

1. Introduction

In the era of internet development and escalating market saturation, enterprises must devise customer-centric marketing strategies (Gengler and Popkowski Leszczyc, 1997; Hossain and Rahman, 2022a, 2022b; C.-H. Liu et al., 2022; Pashchenko et al., 2022; Thomas et al., 2022; Trinh and Wright, 2022; Wu and Padgett, 2004), particularly in challenging domains, such as subscription services across various industries. The prevalent focus of current customer-centric strategies in subscription services revolves around curbing customer churn in sectors like telecommunications (Al-Weshah et al., 2019; Coussement et al., 2017), finance (Berloco et al., 2022; Papouskova and Hajek, 2019; Rozo et al., 2022; Zhou et al., 2021), tourism (Liu et al., 2023; Lu et al., 2020), energy (Moeyersoms and Martens, 2015), and others (De Caigny et al., 2020; Fader and Hardie, 2009; Gamage et al., 2021; Jamal and Bucklin, 1987; Jiang et al., 2024, 2023; Liu et al., 2024; Risselada et al., 2010). Surprisingly, scant attention has been directed toward identifying potential purchasers (PPers) within the customer base. Although acquiring new clients may be more expensive than retaining existing ones (Maldonado et al., 2021), the precise identification of new customers remains lucrative. In contrast to existing customers, the PPer cohort is limited and harbors significant development and value-added potential. Accurately predicting PPers with heightened consumption desires and deploying targeted marketing strategies can minimize marketing expenses and maximize marketing efficacy (Chen et al., 2016). This dual impact saves costs, increases profit margins, and heightens market competitiveness (Lei and Zhu, 2012). Consequently, it is imperative for enterprises to establish robust and accurate PPer prediction technologies to gain a competitive edge under intense market competition.

Researchers employed various machine learning techniques to predict PPers, showcasing diverse and innovative approaches in the field. Martínez et al. (2020) utilize gradient tree boosting to predict PPers in non-contractual settings, revealing superior statistical indicators compared with least absolute shrinkage and selection operator (LASSO) and extreme learning machines. Tanuwijaya, Alamsyah, and Ariyanti (2021) apply K-Means to partition a dataset into three clusters and employ Categorical Boost (CatBoost) to predict PPers within each cluster. Chaudhuri et al. (2021) develop a deep neural network for PPer prediction on e-commerce platforms and conduct interpretable analyses of the input features. Chen et al. (2021) use an enhanced factorization machine to create composite features and cost-sensitive extreme gradient boosting (XGBoost) for PPer prediction. Chou et al. (2022) employ the buy-till die model to extract additional features and subsequently use LASSO or neural networks to enhance the classification accuracy with the newly generated and original features. Xu, Dang, and Wang (2022) introduce a weighted recency, focus, and sentiment model to quantify the probability of a specific individual being a PPer.

Existing PPer prediction research focuses solely on statistical indicators of classifier performance, overlooking the paramount objective of profit maximization. Moreover, there is room for improvement in the current prediction performance of PPer models. To address the unique characteristics of PPer prediction, this study redefines the profit indicator, shifts from customer churn prediction, and introduces a fusion framework that integrates bagging and boosting classifiers. We conduct profit-centric PPer predictions using four business subscription datasets from diverse scenarios to validate the profitability of our proposed fusion framework.

Bagging and boosting are the most commonly employed ensemble algorithms. Bagging involves the construction of multiple bootstrapped training sets from a given dataset, with each base learner trained on these sets to produce a collection of base learners (Breiman, 1996). The classification for each instance is determined through a vote or averaging process across all base learners, resulting in the most voted or averaged classification. The expected misclassification probability of bagging classifiers exhibits equal bias to that of a single bootstrap replicate, yet the variance is reduced by $1/N$ (Fumera et al., 2005). By contrast, boosting assigns weights to training samples and dynamically adjusts these weights based on the classifier performance of the training examples. The weight of a misclassified example increases, enabling adaptive modification of the training set distribution, with each learner's performance influencing the next (Schapire, 1990). Predictive values are estimated through a weighted vote of each learner's predictions, with weights proportional to the learner's accuracy in the training set. By emphasizing misclassified examples, boosting effectively mitigates the bias in newly generated base learners, enhancing performance with high-bias and low-variance data (Kotsianti and Kanellopoulos, 2007).

From the preceding analysis, bagging and boosting diminish variance and mitigate bias, respectively. Drawing inspiration from these distinctive features, we introduce an innovative RF-CatBoost-based fusion framework (RCFF). This framework integrates boosting into the bagging construction, aiming to concurrently reduce both variance and bias.

The main contributions of our study are as follows.

(i) A novel RCFF is introduced for PPer prediction by combining the strengths of a robust bagging classifier, specifically random forest (RF), with an effective boosting approach (CatBoost). In this framework, CatBoosts serve as base learners integrated into the original bootstrap-aggregation structure of the RF, resulting in a unique bootstrap-boosting aggregation framework.

(ii) Based on the characteristics of the PPer prediction, we reformulate the profit indicator from the customer churn prediction. The proposed RCFF is fine-tuned using a grid search to optimize the hyperparameters, aligning with the objective of maximizing profit. In contrast to prior PPer prediction studies that concentrated on prediction accuracy, our approach integrates the goal of profit maximization into model training, enhancing the profitability of business campaigns.

The accurate prediction of PPers is crucial for enabling enterprises to deliver personalized content in precision marketing and formulate effective pricing strategies. Utilizing the proposed RCFF, this study predicts customer purchase intentions when engaged with the provided products or services. Subsequently, the enterprise can offer targeted discounts or gifts based on the prediction outcomes, enhancing the likelihood of successful customer conversions. This study adopts a profit-driven approach to determine the optimal hyperparameter combinations for the RCFF and benchmark classifiers. We evaluate the prediction performance based on profit indicators. Employing Shapley's additive explanations (SHAP) for interpretative analysis, we quantify the influence intensity and direction of each customer feature on profit-driven predictions. This enables enterprises to identify PPers who yield maximum profits in precision marketing scenarios geared toward profit maximization.

In light of the evolving landscape of marketing analytics, where the emphasis has shifted to acquiring potential customers, our research investigates the dynamics of customer purchases and the potential impact of various strategies. In this study, we hypothesize the following:

1. We assume that the predictions are binary; that is, whether customers will make a purchase, rather than on the specific quantity or frequency of purchases. By simplifying purchase behavior into a binary outcome, we can concentrate on analyzing the key factors that influence purchase decisions, such as customer characteristics. This binary prediction assumption allows researchers to focus more on determining whether customers will engage in purchasing activities, laying the foundation for developing more effective marketing strategies.

2. Several factors influence customer purchase decisions. Despite the limited number of variables used in our study, machine learning models can effectively capture the underlying consumer trends and behavioral patterns when dealing with these variables. Meanwhile, companies take major measures for each potential consumer, recognized by predictions, to maximize market share, increase sales, and improve customer satisfaction.

3. Drawing from prior research on profit-driven customer behavior prediction (Höppner et

al., 2020; Jiang et al., 2023; Kozodoi et al., 2019; Lessmann et al., 2021; Liu et al., 2023; Maldonado et al., 2020, 2015), we assume that all orders generate equal profits. This implies disregarding the specific scale, type, or other variations among orders; that is, each order holds an equivalent economic value in terms of profit contribution. This assumption serves to streamline the model, allowing us to focus on the factors influencing customers' purchase behavior without delving into the intricacies of profit differences between orders.

The remainder of this paper is structured as follows. Section 2 outlines the relevant studies on PPer prediction. Section 3 details the methodology used in the study. The experimental setup and analysis of the results are presented in Section 4, and Section 5 concludes the paper with a discussion.

2. Literature Review

PPer prediction can be framed as a binary classification task that specifically assesses whether a client will make a purchase of goods or services. Adopting a binary classification framework is advantageous because of its simplicity in computation and ability to facilitate clear objective goal analysis (S. Chen et al., 2021). PPer prediction typically relies on statistical and machine learning technologies. Statistical approaches, such as logistic regression (LR) (Li, 2019; Zhang, 2021), naive Bayes classifier (NBC) (Das, 2016; Palaniappan et al., 2017), and LASSO (Chou et al., 2022), model the linear relationships between input variables and outcomes in advance. While statistical models offer simplicity and high interpretability, they struggle to effectively capture the complex correlations between customer features and target variables under intricate circumstances.

Hence, machine learning techniques are extensively applied for PPer prediction. Various methods are used for this purpose, including support vector machines (SVM) (Chen et al., 2011; Lei and Zhu, 2012), decision trees (DT) (Avrizal et al., 2018; Coussement et al., 2015; Li et al., 2019; Nilashi et al., 2021; Palaniappan et al., 2017), and neural networks (Basuki, 2016;

Chaudhuri et al., 2021; S.-S. Chen et al., 2021; Choudhury and Nur, 2019; Hu et al., 2016; Kim et al., 2021; Li et al., 2023; Z. Liu et al., 2022; Wen et al., 2018), bagging ensemble classifiers (Esmeli et al., 2022; Jaiswal et al., 2020; Palaniappan et al., 2017; Zhang, 2021), boosting ensemble classifiers (Martínez et al., 2020; Tanuwijaya et al., 2021), stacking ensemble classifiers (Shah et al., 2022; Yeo et al., 2020), and fusion frameworks (Cui et al., 2018). Ensemble classifiers combine multiple individual classifiers to make predictions, thereby reducing the risk of overfitting and increasing generalization performance. By aggregating the decisions of multiple models, ensembles can capture diverse patterns in the data and mitigate the impact of noise or outliers, leading to more robust predictions. Previous research indicates that ensemble classifiers exhibit superior predictive capabilities compared to individual classifiers such as SVM, DT, and neural networks. Notably, fusion frameworks offer flexibility in selecting diverse classifiers, contributing to enhanced accuracy beyond that of ensemble classifiers (Chen et al., 2022). **Table 1** provides a summary of the relevant studies on PPer prediction.

Table 1
Summary of research on PPer prediction.

Reference	Dataset Number	Better Models or Constructed Model	Validation Form	Hyperparameter Selection Method	Hyperparameter Selection Objective	Interpretational Analysis	Statistical Test
Chen et al. (2011)	1	SVM	No	No	No	No	Paired T-test
Lei and Zhu (2012)	1	Improved SVM	No	No	No	No	No
Das (2016)	1	NBC	Cross-validation	No	No	No	No
Coussement et al. (2015)	4	Chi-square automatic interaction detection, classification and regression tree (CART), and neural network	Cross-validation	No	No	No	Non-parametric test of (DeLong et al., 1988)
Basuki (2016)	1	Learning vector quantization neural network	No	Grid Search	Accuracy	No	No
Hu et al. (2016)	1	Heterogeneous social network	Hold-out validation	No	No	No	No
Palaniappan et al. (2017)	1	NBC, random forest (RF), and DT	Cross-validation	No	No	No	No
Avrizal et al. (2018)	1	DT	No	No	No	No	No
Yeo et al. (2020)	1	Predictability-aware aggregation model	No	No	No	No	No
Wen et al. (2018)	1	Spatial, temporal, payment, and product category in probability graphic model	No	Grid Search	Not mentioned	No	No
Cui et al. (2018)	1	Fusion framework based on convolutional neural network and long short-term memory	No	Grid Search	Accuracy	No	No
Choudhury and Nur (2019)	1	Multi-layer perceptron (MLP)	No	No	No	No	No
Li et al. (2019)	1	DT	No	No	No	No	No
YAN (2019)	1	Factor analysis, discriminant analysis	No	No	No	No	No
Li (2019)	1	LR	No	No	No	No	No
Jaiswal et al. (2020)	1	RF	No	No	No	No	No
Martínez et al. (2020)	1	Gradient tree boosting	Cross-validation	No	No	No	No
Tanuwijaya et al. (2021)	1	K-Means, and CatBoost	No	No	No	Feature importance	No
Zhang (2021)	1	LR, and RF	No	No	No	No	No
Rahim et al. (2021)	1	DT	No	No	No	No	No
Chaudhuri et al. (2021)	1	Deep neural network	Cross-validation	Grid Search	Accuracy	Feature importance	Cochran's Q test, and McNemar's test
S. Chen et al. (2021)	1	Improved factorization machine, and XGBoost	No	No	No	No	No
Shah et al. (2022)	1	An ensemble classifier consists of bidirectional representation for transformers, bidirectional long short-term memory, and SVM	No	No	No	No	No
Xu et al. (2022)	1	Weighted recency, focus, and sentiment model	No	No	No	No	No
Esmeli et al. (2022)	4	RF	Cross-validation	No	No	Feature importance	No
Z. Liu et al. (2022)	1	Selective graph convolutional network	No	No	No	No	No
Chou et al. (2022)	1	Buy till you die model, and LASSO	Repeat evaluation based on n-out-of-n bootstrap	Grid Search	Not mentioned	No	t-test
Proposed RCFF	4	RF-CatBoost-based fusion framework	Cross-validation	Grid Search	EMPC	SHAP	Friedman test, Holm post-hoc test, Bayesian A/B test

After reviewing the relevant literature, several conclusions emerge. 1) Most studies employ individual classifiers, such as LR, GNBC, SVM, DT, and neural networks, with only a few leveraging ensemble classifiers for PPer prediction, suggesting room for performance enhancement. The proposed RCFF is constructed based on two ensemble mechanisms including bagging and boosting. Meanwhile, the PPer prediction performance of the proposed RCFF in our study is compared with ensemble models. 2) Given an enterprise's profit-driven objectives, most studies use statistical indicators as criteria for selecting hyperparameters to maximize these indicators, which may not align with the broader operational goals of enterprises. In our study, we reformulate the profit indicator, and optimize the hyperparameters of the proposed RCFF based on grid search with the objective of maximizing profit. 3) Many studies focus narrowly on a single application scenario, evaluating the developed classifiers on a singular dataset, and lacking comprehensiveness in the evaluation process. In this study, datasets under four different industry scenarios are used to verify the PPer prediction performance of the proposed RCFF and benchmark models. 4) Fair comparisons of prediction performance between classifiers often lacking. Effective validation and grid searches for optimal hyperparameter selection are frequently omitted. In this study, grid search is used to determine the hyperparameters of the proposed RCFF and benchmark models, and evaluate the PPer prediction performance based on 10-fold cross-validation. 5) Interpretative analyses of the prediction results and discussions of the features of PPers are generally absent in most studies. The interpretability of classifiers holds crucial value in operational decisions. In this study, SHAP values are used to interpret and analyze the proposed RCFF, to provide policy recommendations for decision makers.

3. Methodology

Bagging classifiers excel in reducing prediction variance, while Boosting classifiers excel in reducing prediction bias. To harness their individual strengths, we opt for a fusion framework that combines bagging and boosting classifiers. The conventional bagging classifier generates new training sets through random sampling (Breiman, 1996). Building on this, RF introduces a random selection of sample features, further diminishing the variance (Breiman, 2001). As an evolved version of the traditional bagging classifier, RF surpasses XGBoost and a light-gradient-boosting machine (LightGBM). CatBoost represents the latest iteration that enhances the gradient-boosting decision tree methodology (Prokhorenkova et al., 2018). CatBoost handles category features in PPer datasets by constructing combined features based on the category features. It introduces ordered boosting to address the prediction shift and bolsters robustness by reducing the gradient bias. Given its notable performance in predicting PPers (Esmeli et al., 2022; Jaiswal et al., 2020; Palaniappan et al., 2017; Tanuwijaya et al., 2021; Zhang, 2021), we propose an RCFF. This section provides a detailed overview of the framework, outlines the integration of RF and CatBoost, and introduces three performance metrics.

3.1. The RF-CatBoost-based Fusion Framework

RF, which is a classical tree-based ensemble algorithm proposed by Breiman (2001), integrates bagging technology into DT generation by employing an aggregation mechanism to determine the final classification results. RF is widely applied in machine learning tasks and is favored for handling high-dimensional data, offering low computational costs and strong parallel computing capabilities (Xia et al., 2020). Furthermore, RF maintains interpretability and facilitates the exploration of influential variables in prediction outcomes. CatBoost, introduced by Prokhorenkova et al. (2018), is an advanced gradient-boosting decision tree (GBDT)-based model. Prokhorenkova et al. (2018) established CatBoost's superiority over other GBDT-based methods, attributing it to two algorithmic advancements: ordered boosting and special consideration of categorical features. Inherently, RF reduces variance, whereas CatBoost is more effective in reducing bias (Breiman, 1996; Kotsianti and Kanellopoulos, 2007). We integrate CatBoost into the RF framework to synergize the strengths of these algorithms and construct a PPer classifier with minimal variance and bias.

The proposed RCFF comprises a collection of independent CatBoosts. During training, bootstrapping is employed to generate multiple subsets from the original training set, and then *m* CatBoosts are individually constructed. Each CatBoost provides a unit vote or prediction probability for the label of each instance. The final label is determined through a majority vote of the *m* CatBoosts, and the final probability is calculated by averaging the prediction probabilities of each CatBoost classifier. The construction of the RCFF involves three stages: partitioning the sub-training sets, building *m* CatBoosts, and aggregating for the final classification. The detailed process is outlined below and depicted in **Fig. 1**.

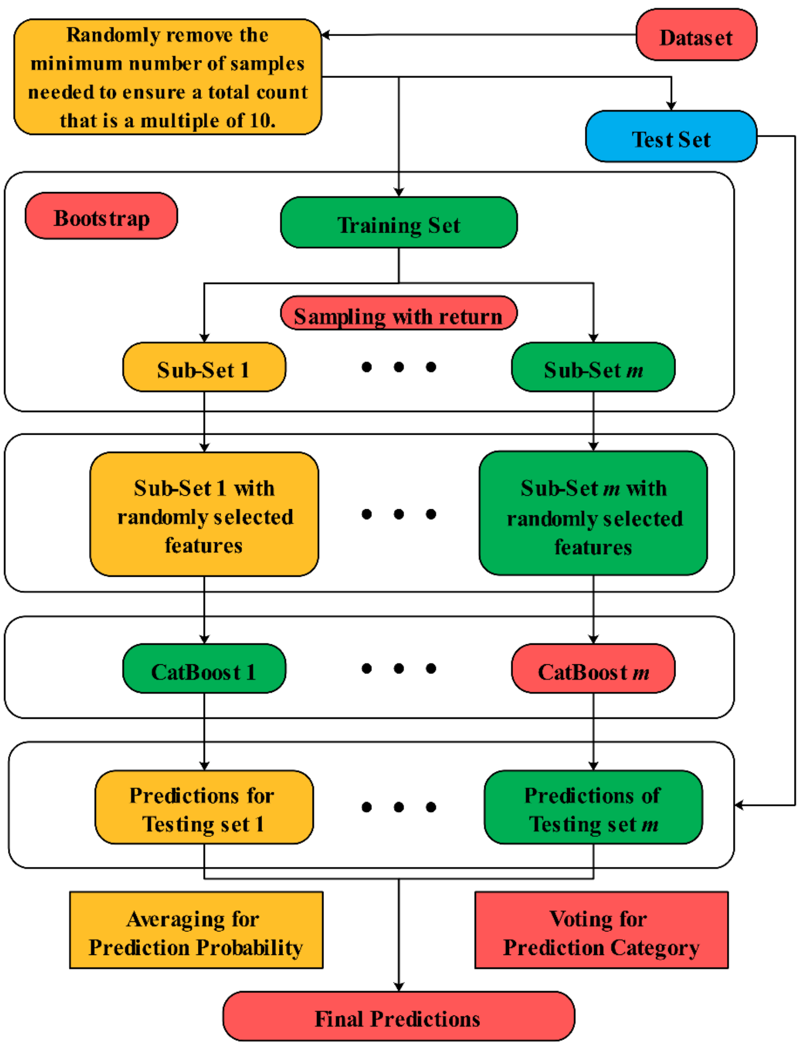

Fig.1. Detailed RCFF process.

Stage I: Partition of sub-training sets

To address the potential correlation issues between different CatBoosts trained on the same complete training set $\boldsymbol{\Omega} = \left\{ \left(\boldsymbol{x}_k , \boldsymbol{y}_k \right) \right\}_{k=1}^{n}$ (with n instances), bootstrapping is employed to create m unrelated training sets $\boldsymbol{\Omega}_j = \left\{ \left(\boldsymbol{x}_k , \boldsymbol{y}_k \right) \right\}_{k=1}^{n} \left(j = 1,2,...,m \right)$. The bootstrap method involves drawing a new training set using uniform sampling with the replacement of the original set. Instances may appear repeatedly in different training subsets $\boldsymbol{\Omega}_j = \left\{ \left(\boldsymbol{x}_k , \boldsymbol{y}_k \right) \right\}_{k=1}^{n} \left(j = 1,2,...,m \right)$, with a fraction $\left(1 - 1/e\right)\left(\approx 63.2\%\right)$ of unique instances, and the remainder being duplicates (Aslam, Popa, and Rivest, 2007). Partitioning sub-training sets using bootstrap technology enhances prediction performance by reducing variance without increasing bias.

Stage II: Construction of m CatBoosts

Multiple CatBoosts are constructed based on newly generated subsets $\boldsymbol{\Omega}_j = \left\{ \left(\boldsymbol{x}_k , \boldsymbol{y}_k \right) \right\}_{k=1}^{n} \left(j = 1,2,...,m \right)$, where $\boldsymbol{x}_k = \left(x_k^1 , x_k^2 , \cdots , x_k^w \right)$ represents a random vector with w features, and $\boldsymbol{y}_k \in \left\{ 0, 1 \right\}$ is a binary variable. $\left(\boldsymbol{x}_k , \boldsymbol{y}_k \right)$ is independently and identically distributed based on an unknown distribution $\Xi\left(\cdot,\cdot\right)$. The objective of the learning process is to train a function $F : R^m \rightarrow R$ to realize the expected loss $L\left(F\right) = \mathbf{E}\, L\left(\boldsymbol{y}, F\left(\boldsymbol{x}\right)\right)$ minimum, where $L\left(\cdot,\cdot\right)$ denotes a smooth loss function and $\left(\boldsymbol{x}, \boldsymbol{y}\right)$ represents a test sample sampled from the training subset $\boldsymbol{\Omega}_j$.

In the iteration process, gradient boosting is used to construct a sequence of approximations $F^t : R^m \rightarrow R \; t = 0,1,...$ in a greedy way. Specifically, F^t can be estimated based on the F^{t-1} additively; that is, $F^t = F^{t-1} + \alpha h^t$. Here, α denotes a step size and $h^t : R^m \rightarrow R$ is a base classifier selected from a set of functions H to minimize the expected loss; that is,

$$h^t = \underset{h \in H}{arg\,min}\, L\left(F^{t-1} + h\right) = \underset{h \in H}{arg\,min}\, \mathbf{E}\, L\left(\boldsymbol{y}, F^{t-1}\left(\boldsymbol{x}\right) + h\left(\boldsymbol{x}\right)\right) \qquad (1)$$

Eq. (1) can be solved via the Newton method using a second order approximation of $L\left(F^{t-1}+h\right)$ at F^{t-1} or via a (negative) gradient step. The gradient step h^t is determined by using $h^t\left(\boldsymbol{x}\right)$ approximates $-g^t\left(\boldsymbol{x},\boldsymbol{y}\right)$, where $g^t\left(\boldsymbol{x},\boldsymbol{y}\right)=\dfrac{\partial L\left(\boldsymbol{y},F^{t-1}\left(\boldsymbol{x}\right)\right)}{\partial F^{t-1}\left(\boldsymbol{x}\right)}$. Generally, the least-squares approximation is employed, and Eq. (1) can be further expressed as

$$h^t=\underset{h\in H}{arg\,min}\,\mathbf{E}\left(-g^t\left(\boldsymbol{x},\boldsymbol{y}\right)-h\left(\boldsymbol{x}\right)\right)^2 \tag{2}$$

There are two advancements in CatBoost: the use of ordered target statistics technology to handle categorical features and the introduction of ordered boosting to overcome gradient bias. For ordered target statistics technology, CatBoost first conducts a random permutation of all instances, and then computes the average label value of the instance with the same category value placed before the given one in the permutation ($\boldsymbol{y}_{\delta_s}$). Given a permutation $\boldsymbol{\delta}=\left(\delta_1,\delta_2,\cdots,\delta_n\right)$, the original *p*-th permutated observation with categorical feature *q* is replaced by $\boldsymbol{x}_{\delta_p,q}$, and the mathematical expression of $\boldsymbol{x}_{\delta_p,q}$ is

$$x_{\delta_p,q}=\frac{\sum_{s=1}^{p-1}\left[x_{\delta_s,q}=x_{\delta_p,q}\right]\boldsymbol{y}_{\delta_s}+\boldsymbol{\xi}P}{\sum_{s=1}^{p-1}\left[x_{\delta_s,q}=x_{\delta_p,q}\right]+\boldsymbol{\xi}} \tag{3}$$

where P and $\boldsymbol{\xi}$ denote the prior value and weight of the prior value, respectively. The former is conducive to decreasing the noise associated with the low-frequency category.

To avoid the gradient bias issue existing in traditional boosting techniques, ordered boosting is introduced into CatBoost to calculate leaf values when determining the tree structure. Specifically, based on a random permutation $\boldsymbol{\delta}=\left(\delta_1,\delta_2,\cdots,\delta_n\right)$, for the *i*-th instance in the permutated dataset x_i, a separate classifier M_i is trained based on the gradients of examples before x_i (i.e., $x_1,x_2,\cdots x_{i-1}$). Subsequently, the gradient of x_i is estimated via M_i to score the tree. Hence, CatBoost effectively overcomes the overfitting of other GBDTs.

Stage III: Aggregating for the final classification

The final predictor is built using multiple CatBoosts constructed on the newly generated sub-training set $\mathbf{\Omega}_j = \{(\boldsymbol{x}_k, \boldsymbol{y}_k)\}_{k=1}^{n} (j = 1,2,...,m)$, which is determined by a majority vote of m CatBoosts, and the final prediction probability is calculated by averaging all prediction probabilities of m CatBoosts.

3.2. Performance Evaluation

We evaluate performance using three indicators: the area under the receiver operating characteristic curve (AUC) and top decile lift (TDL) as statistically oriented indicators, and the expected maximum profit criterion (EMPC) as a profit-oriented indicator. We use these metrics to assess the accuracy and profitability of the proposed RCFF relative to other classifiers. The **Table 2** provides the definitions of these evaluation metrics based on the confusion matrix.

Predicting whether a customer will be a purchaser involves binary label prediction, denoted by **Y**={0 (non-PPer), 1 (PPer)}. Each predictor is trained to produce a probabilistic outcome p within the range [0, 1]. Using a preset threshold t, the outcome p is compared, and if $p \le t$ is used, the instance is classified as a non-PPer (y=1); otherwise, it is classified as a non-PPer (y=0). A confusion matrix is then established to quantify the correct and incorrect classifications. Table 2 defines prior probabilities for class 0 (non-PPer) and class 1 (PPer) as π_0 and π_1, respectively. The probability density functions of the probabilistic outcomes for non-PPers and PPers are $f_0(p)$ and $f_1(p)$, respectively, and their cumulative distribution functions are $F_0(p)$ and $F_1(p)$, respectively. M represents the total number of instances evaluated for each dataset.

Table 2
Confusion matrix definition.

Predicted Class	Actual Class	
	Non-PPer	PPer
Non-PPer	$\pi_0\ F_0(t)\cdot M$	$\pi_1\ F_1(t)\cdot M$
PPer	$\pi_0(1\text{-}F_0(t))\cdot M$	$\pi_1(1\text{-}F_1(t))\cdot M$

Based on the confusion matrix, the AUC indicator can be first indicated as

$$\boldsymbol{AUC} = \int_{-\infty}^{+\infty} F_0\left(p\right) f_1\left(p\right) dp \tag{4}$$

TDL, a statistical indicator, compares the proportion of PPers in the entire dataset with the proportion in the top decile, comprising clients with the highest predicted probabilities of becoming PPers according to the predictor (De Caigny et al., 2018). The TDL indicator is

$$\boldsymbol{TDL}\left(10\%\right) = \frac{\pi_{10\%}}{\pi} \tag{5}$$

where $\pi_{10\%}$ is the proportion of PPers in the top 10% of ordered posterior probabilities of becoming a PPer and π represents the proportion of PPers in the entire dataset. TDL is a crucial performance indicator from a managerial standpoint, as it focuses on PPers with heightened consumption intentions, increasing the likelihood of conversion to formal consumers, and thereby contributing to higher profits with lower costs.

However, the classical indicator AUC incorporates implicit assumptions about misclassification costs, and using this indicator in a business environment can result in suboptimal profit outcomes. The EMPC is a classical profit-driven evaluation indicator proposed by Verbraken, Verbeke, and Baesens (2013). We adopt the EMPC because profit maximization is a fundamental goal for businesses, including those in the retail and consumer services sectors. Considering that solving a specific business problem requires us to incorporate characteristics inherent in the cost and benefit structure, the EMPC, is derived for customer churn. Specifically, the average classification profit under a preset threshold t is

$$\begin{aligned} P_C\left(t;\gamma,CLV,\delta,\phi\right) = CLV\cdot\left[\gamma\cdot\left(1-\delta\right)-\phi\right]\cdot\pi_1\cdot\left(1-F_1\left(t\right)\right) \\ -CLV\cdot\left(\delta+\phi\right)\cdot\pi_0\cdot\left(1-F_0\left(t\right)\right) \end{aligned} \tag{6}$$

In customer churn prediction, CLV denotes the customer life value, γ is the probability of successfully retaining a churner, and $\delta = d/CLV$ and $\phi = f/CLV$ represent the cost of retaining and contacting customers, respectively. In PPer prediction, *CLV* is replaced by the average profit on purchase orders (*AP*); γ is the probability of successfully attracting a PPer; and *d* and *f* denote the costs of the incentive for a PPer to convert and contact a customer, respectively. Therefore, we can redefine the average classification profit as

$$\begin{aligned} P_C\left(t;\gamma,AP,\delta,\phi\right) = AP\cdot\left[\gamma\cdot\left(1-\delta\right)-\phi\right]\cdot\pi_1\cdot\left(1-F_1\left(t\right)\right) \\ -AP\cdot\left(\delta+\phi\right)\cdot\pi_0\cdot\left(1-F_0\left(t\right)\right) \end{aligned} \tag{7}$$

Notably, the probability of successfully attracting a PPer varies, and as per Verbraken, Verbeke, and Baesens (2013), γ follows a beta distribution. Consequently, we can express the EMPC is

$$\boldsymbol{EMPC} = \int_\gamma P_C\left(T\left(\gamma\right);\gamma,AP,\delta,\phi\right)\cdot h\left(\gamma\right)\mathrm{d}\gamma \tag{8}$$

$$\boldsymbol{EMPC} = \int_\gamma \left\{\begin{array}{l} AP\cdot\left[\gamma\cdot\left(1-\delta\right)-\phi\right]\cdot\pi_1\cdot\left(1-F_1\left(t\right)\right) \\ -AP\cdot\left(\delta+\phi\right)\cdot\pi_0\cdot\left(1-F_0\left(t\right)\right) \end{array}\right\}\cdot h\left(\gamma\right)\mathrm{d}\gamma \tag{9}$$

$$\boldsymbol{EMPC} = \int_\gamma \left\{\begin{array}{l} \left\{\begin{array}{l} AP\cdot\left[\gamma\cdot\left(1-\delta\right)-\phi\right] \\ \cdot\pi_1\cdot\left(1-F_1\left(t\right)\right) \end{array}\right\} \\ -\left\{\begin{array}{l} AP\cdot\left(\delta+\phi\right) \\ \cdot\pi_0\cdot\left(1-F_0\left(t\right)\right) \end{array}\right\} \end{array}\right\}\cdot\frac{\gamma^{\alpha-1}\left(1-\gamma\right)^{\beta-1}}{\int_0^1 u^{\alpha-1}\left(1-u\right)^{\beta-1}du}\mathrm{d}\gamma \tag{10}$$

where $h\left(\gamma\right) = h\left(\gamma;\alpha,\beta\right) = \dfrac{\gamma^{\alpha-1}\left(1-\gamma\right)^{\beta-1}}{\int_0^1 u^{\alpha-1}\left(1-u\right)^{\beta-1}du}$ is the probability density function of the beta distribution and $T\left(\gamma\right)$ represents the optimal threshold, expressed mathematically as in Eqs. (11) -(12).

$$T = \underset{t\in[0,\,1]}{argmax}\left\{P_C\left(t;\gamma,CLV,\delta,\phi\right)\right\} \tag{11}$$

$$T = \underset{t \in [0,1]}{argmax} \begin{Bmatrix} AP \cdot \left[\gamma \cdot (1-\delta) - \phi \right] \cdot \pi_1 \cdot \left(1 - F_1(t)\right) \\ -AP \cdot (\delta + \phi) \cdot \pi_0 \cdot \left(1 - F_0(t)\right) \end{Bmatrix} \tag{12}$$

The EMPC is not only directly used to evaluate model performance, but can also be used as a profit measure in model construction (Höppner et al., 2020; Stripling et al., 2018). In addition to applying the EMPC to the customer churn problem, a profit-driven performance measure would help solve other business problems, such as consumer credit scoring (Verbraken et al., 2014). In this study, the EMPC is optimized to attract potential consumers. The hyperparameter settings for EMPC are adjusted accordingly. When a customer is predicted to be a PPer, the associated costs are realized at a specific juncture. The company initiates a sequence of actions for contacting leads, thereby incurring additional costs. Companies routinely make concerted efforts to establish contact with PPers to encourage them to purchase their products. Contact costs arise because the company needs to communicate with PPers through various channels, such as phone calls and email. This process requires the allocation of resources such as staffing, technology, and time; hence, contact costs materialize when a customer is predicted to be a PPer. Retention costs materialize arise because the company aims to ensure the conversion of PPers into actual purchasers. To achieve this objective, incentive measures such as special discounts and personalized services may be necessary, constituting additional costs that contribute to retention. Thus, behind the predictions of potential consumers, the timing and reasons for cost realization are integral concepts closely tied to actual business operations. Drawing insights from the literature on customer churn prediction and customer cancellation prediction (Höppner et al., 2020; Liu et al., 2023; Maldonado et al., 2020; Stripling et al., 2018; Verbraken et al., 2013), we maintain the average profit on purchase orders (*AP*) for PPer prediction at €200, with the cost of contacting a customer set at €1. Additionally, the literature suggests that the cost of acquiring a new customer is approximately five to six times that of retaining a would-be churner (Ganesh et al., 2000; Van den Poel and Larivière, 2004). Consequently, we set the cost of acquiring a new customer in the PPer prediction (d) to €50, in line with the cost of retaining a would-be churner (previously set at €10) in prior research on customer churn prediction. Retaining a potential churner involves customer intentions, whereas converting PPers into subscribers means translating intentions into practical action. Given the varying difficulty of the two operations, the beta distribution that characterizes the probability of would-be churners accepting retention is inappropriate for PPers accepting invitations. Recognizing that PPers are more inclined to subscribe to a service, and that the probability of accepting invitations is significantly higher than that of retaining would-be churners, we assume that the probability of PPers accepting invitations follows a symmetric beta distribution. Specifically, parameter γ follows a beta distribution with parameter values α=14 and β=6.

4. Experiment Setup and Results

This section details the experimental setup and provides a comprehensive analysis of the results, including the studied dataset, parameter configuration, comparative outcomes between the proposed RCFF and other classifiers, interpretability analysis, and Bayesian A/B analysis.

4.1. Dataset Description

We analyze four distinct PPer prediction scenarios: time deposits, health insurance, 5G packages, and credit cards. We chose datasets from diverse fields to validate the applicability of the proposed RCFF. Datasets 1, 2, and 4 are sourced from Kaggle (https://www.kaggle.com), and Dataset 3 is obtained from Heywhale (https://www.heywhale.com). **Table 3** presents an overview of the data. When selecting datasets from the four distinct industries, we carefully considered broad applicability and in-depth insights. Several factors explain the rationale for choosing these domains.

- Credit card services, time deposits, 5G packages and health insurance play pivotal and widespread roles in the socioeconomic landscape. These industries directly impact individuals' daily lives and financial decisions, adding practical relevance to the study of customer behavior in these critical sectors.

- The selected industries span different stages of the consumer life cycle. For instance, credit card services involve day-to-day transactions, time deposits relate to savings and investments, 5G packages address communication needs, and health insurance focuses on healthcare management. This comprehensive approach allows researchers to holistically understand consumer purchasing decisions across various lifecycle stages.
- Choosing datasets from multiple industries facilitates cross-industry comparisons and reveals common trends and unique patterns. This approach provides a broader understanding that extends beyond the confines of specific industrial contexts.

These factors collectively drove the selection of datasets from these specific industries, ensuring that the research design comprehensively investigates customer ordering decision behaviors across diverse domains. The goal is to provide robust guidance for future business and marketing strategies.

We conduct data preprocessing from two perspectives: data features and data size. To address the data features, we first removed the ID feature from each dataset, and the categorical features underwent one-hot encoding. This resulted in 38, 80, 48, and 52 for Datasets 1–4, respectively. Regarding data size processing, instances were randomly removed from the original datasets. This is necessary for the application of 10-fold cross-validation to assess the predictor performance while maintaining integer samples for each cross-validation fold. Consequently, the final sample for Datasets 1-4 are 31,640, 50,880, 140,000, and 245,720, respectively, all integer multiples of 10. The training-to-testing ratio is set at 9:1, and the testing data points used to validate the training results. Additionally, the training set is normalized to evaluate the prediction outcomes. Although tree-based classifiers do not theoretically require normalization, the feature values for both the training and testing sets are normalized based on the training data for a fair comparison. **Table 4** lists the label distributions of the test sets used for cross-validation

Table 3

Sample data summary.

Detailed Information [1]	Dataset 1 [2]	Dataset 2 [3]	Dataset 3 [4]	Dataset 4 [5]
Service	Time deposit	Health insurance	5G package	Credit card
Initial Sample Number	31647	50882	140000	245725
Initial Feature Number	17	12	43	9
Final Sample Number	31640	50880	140000	245720
Final Feature Number	38	80	48	52
Label Distribution	10.72%	24.00%	20.00%	23.72%

Table 4

Label distribution of test sets

Folds	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Fold 1	10.05%	23.90%	19.89%	23.78%
Fold 2	10.75%	24.17%	20.39%	23.83%
Fold 3	11.09%	23.98%	19.71%	23.61%
Fold 4	10.84%	23.49%	20.12%	23.93%
Fold 5	10.40%	24.33%	20.32%	23.18%
Fold 6	10.24%	23.80%	19.97%	23.75%
Fold 7	10.49%	24.55%	20.25%	24.16%
Fold 8	10.37%	24.69%	19.80%	23.77%
Fold 9	10.97%	23.43%	19.74%	23.72%
Fold 10	12.04%	23.62%	19.81%	23.48%
Ave (Std)	10.72% (0.54%)	24.00% (0.41%)	20.00% (0.24%)	23.72% (0.25%)

4.2. Parameter Settings

Eleven benchmark classifiers, comprising eleven individual classifiers: CART (Breiman et al., 1984), CatBoost (Prokhorenkova et al., 2018), convolutional neural network (CNN), convolutional neural network with dropout layer and normalization layer (CNN-DN), Gaussian naive Bayes classifier (GNBC) (Domingos and Pazzani, 1997), K-nearest neighbor (KNN) (Keramati et al., 2014), LASSO (Tibshirani, 1996), LightGBM (Ke et al., 2017), multi-layer perceptron (MLP) (Sundarkumar and Ravi, 2015), RF (Breiman, 2001), and XGBoost (Chen and Guestrin, 2016). We compare two fusion frameworks (RF-LightGBM-based and RF-XGBoost-based, or RLFF and RXFF, respectively) to the proposed RCFF. We include various comparison algorithms to validate the efficacy of the proposed framework. **Table 5** provides an overview of the parameter settings for both the benchmark classifiers and the proposed RCFF.

1 The variable description of datasets can be found in the following links.

[2] https://www.kaggle.com/datasets/jinxzed/av-hacklive-guided-hackathon

[3] https://www.kaggle.com/datasets/sureshmecad/health-insurance-lead-prediction

[4] https://www.heywhale.com/mw/dataset/6017a810c966020016eaf5e1

[5] https://www.kaggle.com/datasets/shelvigarg/credit-card-buyers

Table 5
Candidate parameter setting of all classifiers.

Classifier	Parameters	Candidate Values
CART	Maximum proportion for randomly selected features	[0.01,0.25,0.5,0.75,1]
	Maximum depth	[3,5,7,9]
	Minimum number of samples required for node splitting	[2,25,50,75,100]
	Minimum number of samples contained in the nodes	[1,25,50,75,100]
CatBoost	Number of DTs	100
	Learning rate	[0.01,0.25,0.5,0.75,1]
	Depth of DTs	[3,5,7,9]
	Minimum number of samples contained in the nodes of DTs	[1,25,50,75,100]
CNN	Epochs	100
	Optimizer	adam
	Batch size	[128,256,512,1024,2048]
	Filter size in first Conv1D	2
	Filter size in second Conv1D	1
	Number of filters	Number of features
	Size of hidden layer	[[1, 1], [25, 25], [50, 50], [75, 75], [100, 100]]
CNN-DN	Epochs	100
	Optimizer	adam
	Batch size	[128,256,512,1024,2048]
	Dropout size	[0.01,0.25,0.5,0.75,0.99]
	Filter size in first Conv1D	2
	Filter size in second Conv1D	1
	Number of filters	Number of features
	Size of hidden layer	[[1, 1], [25, 25], [50, 50], [75, 75], [100, 100]]
GNBC	-	-
KNN	Number of neighboring samples considered	[1,25,50,75,100]
	Leaf size	[1,25,50,75,100]
LASSO	Penalty coefficient for solving multi collinearity of input features	[0.0001,0.001,0.01, 0.25,0.5,0.75,1]
LightGBM	Number of DTs	100
	Nodes number of DTs	[2,25,50,75,100]
	Learning rate	[0.01,0.25,0.5,0.75,1]
	Maximum depth of DTs	[3,5,7,9]
	Minimum value of the sum of sample weights in each node of DTs	[0.01,0.25,0.5,0.75,1]
MLP	Maximum iterations	100
	Activation function	relu
	Optimizer	adam
	Batch size	[128,256,512,1024,2048]
	Initial learning rate	[0.01,0.25,0.5,0.75,1]
	Size of hidden layer	[1,25,50,75,100]
RF	Number of DTs	100
	bootstrap	True
	Minimum number of samples required for node splitting of DTs	[2,25,50,75,100]
	Minimum number of samples contained in the nodes of DTs	[1,25,50,75,100]
	Maximum depth of DTs	[3,5,7,9]
XGBoost	Number of DTs	100
	Learning rate	[0.01,0.25,0.5,0.75,1]
	Maximum depth of DTs	[3,5,7,9]
RLFF	Number of LightGBMs	100
	bootstrap	True
	Number of DTs in LightGBMs	[2,25,50,75,100]
	Nodes number of DTs in LightGBMs	[2,25,50,75,100]
	Learning rate in LightGBMs	[0.01,0.25,0.5,0.75,1]
	Maximum depth of DTs in LightGBMs	[3,5,7,9]
	Minimum value of the sum of sample weights in each node of DTs in LightGBMs	[0.01,0.25,0.5,0.75,1]
RXFF	Number of XGBoosts	100
	bootstrap	True
	Number of DTs in XGBoosts	[2,25,50,75,100]
	Learning rate in XGBoosts	[0.01,0.25,0.5,0.75,1]
	Maximum depth of DTs in XGBoosts	[3,5,7,9]

RCFF	Number of CatBoosts	100
	bootstrap	True
	Number of DTs in CatBoosts	[2,25,50,75,100]
	Learning rate in CatBoosts	[0.01,0.25,0.5,0.75,1]
	Depth of DTs in CatBoosts	[3,5,7,9]
	Minimum number of samples contained in the nodes of DTs in CatBoosts	[1,25,50,75,100]

All the candidate parameters are selected through a trial-and-error process based on the training set. Using a grid search, the optimal parameter values are selected from a set of candidate values. The search objective is to maximize the profit metric (EMPC) value, which can be shown as:

$$\begin{cases} \{\boldsymbol{Hypers}\} = arg \max_{\{\boldsymbol{Hypers}\}} EMPC \\ \qquad = arg \max_{\{\boldsymbol{Hypers}\}} \int_{\gamma} P_C\left(T(\gamma);\gamma, AP, \delta, \phi\right) \cdot h(\gamma)\,\mathrm{d}\gamma \\ s.t.\ h(\gamma) = h(\gamma;\alpha,\beta) = \dfrac{\gamma^{\alpha-1}(1-\gamma)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du} \\ AP = € \ 200, d = € \ 50, f = € \ 1 \\ \delta = d/AP, \varphi = f/AP, \alpha = 14, \beta = 6 \end{cases} \tag{13}$$

This approach minimizes the impact of varying parameter settings on the prediction performance, ensuring a fair comparison among classifiers of the same type.

4.3. Empirical Results

Table 6 presents the average results of the 10-fold cross-validation across the four prediction scenarios, encompassing three evaluation indicators. Corresponding standard deviations are provided in parentheses. The AUC, EMPC, and TDL (10%) are all positive indicators, where a larger value indicates superior prediction performance. Meanwhile, we listed the values of classical statistical evaluation indicators (F1-score, Precision, and Recall) based on classifiers in the **Table A2** of **Appendix**. As our focus is on assessing the business capability of the proposed RCFF in the context of potential buyer prediction, our evaluation centers around its performance against key business indicators in real-world scenarios. The statistical evaluation indicators mentioned above are provided solely for reference purposes. Our study selects the hyperparameters for the proposed frameworks, guided by EMPC, to realize profit-oriented PPer prediction. Two significant conclusions are drawn: (1) Under the profit-oriented training paradigm, the proposed fusion frameworks demonstrate adeptness in harmonizing classification performance across diverse sample categories, thereby augmenting enterprise profitability. (2) Although the conventional metrics applied to evaluate the proposed fusion frameworks may not surpass those of benchmark models, this does not infer the inferiority of the proposed frameworks within the traditional indicators-driven training paradigm.

Furthermore, we conduct a statistical test on the prediction performance of the classifiers (De Bock and De Caigny, 2021). The procedure of our statistical test is as follows. The average ranks across the 10-fold cross-validation results for the four datasets are computed, where a higher average rank indicates superior performance. Subsequently, the Friedman test is applied based on the average ranks. If the Friedman test rejects the null hypothesis, then we find a significant difference in the prediction performance among at least two classifiers. Finally, the Holm post-hoc test is employed to assess the adjusted statistical significance between the control classifier and each benchmark classifier, effectively mitigating family wise errors.

Assessing the performance evaluation metrics raises two crucial questions:

- Can the proposed RCFF outperform all benchmark classifiers in terms of profitability?
- Does the fusion mechanism, incorporating both bagging and boosting classifiers, enhance the profitability of PPer prediction?

On the first question, the proposed RCFF demonstrated remarkable profit gains, surpassing all competitive classifiers, as evidenced by its consistently high EMPC values across the four datasets. As shown in **Table 6**, the average EMPC values for the proposed RCFF are € 4.6354, € 2.4564, € 11.5845, and € 13.2431. Notably, the AUC and TDL (10%) values of this framework may not always rank the highest, indicating that the most accurate classifier does not necessarily translate into the most profitable one.

Subsequently, we conduct a statistical test of prediction performance against benchmark classifiers using the proposed RCFF as a control classifier. The Friedman statistic, yielding 206.1475 (*p*-value: 0.0000) based on the AUC ranks, 172.5585 (*p*-value: 0.0000) based on the EMPC ranks, and 91.1866 (*p*-value: 0.0000) based on the TDL (10%) ranks, underscores the significant differences in prediction ability among the classifiers. **Table 7** lists the average ranking of the proposed RCFF and benchmark classifiers, along with the adjusted *p*-values from the Holm post-hoc test.

Table 6
Comparison of accuracy and profitability for each dataset.

Dataset	Classifier	AUC	EMPC	TDL (10%)
Dataset 1	**CART**	0.9045 (0.0066)	3.9374 (0.4488)	5.2891 (0.2275)
	CatBoost	0.9262 (0.0046)	4.5152 (0.6010)	5.5576 (0.1668)
	CNN	0.9214 (0.0034)	4.3632 (0.4989)	5.4447 (0.1928)
	CNN-DN	0.9197 (0.0043)	4.2644 (0.5042)	5.4288 (0.1446)
	GNBC	0.8078 (0.0134)	2.0243 (0.3016)	4.0487 (0.1669)
	KNN	0.8412 (0.0082)	2.5681 (0.3502)	4.5635 (0.1978)
	LASSO	0.9098 (0.0044)	4.1041 (0.5396)	5.3432 (0.2381)
	LightGBM	0.9241 (0.0047)	4.4308 (0.5465)	5.4834 (0.2035)
	MLP	0.7062 (0.2059)	2.1311 (2.1632)	3.1605 (2.2103)
	RF	0.9202 (0.0046)	4.3725 (0.5750)	5.4524 (0.2351)
	XGBoost	0.9267 (0.0047)	4.5119 (0.5984)	5.5409 (0.2035)
	RLFF	0.9269 (0.0044)	4.4454 (0.5593)	5.4912 (0.2025)
	RXFF	0.9276 (0.0046)	4.5579 (0.5942)	**5.6005** (0.1859)
	RCFF	**0.9283** (0.0048)	**4.6354** (0.6132)	5.5948 (0.1916)
Dataset 2	**CART**	0.6131 (0.0082)	1.5391 (0.3071)	1.7231 (0.0565)
	CatBoost	0.6690 (0.0054)	2.2630 (0.3413)	1.8067 (0.0605)
	CNN	0.6253 (0.0087)	1.2059 (0.2445)	1.5796 (0.0887)
	CNN-DN	0.6257 (0.0085)	1.2055 (0.2454)	1.5642 (0.1014)
	GNBC	0.6027 (0.0091)	0.8022 (0.2340)	1.3950 (0.0854)
	KNN	0.6417 (0.0071)	1.4504 (0.3353)	1.6382 (0.0709)
	LASSO	0.6260 (0.0085)	1.2321 (0.2611)	1.5714 (0.0752)
	LightGBM	0.6436 (0.0106)	1.7847 (0.2399)	1.7498 (0.0367)
	MLP	0.6438 (0.0495)	1.8217 (0.7414)	1.6727 (0.2805)
	RF	0.6536 (0.0072)	1.9836 (0.3062)	1.7722 (0.0792)
	XGBoost	0.6508 (0.0084)	1.9214 (0.2053)	1.7536 (0.0597)
	RLFF	0.6686 (0.0054)	2.1855 (0.2514)	1.7967 (0.0793)
	RXFF	0.6650 (0.0092)	2.2585 (0.3038)	1.8018 (0.0956)
	RCFF	**0.6737** (0.0062)	**2.4564** (0.3425)	**1.8496** (0.0865)
Dataset 3	**CART**	0.8919 (0.0037)	10.9009 (0.2155)	3.8630 (0.0680)
	CatBoost	0.9066 (0.0028)	11.4956 (0.2095)	4.0079 (0.0454)
	CNN	0.8961 (0.0032)	10.8633 (0.1740)	3.8948 (0.0594)
	CNN-DN	0.8931 (0.0034)	10.6493 (0.2213)	3.8654 (0.0622)
	GNBC	0.8405 (0.0033)	7.9662 (0.2219)	2.6133 (0.0437)
	KNN	0.8508 (0.0041)	8.4245 (0.3162)	3.3723 (0.0457)
	LASSO	0.8711 (0.0029)	9.6743 (0.2378)	3.3073 (0.0442)
	LightGBM	0.9061 (0.0026)	11.4661 (0.1868)	3.9826 (0.0430)
	MLP	0.8973 (0.0031)	10.9491 (0.2034)	3.8968 (0.0516)
	RF	0.8990 (0.0026)	10.9888 (0.1725)	3.9294 (0.0492)
	XGBoost	0.9070 (0.0026)	11.4863 (0.1970)	3.9878 (0.0426)
	RLFF	0.9085 (0.0027)	11.5470 (0.1937)	4.0125 (0.0451)

	RXFF	**0.9087** (0.0026)	11.5590 (0.2050)	4.0293 (0.0452)
	RCFF	0.9082 (0.0031)	**11.5845** (0.2099)	**4.0426** (0.0474)
Dataset 4	**CART**	0.8713 (0.0024)	13.0378 (0.2283)	3.6812 (0.0317)
	CatBoost	0.8736 (0.0023)	13.2169 (0.1800)	3.6857 (0.0446)
	CNN	0.8666 (0.0025)	12.8184 (0.1888)	3.6220 (0.0399)
	CNN-DN	0.8635 (0.0022)	12.6751 (0.1902)	3.6125 (0.0385)
	GNBC	0.7981 (0.0032)	10.0589 (0.1870)	3.5017 (0.0364)
	KNN	0.8612 (0.0024)	12.5715 (0.2175)	3.6171 (0.0394)
	LASSO	0.8574 (0.0020)	12.2691 (0.1852)	3.5941 (0.0346)
	LightGBM	0.8737 (0.0021)	13.1895 (0.1806)	3.6877 (0.0433)
	MLP	0.8701 (0.0026)	13.0688 (0.2181)	3.6684 (0.0402)
	RF	0.8701 (0.0024)	13.0659 (0.1748)	3.6263 (0.0399)
	XGBoost	0.8734 (0.0024)	13.1867 (0.1832)	3.6896 (0.0418)
	RLFF	0.8739 (0.0024)	13.2347 (0.1940)	3.6897 (0.0367)
	RXFF	**0.8741** (0.0022)	13.2382 (0.1884)	**3.6901** (0.0392)
	RCFF	0.8740 (0.0022)	**13.2431** (0.1894)	3.6846 (0.0398)

Table 7 shows that the proposed RCFF secured optimal average ranks across all evaluation indicators. Notably, for EMPC, the proposed framework significantly outperformed most of the benchmark classifiers at the 95% confidence level, except for RXFF at the 80% confidence level. Regarding AUC, the proposed framework exhibits significant differences from almost all benchmark classifiers, except CatBoost and the fusion frameworks, at the 95% confidence level. In terms of TDL (10%), the proposed framework shows significant differences from all benchmark classifiers, except RXFF, RLFF, and CatBoost, at the 85% confidence level. Notably, there is no significant difference between the fusion frameworks for the statistical evaluation indicators. Moreover, the proposed RCFF significantly surpasses all benchmark classifiers, including the fusion framework, at least at the 80% confidence level in terms of profitability, but not in terms of the statistical indicators. Given that the primary objective of identifying PPers is profit maximization rather than minimizing misclassification errors, these statistical test results affirm the framework's ability to achieve the most profitable predictions.

Table 7
Average rank adjusted *p*-values of Holm post-hoc test of the proposed RCFF and benchmark classifiers.

Set	Classifier	AUC	EMPC	TDL (10%)
Control	**RCFF**	12.9750	13.5250	12.2375
Benchmarks	**CART**	4.5250*** (0.0000)	5.5250*** (0.0000)	6.9875*** (0.0000)
	CatBoost	10.8500** (0.0693)	10.9500*** (0.0177)	10.9625 (0.3457)
	CNN	5.7750*** (0.0000)	5.5750*** (0.0000)	5.7750*** (0.0000)
	CNN-DN	4.6750*** (0.0000)	4.4000*** (0.0000)	4.5750*** (0.0000)
	GNBC	1.1750*** (0.0000)	1.1500*** (0.0000)	1.1750*** (0.0000)
	KNN	3.6000*** (0.0000)	3.3250*** (0.0000)	3.7125*** (0.0000)
	LASSO	3.4250*** (0.0000)	3.4250*** (0.0000)	3.3500*** (0.0000)

LightGBM	9.2250*** (0.0003)	9.1750*** (0.0000)	9.4250*** (0.0132)
MLP	6.4000*** (0.0000)	6.5750*** (0.0000)	6.5000*** (0.0000)
RF	7.5500*** (0.0000)	8.0250*** (0.0000)	7.5750*** (0.0000)
XGBoost	10.0500*** (0.0071)	9.9250*** (0.0005)	10.2125* (0.1216)
RLFF	12.3000 (0.9411)	11.2250*** (0.0279)	10.4625 (0.1733)
RXFF	12.4750 (0.9411)	12.2000 (0.1566)	12.0500 (0.8411)

*, **, and *** indicate rejection of the null hypothesis in the Holm post-hoc test at the 85%, 90%, and 95% confidence levels, respectively.
Higher average ranks indicate higher prediction effectiveness.
The adjusted *p*-values are shown in parentheses.

To address the second question, we investigate whether the fusion mechanism based on bagging and boosting classifiers enhanced the profitability of the fusion frameworks, and conducted a comparative analysis of profitability between the three fusion frameworks and their constituent classifiers. **Table 6** reveals that all three fusion frameworks outperformed their component classifiers in terms of AUC and EMPC values. For instance, in Dataset 1, the RLFF achieved better average AUC (0.9269) and EMPC (€4.4454) values than the RF (average AUC: 0.9202, average EMPC: €4.3725) and LightGBM (average AUC: 0.9241, average EMPC: €4.4308). We observe similar trends when comparing RXFF with RF and XGBoost or RCFF with RF and CatBoost. Additionally, the fusion frameworks generally have the highest overall TDL (10%) value, except for RCFF in Dataset 4. In this dataset, RCFF achieves an average TDL (10%) value of 3.6846, surpassing RF (average TDL (10%): 3.6263), but slightly trailing CatBoost (average TDL (10%): 3.6857). This result suggests that CatBoost is adept at identifying the top 10% of PPers with the highest subscription potential compared with the proposed RCFF. However, the overall prediction profitability of CatBoost was notably inferior to that of RCFF, as evidenced by its lower EMPC value. Furthermore, the average EMPC value in the RCFF (€13.2431) is higher than that in CatBoost (€13.2169), confirming that the proposed fusion framework provides superior PPer prediction performance over CatBoost. To further substantiate the efficacy of the proposed fusion strategy, we apply Holm's post-hoc tests for each evaluation indicator among the three fusion frameworks and their corresponding component classifiers. The average ranks and test results are summarized in **Table 8**.

Table 8

Average rank and the adjusted *p*-values of Holm post-hoc test of the fusion frameworks and their component classifiers.

Set	Classifier	AUC	EMPC	TDL (10%)
Control	**RCFF**	2.9250	2.9250	2.5500
Benchmarks	**RF**	1.0000*** (0.0000)	1.0750*** (0.0000)	1.1625*** (0.0000)
	CatBoost	2.0750*** (0.0001)	2.0000*** (0.0000)	2.2875 (0.2404)
Control	**RLFF**	2.8750	2.7000	2.4000
Benchmarks	**RF**	1.2750*** (0.0000)	1.4250*** (0.0000)	1.5250*** (0.0002)
	LightGBM	1.8500*** (0.0000)	1.8750*** (0.0002)	2.0750* (0.1461)
Control	**RXFF**	2.8750	2.7750	2.6250
Benchmarks	**RF**	1.0750*** (0.0000)	1.2500*** (0.0000)	1.3375*** (0.0000)
	XGBoost	2.0500*** (0.0002)	1.9750*** (0.0003)	2.0375*** (0.0086)

*, **, and *** indicate rejection of the null hypothesis in the Holm post-hoc test at the 85%, 90%, and 95% confidence levels, respectively.
Higher average ranks indicate higher prediction effectiveness.
The adjusted *p*-values are shown in parentheses.
All comparisons of the evaluation indicators reject the null hypothesis of the Friedman test, with *p*-value=0.0000.

Most of the adjusted *p*-values from the Holm post-hoc test strongly suggest rejecting the null hypothesis at the 95% confidence level. Consequently, significant differences between the fusion frameworks and their respective component classifiers are evident in almost all scenarios. Hence, the fusion strategy successfully integrates the advantages of the bagging and boosting classifier, consistently delivering more profitable prediction results. These findings confirm the effectiveness of the developed fusion strategy.

Remark. Comparison and statistical analyses confirm the superiority of the proposed RCFF as the most profitable classifier for identifying PPers among customers. Moreover, the fusion strategy integrating the bagging and boosting classifiers enhances profitability relative to standalone bagging and boosting classifiers.

4.4. Interpretability Analysis

In addition to predictive performance, understanding the interpretability of the classifier outputs is crucial for practical applications. Previous studies explore the influence of specific variables on consumer purchase decisions (Scholz et al., 2018; von Helversen et al., 2018). Our study integrates SHAP to uncover the rationale behind predictions and assess the impact of each feature on prediction outcomes (Lundberg et al., 2020, 2018; Lundberg and Lee, 2017). Using Dataset 1 as an illustrative example, we aim to determine the influence of each attribute on the prediction results. The definitions of the features of Dataset 1 are presented in **Table A1** in the

Appendix. The SHAP values for each feature, calculated based on the predictions of the proposed RCFF corresponding to the first selected test set during ten-fold cross-validation of Dataset 1, are presented in **Fig. 2**. In **Fig. 2**, each dot represents a customer (instance) from the first test set in the 10-fold cross-validation of Dataset 1. The density of the points indicates the SHAP value for each feature. The vertical axis ranks the 38 features in descending order of their impact on the prediction results, whereas the horizontal axis indicates the positive or negative influence of the SHAP value on the outcome. As the feature value increases, the dot color transitions to red, whereas a decrease results in a blue dot. More red dots on the positive axis and more blue dots on the negative axis suggest a positive association with the outcome, indicating that higher feature values correspond to a greater likelihood of becoming a purchaser. Conversely, more blue dots on the positive axis and more red dots on the negative axis indicate a negative association, suggesting that higher feature values are associated with a lower probability of becoming a purchaser with a stronger consumption intention.

Analyzing the SHAP values for all features in Dataset 1, we find that the variable *last_contact_duration* has the most substantial impact on prediction output, exhibiting a positive correlation. *last_contact_duration* represents the duration of the last contact made with the customer. A longer last contact duration indicates increased customer interest in the product or service, correlating with a higher likelihood of the customer having stronger consumption intention. For practitioners, focusing on customers with an longer last contact duration can enhance the chances of marketing to individuals with robust consumption intentions. Following closely, *housing_loan* is the second crucial factor influencing the identification of PPers. This variable indicates whether a customer has taken a housing loan. **Fig. 2** illustrates that customers without housing loans are more likely to be purchasers. This observation suggests that customers without housing loans may possess more disposable income, which is potentially directed toward endeavors such as time deposits. Consequently, banks should pay careful attention to customers' housing loan statuses as a reference for affordability, aiding in the precise identification of purchasers.

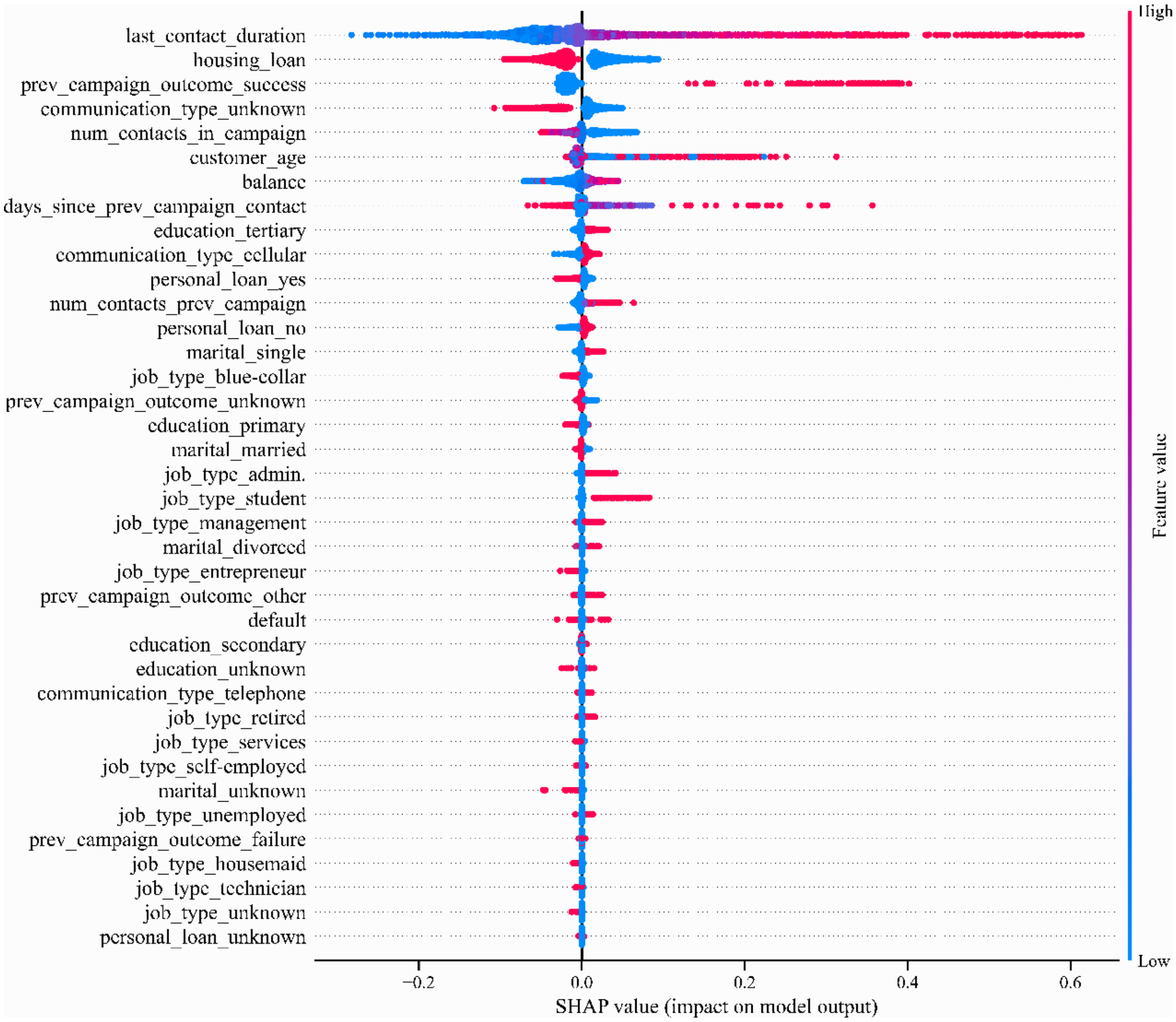

Fig.2. SHAP values of all features in Dataset 1.

Additionally, *prev_campaign_outcome_success* (success in the previous campaign) also influences prediction results. A customer retained in the previous customer relationship management instance successfully has a higher likelihood of becoming a purchaser. Success in the previous campaign signifies customer satisfaction with the time deposit service offered by the bank, indicating continued interest in subscribing to this service. In response, banks can strategically present attractive terms to customers who have been successfully retained in previous customer relationship management, further bolstering the likelihood of their continued engagement with time deposits. **Fig. 2** visually represents the contributions of the other features to the outcomes in Dataset 1.

Remark. SHAP values offer a visually interpretable representation of the contribution of each feature to the prediction outcome in the proposed RCFF. This result can help decision makers distinguish purchasers from customers by highlighting key features for reference purposes.

4.5. Bayesian A/B test

We used frequency statistics in Section 4.3 to validate the experiment (Friedman test, Holm post-hoc test). The results confirm the superior profitability of the proposed RCFF over benchmark classifiers. Here, we employ Bayesian A/B testing to assess the immediate probability of the proposed framework's superiority and quantify the risk associated with its implementation compared to other classifiers. Unlike traditional statistical tests that rely on frequentist statistics, Bayesian A/B testing offers enhanced interpretability and applicability. Traditional frequentist tests determine *p*-values and assess the null hypothesis rejection without providing specific probabilities of one group's superiority. Moreover, the interpretive power of Bayesian A/B testing is superior and does not require a large sample size, offering a cost-effective analysis using existing information.

In our study, we adopted the Bayesian A/B testing framework, utilizing the EMPC ranks of models as our evaluation metric. We conducted 10-fold cross-validation on four datasets to calculate the EMPC ranks for each model, which served as proxies for the profits gained by enterprises relying on the predictions of these models. The Bayesian A/B test was executed using the Python package [6]. Given our criterion, which focuses on the rank of the models' EMPC value, we utilized the discrete variant of the Bayesian A/B test class for our analysis. **Table 8** presents the average ranks for the EMPC and Bayesian A/B test outcomes, focusing on the EMPC ranks between the proposed RCFF and the benchmark classifiers. The RCFF has the highest average

---

[6] https://github.com/Matt52/bayesian-testing

rank at 13.5250, surpassing all comparators, indicating a superior EMPC performance. The Bayesian A/B test results indicate a 93.96% probability that the proposed framework outperformed the comparative classifiers, translating to a mere 6.04% likelihood of making a suboptimal decision when employing the proposed RCFF for PPer prediction. In contrast, employing CART, CNN, CNN-DN, GNBC, KNN, LASSO, LightGBM, MLP, RF, and XGBoost for PPer prediction yield a 0.00% probability of outperforming the proposed RCFF. Even in the rare instances that RLFF or RXFF may slightly exceed the proposed RCFF, the probabilities are minimal (0.17% and 5.85 %, respectively). Furthermore, the risk associated with implementing the proposed RCFF, measured as the expected loss, was 0.0176, which is significantly lower than that of all prediction models. This robustly affirms the likelihood that the proposed framework delivers superior EMPC values with minimal anticipated loss compared to the individual benchmark classifiers and other fusion frameworks.

Table 8
Bayesian A/B test based on 40 EMPC ranks for the proposed RCFF and comparative classifiers.

Classifier	Average Ranks	Probability Being Best	Expected Loss
CART	5.5250	0.00%	5.9380
CatBoost	10.9500	0.03%	1.9306
CNN	5.5750	0.00%	5.8999
CNN-DN	4.4000	0.00%	6.7758
GNBC	1.1500	0.00%	9.1860
KNN	3.3250	0.00%	7.5730
LASSO	3.4250	0.00%	7.5032
LightGBM	9.1750	0.00%	3.2402
MLP	6.5750	0.00%	5.1634
RF	8.0250	0.00%	4.0901
XGBoost	9.9250	0.00%	2.6826
RLFF	11.2250	0.17%	1.7198
RXFF	12.2000	5.85%	0.9986
RCFF	13.5250	93.96%	0.0176

Note: In principle, the sum of the probabilities of being the best of all models should be 1. However, the sum is typically not equal to 1 after leaving four decimal places.

Table 9 employs Bayesian A/B tests based on the EMPC ranks to further assess the superiority of the fusion frameworks resulting from integrating bagging and boosting over their individual component classifiers. The average ranks of the EMPC and the comparative classifiers are presented. Similar to **Table 8**, the average rank values for the RCFF, RLFF, and RXFF are 2.9250, 2.7000, and 2.7750, respectively, indicating the top positions among the individual benchmarks. Bayesian A/B tests convincingly indicate a 100% probability that all three fusion frameworks outperform their component classifiers, with an associated expected loss of 0.0000. Thus, the proposed fusion strategy significantly surpasses individual classifiers in profit generation, posing minimal risk to decision makers compared with employing component classifiers.

Table 9
Bayesian A/B test based on 40 EMPC ranks for the fusion frameworks and their component classifiers.

Classifier	Average Ranks	Probability Being Best	Expected Loss
RF	1.0750	0.00%	1.7216
CatBoost	2.0000	0.00%	0.8611
RCFF	2.9250	100.00%	0.0000
Classifier	**Average Ranks**	**Probability Being Best**	**Expected Loss**
RF	1.4250	0.00%	1.1857
LightGBM	1.8750	0.00%	0.7665
RLFF	2.7000	100.00%	0.0000
Classifier	**Average Ranks**	**Probability Being Best**	**Expected Loss**
RF	1.2500	0.00%	1.4206
XGBoost	1.9750	0.00%	0.7453
RXFF	2.7750	100.00%	0.0000

Remark. The Bayesian A/B test confirms the high probability that the proposed RCFF outperforms the comparative classifiers and that the proposed fusion strategy demonstrates a 100% likelihood of surpassing its individual component classifiers.

5. Conclusion and Discussion

5.1. Conclusion

Profit-driven decision making is crucial for enterprises seeking long-term success and market competitiveness. In maximizing profit, identifying PPers is paramount. This study introduces an innovative RCFF for profit-oriented PPer prediction by integrating CatBoost into RF. Notably, profit-based indicators guide hyperparameter optimization via a grid search, thereby elevating the profitability of the proposed framework. Extensive validation using four diverse real-world datasets compares the performance of the RCFF against benchmark classifiers. The SHAP values reveal the intricate relationships between the input features and prediction outcomes. In addition, Bayesian A/B tests quantitatively assess the likelihood of the proposed fusion framework surpassing both the benchmark classifiers and their individual component classifiers.

The key findings can be summarized as follows. (i) Empirical experiments and statistical tests confirm the superior profitability of the proposed RCFF, as evidenced by its higher EMPC values compared to other classifiers. The integration of bagging and boosting through the developed fusion strategy surpasses standalone classifiers in profit generation. (ii) Interpretability analysis identifies the crucial factors that influence a customer's likelihood of becoming a purchaser of a bank's time deposit. This insight-rich information aids decision makers identify PPers precisely and execute targeted marketing with minimal time and cost. (iii) Bayesian A/B tests confirm the high probability of the proposed RCFF outperforming all comparative classifiers, whereas the fusion strategy consistently outperformed the standalone classifiers with a 100% probability of superiority.

5.2. Academic Applications for Retailing and Consumer Services

This study makes significant academic contributions to retail and consumer services on several fronts. First, the introduction of the RCFF effectively predicts PPers, yielding increased profits. Second, by comparing the fusion frameworks with their component classifiers, we establish that the proposed fusion strategy enhances the prediction performance for PPers by integrating bagging and boosting classifiers. Third, this study advances the exploration of PPer predictions from a profit-driven perspective, encouraging businesses to select classifiers based on profitability for precision marketing aligned with their objectives. Finally, an interpretative analysis of profit-driven classifier prediction results offers a novel perspective for researchers.

5.3. Managerial Applications for Retailing and Consumer Services

The managerial applications of this study for retailing and consumer services are twofold:

1. Enterprises can leverage the proposed RCFF to predict PPers within customer groups and identify those with stronger consumption intentions. For these identified prospects targeted

strategies, such as price discounts or consumer gifts, can be employed to encourage subscriptions or purchases, thereby enhancing the overall experience. This prevents the allocation of resources to customers who are not inclined to consume. Through precise customer targeting, businesses can more effectively utilize limited resources by concentrating them on the most likely to convert customers, thereby increasing the efficiency and return on investment of promotional activities. Based on the proposed RCFF, companies can identify the most promising customer segments and deliver targeted promotional messages to them. This not only reduces disturbances to customers who are not interested or lack purchasing power but also allows resources to be focused on customers with genuine purchasing intent, thereby improving the conversion rate of promotional activities.

2. Enterprises can conduct explanatory analyses of profit-driven RCFF prediction results. This study provides insights into the influence of each customer feature on the consumption propensity, enabling the determination of consumption propensities within specific customer groups. For instance, in Dataset 1, which is related to fixed deposit subscriptions, customers with longer contact times tend to have a stronger inclination to subscribe, whereas those with housing loans may lack the capacity for time deposit subscriptions. For those customers with longer contact times and a propensity to subscribe to fixed deposits, special discounts or customized services for fixed deposit products can be offered to further attract their purchases. As for customers with housing loans who may lack the capacity for time deposit subscriptions, other types of deposit products can be introduced or corresponding financial advisory services provided to meet their financial needs.

5.4. Specific Implications for Retailing and Consumer Services

The proposed fusion framework predicts PPers of credit card services, fixed-term deposit services, 5G packages, and health insurance services. The specific implications of our study for retail and consumer services are as follows:

Credit card services:

Implications for retailing: Anticipating potential credit card purchasers enables retailers to target promotions for credit card products and provide customized incentives and rewards to boost credit card sales.

Implications for consumer services: Consumers benefit from customized credit card products, which enhance their shopping experiences. Personalized credit card services also help prevent fraud and strengthen consumer financial security.

Fixed-term deposit services:

Implications for retailing: Predicting fixed-term deposit purchasers helps retailers design more attractive savings plans, thereby driving growth in financial product sales.

Implications for consumer services: Consumers gain access to more appealing interest rates and other benefits while better managing their finances for wealth appreciation.

5G package services:

Implications for retailing: Predicting 5G package purchasers will assist retailers in promoting high-speed data services, personalized plans, and value-added services.

Implications for consumer services: Consumers experience faster and more reliable network services and can choose personalized 5G plans that suit their needs.

Health insurance services:

Implications for retailing: Predicting health insurance purchasers helps retailers design more attractive insurance products and provide targeted health management services.

Implications for consumer services: Consumers benefit from more suitable health insurance plans and personalized health management advice and services to enhance their overall well-being.

In summary, by predicting PPers, retailers and consumer service providers can better meet individual needs, improve sales efficiency, reduce marketing costs, and create more valuable shopping and service experiences for consumers. This personalized and precise marketing approach helps build customer loyalty and satisfaction.

5.5. Limitations and Future Research

Several limitations of this study merit further exploration in future research. First, the constructed fusion framework relies on two types of ensemble classifiers with tree structures, limiting its ability to process unstructured information such as text or images. In PPer prediction tasks, it may be necessary to mine the intrinsic information in unstructured data. Owing to this limitation, we can extract structured features from unstructured data, such as the emotional scoring of customer review texts or extracting features of customer review pictures. Future studies could consider employing deep-learning techniques to mine the intrinsic information of unstructured data to improve the PPer prediction of the proposed fusion framework, particularly in scenarios involving multimodal information. For example, delving deeper into promotional conversations with customers, using pre-trained BERT to assess customer sentiment, thereby assisting in PPer predictions. Second, the fusion framework was primarily tested within the confines of PPer prediction, introducing a notable limitation to its generalizability across diverse fields. This study focuses on profit-driven predictive tasks by employing profit evaluation metrics tailored to predict PPers. However, other domains may have different profit evaluation metrics. Therefore, the application of the proposed fusion framework to other predictive scenarios requires, as a first step, fine-tuning existing research evaluation metrics or designing evaluation metrics capable of assessing predictive profits. This requires researchers to have a sufficient understanding of current predictive scenarios. Regarding the model itself, feature specifications were not explicitly set because the current prediction scenario concentrated on forecasting PPers. In future studies, we plan to assess the performance of the proposed fusion framework across various prediction scenarios to gauge its applicability and universality beyond the context of forecasting PPers.

Appendix

Feature	Meaning for customer
customer_age	Age in years
job_type_blue-collar	The job type is blue-collar
job_type_management	The job type is management
job_type_admin	The job type is admin
job_type_student	The job type is student
job_type_entrepreneur	The job type is entrepreneur
job_type_retired	The job type is retired
job_type_services	The job type is service
job_type_self-employed	The job type is self-employed
job_type_unemployed	The job type is unemployed
job_type_housemaid	The job type is housemaid
job_type_technician	The job type is technician
job_type_unknown	The job type is unknown
marital_single	Marital status is single
marital_married	Marital status is married
marital_divorced	Marital status is divorced
marital_unknown	Marital status is unknown
education_tertiary	Education level is tertiary
education_primary	Education level is primary
education_secondary	Education level is secondary
education_unknown	Education level is unknown
default	Whether customer has defaulted in past
balance	Current balance in bank
housing_loan	Has customer taken a housing loan
personal_loan_yes	Customer has a personal loan
personal_loan_no	Customer does not have a personal loan
personal_loan-unknown	Whether customer has personal loan is unknown
communication_type_unknown	The communication type is unknown
communication_type_cellular	The communication type is cellular
communication_type_telephone	The communication type is telephone
last_contact_duration	Last contact duration made (in seconds)
number_contects_in_campaign	Number of contacts made during the current campaign
Days_since_prev_campaign_contact	Number of days passed since was contacted in previous campaign
Num_contacts_prev_campaign	Number of contacts made during the previous campaign
prev_campaign_outcome_success	Success in previous campaign
prev_campaign_outcome_failure	Failure in previous campaign
prev_campaign_outcome_unknown	The outcome in previous campaign is unknown
prev_campaign_outcome_other	The outcome in previous campaign is other

Table A1: The meanings of features in Dataset 1

Dataset	Classifier	F1-Score	Precision	Recall
Dataset 1	**CART**	0.9505 (0.0037)	0.9303 (0.0064)	0.9717 (0.0041)
	CatBoost	**0.9533** (0.0021)	0.9347 (0.0037)	0.9726 (0.0032)
	CNN	0.9521 (0.0019)	0.9309 (0.0032)	0.9743 (0.0044)
	CNN-DN	0.9516 (0.0019)	0.9279 (0.0041)	0.9766 (0.0042)
	GNBC	0.8966 (0.0046)	0.9432 (0.0038)	0.8545 (0.0068)
	KNN	0.9479 (0.0032)	0.9113 (0.0053)	0.9877 (0.0019)
	LASSO	0.9505 (0.0030)	0.9214 (0.0048)	0.9817 (0.0023)
	LightGBM	0.9088 (0.0042)	0.9813 (0.0020)	0.8463 (0.0073)
	MLP	0.9465 (0.0032)	0.9185 (0.0249)	0.9775 (0.0226)
	RF	0.9506 (0.0028)	0.9134 (0.0049)	**0.9909** (0.0016)
	XGBoost	0.9529 (0.0018)	0.9335 (0.0031)	0.9731 (0.0035)
	RLFF	0.9110 (0.0041)	**0.9814** (0.0013)	0.8500 (0.0073)
	RXFF	0.9531 (0.0021)	0.9338 (0.0032)	0.9732 (0.0034)
	RCFF	0.9530 (0.0028)	0.9351 (0.0036)	0.9717 (0.0039)
Dataset 2	**CART**	0.8615 (0.0026)	0.7652 (0.0045)	0.9855 (0.0048)
	CatBoost	0.8633 (0.0028)	0.7634 (0.0037)	0.9931 (0.0025)
	CNN	0.8632 (0.0027)	0.7615 (0.0042)	0.9961 (0.0014)
	CNN-DN	0.8636 (0.0026)	0.7608 (0.0044)	0.9986 (0.0016)
	GNBC	0.4216 (0.0117)	**0.8508** (0.0087)	0.2803 (0.0101)
	KNN	**0.8637** (0.0026)	0.7602 (0.0041)	0.9996 (0.0004)
	LASSO	0.8636 (0.0026)	0.7606 (0.0041)	0.9990 (0.0008)

	LightGBM	0.7332 (0.0323)	0.8149 (0.0098)	0.6690 (0.0553)
	MLP	0.8561 (0.0053)	0.7716 (0.0075)	0.9617 (0.0201)
	RF	**0.8637** (0.0027)	0.7600 (0.0041)	**1.0000** (0.0000)
	XGBoost	0.8632 (0.0026)	0.7629 (0.0043)	0.9940 (0.0026)
	RLFF	0.7042 (0.0128)	0.8367 (0.0059)	0.6083 (0.0198)
	RXFF	0.8632 (0.0027)	0.7626 (0.0044)	0.9944 (0.0027)
	RCFF	0.8632 (0.0024)	0.7631 (0.0040)	0.9937 (0.0017)
Dataset 3	**CART**	0.9225 (0.0018)	0.9087 (0.0035)	0.9369 (0.0033)
	CatBoost	0.9257 (0.0018)	0.9114 (0.0025)	0.9405 (0.0030)
	CNN	0.9215 (0.0015)	0.9023 (0.0034)	0.9416 (0.0046)
	CNN-DN	0.9197 (0.0017)	0.9017 (0.0062)	0.9386 (0.0079)
	GNBC	0.8707 (0.0023)	0.9217 (0.0023)	0.8250 (0.0042)
	KNN	0.9026 (0.0017)	0.8728 (0.0020)	0.9344 (0.0027)
	LASSO	0.9019 (0.0013)	0.8533 (0.0022)	**0.9563** (0.0013)
	LightGBM	0.8940 (0.0016)	**0.9500** (0.0024)	0.8442 (0.0028)
	MLP	0.9212 (0.0019)	0.9079 (0.0070)	0.9350 (0.0098)
	RF	0.9240 (0.0017)	0.9047 (0.0026)	0.9442 (0.0024)
	XGBoost	0.9257 (0.0018)	0.9132 (0.0028)	0.9385 (0.0027)
	RLFF	0.8996 (0.0018)	0.9484 (0.0030)	0.8556 (0.0039)
	RXFF	**0.9264** (0.0018)	0.9132 (0.0027)	0.9400 (0.0027)
	RCFF	0.9263 (0.0016)	0.9122 (0.0027)	0.9408 (0.0027)
Dataset 4	**CART**	0.9125 (0.0012)	0.8660 (0.0044)	0.9644 (0.0053)
	CatBoost	0.9128 (0.0013)	0.8723 (0.0024)	0.9573 (0.0012)
	CNN	0.9106 (0.0017)	0.8702 (0.0033)	0.9549 (0.0033)
	CNN-DN	0.9111 (0.0013)	0.8603 (0.0034)	0.9682 (0.0038)
	GNBC	0.7771 (0.0024)	0.9016 (0.0020)	0.6829 (0.0031)
	KNN	0.9096 (0.0015)	0.8644 (0.0028)	0.9597 (0.0018)
	LASSO	0.9076 (0.0014)	0.8480 (0.0023)	**0.9762** (0.0010)
	LightGBM	0.8765 (0.0021)	**0.9226** (0.0021)	0.8347 (0.0031)
	MLP	0.9119 (0.0014)	0.8720 (0.0033)	0.9557 (0.0054)
	RF	0.9128 (0.0012)	0.8586 (0.0022)	0.9744 (0.0007)
	XGBoost	**0.9132** (0.0014)	0.8704 (0.0026)	0.9604 (0.0014)
	RLFF	0.8792 (0.0017)	0.9216 (0.0021)	0.8404 (0.0026)
	RXFF	0.9131 (0.0013)	0.8709 (0.0025)	0.9596 (0.0010)
	RCFF	0.9130 (0.0015)	0.8726 (0.0028)	0.9574 (0.0013)

Table A2: The statistical evaluation indicators of classifiers

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

Al-Weshah, G.A., Al-Manasrah, E., Al-Qatawneh, M., 2019. Customer relationship management systems and organizational performance: Quantitative evidence from the Jordanian telecommunication industry. J. Mark. Commun. https://doi.org/10.1080/13527266.2018.1449007

Aslam, J.A., Popa, R.A., Rivest, R.L., 2007. On estimating the size and confidence of a statistical audit, in: EVT 2007 - 2007 USENIX/ACCURATE Electronic Voting Technology Workshop.

Avrizal, R., Wibowo, A., Yuniarti, A.S., Sandy, D.A., Prihandani, K., 2018. Analysis Comparison of The Classification Data Mining Method to Predictthe Decisions of Potential Customer Insurance. Int. J. Comput. Tech. 5, 15–20.

Basuki, A., 2016. Customer classification using Learning Vector Quantization Neural Network. J. Telecommun. Electron. Comput. Eng. 8, 131–135.

Berloco, C., Argiento, R., Montagna, S., 2022. Forecasting short-term defaults of firms in a commercial network via Bayesian spatial and spatio-temporal methods. Int. J. Forecast. https://doi.org/10.1016/j.ijforecast.2022.05.003

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140. https://doi.org/10.1007/BF00058655

Breiman, L., Friedman, J., Stone, C., Olshen, R., 1984. Classification and Regression Trees (Wadsworth Statistics/Probability). New York CRC Press.

Chaudhuri, N., Gupta, G., Vamsi, V., Bose, I., 2021. On the platform but will they buy? Predicting customers' purchase behavior using deep learning. Decis. Support Syst. 149, 113622. https://doi.org/10.1016/j.dss.2021.113622

Chen, S.-S., Choubey, B., Singh, V., 2021. A neural network based price sensitive recommender model to predict customer choices based on price effect. J. Retail. Consum. Serv. 61, 102573. https://doi.org/https://doi.org/10.1016/j.jretconser.2021.102573

Chen, S., Wang, X., Zhang, H., Wang, J., 2021. Customer purchase prediction from the perspective of imbalanced data: A machine learning framework based on factorization machine. Expert Syst. Appl. 173, 114756. https://doi.org/10.1016/j.eswa.2021.114756

Chen, T.-Y., Liu, Y.-C., Chen, Y.-M., 2016. A method of potential customer searching from opinions of network villagers in virtual communities. Online Inf. Rev. 40, 146–167. https://doi.org/10.1108/OIR-12-2014-0295

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794. https://doi.org/10.1145/2939672.2939785

Chen, W.C., Hsu, C.C., Hsu, J.N., 2011. Optimal selection of potential customer range through the union sequential pattern by using a response model. Expert Syst. Appl. 38, 7451–7461. https://doi.org/10.1016/j.eswa.2010.12.078

Chen, Y., Li, Y., Wu, M., Lu, F., Hou, M., Yin, Y., 2022. Differentiating Crohn's disease from intestinal tuberculosis using a fusion correlation neural network. Knowledge-Based Syst. 244, 108570. https://doi.org/10.1016/j.knosys.2022.108570

Chou, P., Chuang, H.H.-C., Chou, Y.-C., Liang, T.-P., 2022. Predictive analytics for customer repurchase: Interdisciplinary integration of buy till you die modeling and machine learning. Eur. J. Oper. Res. 296, 635–651. https://doi.org/10.1016/j.ejor.2021.04.021

Choudhury, A.M., Nur, K., 2019. A machine learning approach to identify potential customer based on purchase behavior. 1st Int. Conf. Robot. Electr. Signal Process. Tech. ICREST

2019 242–247. https://doi.org/10.1109/ICREST.2019.8644458
Coussement, K., Harrigan, P., Benoit, D.F., 2015. Improving direct mail targeting through customer response modeling. Expert Syst. Appl. 42, 8403–8412. https://doi.org/10.1016/j.eswa.2015.06.054
Coussement, K., Lessmann, S., Verstraeten, G., 2017. A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. Decis. Support Syst. 95, 27–36. https://doi.org/10.1016/j.dss.2016.11.007
Cui, Y., Tobossi, R., Vigouroux, O., 2018. Modelling customer online behaviours with neural networks: applications to conversion prediction and advertising retargeting.
Das, T.K., 2016. A customer classification prediction model based on machine learning techniques. Proc. 2015 Int. Conf. Appl. Theor. Comput. Commun. Technol. iCATccT 2015 321–326. https://doi.org/10.1109/ICATCCT.2015.7456903
De Bock, K.W., De Caigny, A., 2021. Spline-rule ensemble classifiers with structured sparsity regularization for interpretable customer churn modeling. Decis. Support Syst. 150, 113523. https://doi.org/10.1016/j.dss.2021.113523
De Caigny, A., Coussement, K., De Bock, K.W., 2018. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. Eur. J. Oper. Res. 269, 760–772. https://doi.org/10.1016/j.ejor.2018.02.009
De Caigny, A., Coussement, K., De Bock, K.W., Lessmann, S., 2020. Incorporating textual information in customer churn prediction models based on a convolutional neural network. Int. J. Forecast. 36, 1563–1578. https://doi.org/10.1016/j.ijforecast.2019.03.029
DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics 44, 837–845. https://doi.org/10.2307/2531595
Domingos, P., Pazzani, M., 1997. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Mach. Learn. 29, 103–130. https://doi.org/10.1023/a:1007413511361
Esmeli, R., Bader-El-Den, M., Abdullahi, H., 2022. An analyses of the effect of using contextual and loyalty features on early purchase prediction of shoppers in e-commerce domain. J. Bus. Res. 147, 420–434. https://doi.org/10.1016/j.jbusres.2022.04.012
Fader, P.S., Hardie, B.G.S., 2009. Probability Models for Customer-Base Analysis. J. Interact. Mark. 23, 61–69. https://doi.org/10.1016/j.intmar.2008.11.003
Fumera, G., Roli, F., Serrau, A., 2005. Dynamics of variance reduction in bagging and other techniques based on randomisation, in: Lecture Notes in Computer Science. https://doi.org/10.1007/11494683_32
Gamage, T.C., Gnanapala, A., Ashill, N.J., 2021. Understanding social customer relationship management adoption: qualitative insights. J. Strateg. Mark. https://doi.org/10.1080/0965254X.2021.1923056
Ganesh, J., Arnold, M.J., Reynolds, K.E., 2000. Understanding the Customer Base of Service Providers: An Examination of the Differences between Switchers and Stayers. J. Mark. 64, 65–87. https://doi.org/10.1509/jmkg.64.3.65.18028
Gengler, C.E., Popkowski Leszczyc, P.T.L., 1997. Using customer satisfaction research for relationship marketing: A direct marketing approach. J. Direct Mark. 11, 23–29. https://doi.org/10.1002/(SICI)1522-7138(199724)11:1<23::AID-DIR3>3.0.CO;2-0
Höppner, S., Stripling, E., Baesens, B., Broucke, S. vanden, Verdonck, T., 2020. Profit driven decision trees for churn prediction. Eur. J. Oper. Res. 284, 920–933. https://doi.org/10.1016/j.ejor.2018.11.072
Hossain, M.S., Rahman, M.F., 2022a. Detection of potential customers’ empathy behavior towards customers’ reviews. J. Retail. Consum. Serv. 65, 102881. https://doi.org/10.1016/j.jretconser.2021.102881

Hossain, M.S., Rahman, M.F., 2022b. Customer Sentiment Analysis and Prediction of Insurance Products' Reviews Using Machine Learning Approaches. FIIB Bus. Rev. 12, 386–402. https://doi.org/10.1177/23197145221115793

Hu, Q., Xie, S., Zhang, J., Zhu, Q., Guo, S., Yu, P.S., 2016. Hetero sales: Utilizing heterogeneous social networks to identify the next enterprise customer. 25th Int. World Wide Web Conf. WWW 2016 41–50. https://doi.org/10.1145/2872427.2883000

Jaiswal, D.P., Kumar, S., Mukherjee, P., 2020. Customer transaction prediction system. Procedia Comput. Sci. 168, 49–56. https://doi.org/10.1016/j.procs.2020.02.256

Jamal, Z., Bucklin, R.E., 1987. Improving the diagnosis and prediction of customer churn: A heterogeneous hazard modeling approach. J. Interact. Mark. 20, 16–29. https://doi.org/10.1002/dir.20064

Jiang, P., Liu, Z., Abedin, M.Z., Wang, J., Yang, W., Dong, Q., 2024. Profit-driven weighted classifier with interpretable ability for customer churn prediction. Omega 125, 103034. https://doi.org/10.1016/j.omega.2024.103034

Jiang, P., Liu, Z., Zhang, L., Wang, J., 2023. Hybrid model for profit-driven churn prediction based on cost minimization and return maximization. Expert Syst. Appl. 228, 120354. https://doi.org/10.1016/j.eswa.2023.120354

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. LightGBM: A highly efficient gradient boosting decision tree, in: Advances in Neural Information Processing Systems. pp. 3147–3155.

Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., Abbasi, U., 2014. Improved churn prediction in telecommunication industry using data mining techniques. Appl. Soft Comput. J. 24, 994–1012. https://doi.org/10.1016/j.asoc.2014.08.041

Kim, J., Ji, H., Oh, S., Hwang, S., Park, E., del Pobil, A.P., 2021. A deep hybrid learning model for customer repurchase behavior. J. Retail. Consum. Serv. 59, 102381. https://doi.org/10.1016/j.jretconser.2020.102381

Kotsianti, S.B., Kanellopoulos, D., 2007. Combining bagging, boosting and dagging for classification problems, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-540-74827-4_62

Kozodoi, N., Lessmann, S., Papakonstantinou, K., Gatsoulis, Y., Baesens, B., 2019. A multi-objective approach for profit-driven feature selection in credit scoring. Decis. Support Syst. 120, 106–117. https://doi.org/10.1016/j.dss.2019.03.011

Lei, S., Zhu, D., 2012. Web potential customer classification based on SVM, in: Proceedings of the 2012 International Conference on Industrial Control and Electronics Engineering, ICICEE 2012. pp. 568–570. https://doi.org/10.1109/ICICEE.2012.155

Lessmann, S., Haupt, J., Coussement, K., De Bock, K.W., 2021. Targeting customers for profit: An ensemble learning framework to support marketing decision-making. Inf. Sci. (Ny). 557, 286–301. https://doi.org/10.1016/j.ins.2019.05.027

Li, B., Liao, M., Yuan, J., Zhang, J., 2023. Green consumption behavior prediction based on fan-shaped search mechanism fruit fly algorithm optimized neural network. J. Retail. Consum. Serv. 75, 103471. https://doi.org/10.1016/j.jretconser.2023.103471

Li, J., Pan, S., Huang, L., Zhu, X., 2019. A machine learning based method for customer behavior prediction. Tech. Gaz. 26, 1670–1676. https://doi.org/10.17559/TV-20190603165825

Li, Y., 2019. Customer Identification of Potential Energy Substitution Based on Big Data Method, in: Proceedings of the 2019 International Conference on Computer, Network, Communication and Information Systems (CNCI 2019). Atlantis Press, Paris, France, pp. 63–73. https://doi.org/10.2991/cnci-19.2019.9

Liu, C.-H., Gan, B., Ko, W.-H., Teng, C.-C., 2022. Comparison of localized and foreign restaurant brands for consumer behavior prediction. J. Retail. Consum. Serv. 65, 102868. https://doi.org/10.1016/j.jretconser.2021.102868
Liu, Z., Jiang, P., De Bock, K.W., Wang, J., Zhang, L., Niu, X., 2024. Extreme gradient boosting trees with efficient Bayesian optimization for profit-driven customer churn prediction. Technol. Forecast. Soc. Change 198, 122945. https://doi.org/10.1016/j.techfore.2023.122945
Liu, Z., Jiang, P., Wang, J., Du, Z., Niu, X., Zhang, L., 2023. Hospitality order cancellation prediction from a profit-driven perspective. Int. J. Contemp. Hosp. Manag. 35, 2084–2112. https://doi.org/10.1108/IJCHM-06-2022-0737
Liu, Z., Wang, X., Li, Y., Yao, L., An, J., Bai, L., Lim, E.-P., 2022. Face to purchase: Predicting consumer choices with structured facial and behavioral traits embedding. Knowledge-Based Syst. 235, 107665. https://doi.org/10.1016/j.knosys.2021.107665
Lu, X., He, S., Lian, S., Ba, S., Wu, J., 2020. Is user-generated content always helpful? The effects of online forum browsing on consumers' travel purchase decisions. Decis. Support Syst. 137, 113368. https://doi.org/10.1016/j.dss.2020.113368
Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9
Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions, in: Proceedings of the 31st International Conference on Neural Information Processing Systems. pp. 4768–4777. https://doi.org/10.5555/3295222.3295230
Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.W., Newman, S.F., Kim, J., Lee, S.I., 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat. Biomed. Eng. 2, 749–760. https://doi.org/10.1038/s41551-018-0304-0
Maldonado, S., Domínguez, G., Olaya, D., Verbeke, W., 2021. Profit-driven churn prediction for the mutual fund industry: A multisegment approach. Omega 100, 102380. https://doi.org/10.1016/j.omega.2020.102380
Maldonado, S., Flores, Á., Verbraken, T., Baesens, B., Weber, R., 2015. Profit-based feature selection using support vector machines - General framework and an application for customer retention. Appl. Soft Comput. J. 35, 740–748. https://doi.org/10.1016/j.asoc.2015.05.058
Maldonado, S., López, J., Vairetti, C., 2020. Profit-based churn prediction based on Minimax Probability Machines. Eur. J. Oper. Res. 284, 273–284. https://doi.org/10.1016/j.ejor.2019.12.007
Martínez, A., Schmuck, C., Pereverzyev, S., Pirker, C., Haltmeier, M., 2020. A machine learning framework for customer purchase prediction in the non-contractual setting. Eur. J. Oper. Res. 281, 588–596. https://doi.org/10.1016/j.ejor.2018.04.034
Moeyersoms, J., Martens, D., 2015. Including high-cardinality attributes in predictive models: A case study in churn prediction in the energy sector. Decis. Support Syst. https://doi.org/10.1016/j.dss.2015.02.007
Nilashi, M., Ahmadi, H., Arji, G., Alsalem, K.O., Samad, S., Ghabban, F., Alzahrani, A.O., Ahani, A., Alarood, A.A., 2021. Big social data and customer decision making in vegetarian restaurants: A combined machine learning method. J. Retail. Consum. Serv. 62, 102630. https://doi.org/10.1016/j.jretconser.2021.102630
Palaniappan, S., Mustapha, A., Foozy, C.F.M., Atan, R., 2017. Customer profiling using classification approach for bank telemarketing. Int. J. Informatics Vis. 1, 214–217.

https://doi.org/10.30630/joiv.1.4-2.68
Papouskova, M., Hajek, P., 2019. Two-stage consumer credit risk modelling using heterogeneous ensemble learning. Decis. Support Syst. 118, 33–45. https://doi.org/10.1016/j.dss.2019.01.002
Pashchenko, Y., Rahman, M.F., Hossain, M.S., Uddin, M.K., Islam, T., 2022. Emotional and the normative aspects of customers' reviews. J. Retail. Consum. Serv. 68, 103011. https://doi.org/10.1016/j.jretconser.2022.103011
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. Catboost: Unbiased boosting with categorical features, in: Advances in Neural Information Processing Systems. pp. 6639–6649. https://doi.org/10.5555/3327757.3327770
Rahim, M.A., Mushafiq, M., Khan, S., Arain, Z.A., 2021. RFM-based repurchase behavior for customer classification and segmentation. J. Retail. Consum. Serv. 61, 102566. https://doi.org/10.1016/j.jretconser.2021.102566
Risselada, H., Verhoef, P.C., Bijmolt, T.H.A., 2010. Staying Power of Churn Prediction Models. J. Interact. Mark. 24, 198–208. https://doi.org/10.1016/j.intmar.2010.04.002
Rozo, B.J.G., Crook, J., Andreeva, G., 2022. The role of web browsing in credit risk prediction. Decis. Support Syst. 113879. https://doi.org/10.1016/j.dss.2022.113879
Schapire, R.E., 1990. The strength of weak learnability. Mach. Learn. https://doi.org/10.1007/bf00116037
Scholz, M., Schnurbus, J., Haupt, H., Dorner, V., Landherr, A., Probst, F., 2018. Dynamic effects of user- and marketer-generated content on consumer purchase behavior: Modeling the hierarchical structure of social media websites. Decis. Support Syst. 113, 43–55. https://doi.org/10.1016/j.dss.2018.07.001
Shah, M., Skandan, K.A., Shivani Sweta, S., Gaur, G., Prajwala, T.R., 2022. Customer Purchase Intention Prediction Using Text Analytical Models, in: 2022 IEEE 7th International Conference for Convergence in Technology, I2CT 2022. IEEE, pp. 23–26. https://doi.org/10.1109/I2CT54291.2022.9825140
Stripling, E., vanden Broucke, S., Antonio, K., Baesens, B., Snoeck, M., 2018. Profit maximizing logistic model for customer churn prediction using genetic algorithms. Swarm Evol. Comput. 40, 116–130. https://doi.org/10.1016/j.swevo.2017.10.010
Sundarkumar, G.G., Ravi, V., 2015. A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. Eng. Appl. Artif. Intell. 37, 368–377. https://doi.org/10.1016/j.engappai.2014.09.019
Tanuwijaya, S., Alamsyah, A., Ariyanti, M., 2021. Mobile Customer Behaviour Predictive Analysis for Targeting Netflix Potential Customer. 2021 9th Int. Conf. Inf. Commun. Technol. ICoICT 2021 348–352. https://doi.org/10.1109/ICoICT52021.2021.9527487
Thomas, J.S., Chen, C., Iacobucci, D., 2022. Email Marketing as a Tool for Strategic Persuasion. J. Interact. Mark. 57, 377–392. https://doi.org/10.1177/10949968221095552
Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. J. R. Stat. Soc. Ser. B 58, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
Trinh, G., Wright, M.J., 2022. Predicting future consumer purchases in grocery retailing with the condensed Poisson lognormal model. J. Retail. Consum. Serv. 64, 102809. https://doi.org/10.1016/j.jretconser.2021.102809
Van den Poel, D., Larivière, B., 2004. Customer attrition analysis for financial services using proportional hazard models. Eur. J. Oper. Res. 157, 196–217. https://doi.org/10.1016/S0377-2217(03)00069-9
Verbraken, T., Bravo, C., Weber, R., Baesens, B., 2014. Development and application of consumer credit scoring models using profit-based classification measures. Eur. J. Oper. Res. 238, 505–513. https://doi.org/10.1016/j.ejor.2014.04.001

Verbraken, T., Verbeke, W., Baesens, B., 2013. A novel profit maximizing metric for measuring classification performance of customer churn prediction models. IEEE Trans. Knowl. Data Eng. 25, 961–973. https://doi.org/10.1109/TKDE.2012.50

von Helversen, B., Abramczuk, K., Kopeć, W., Nielek, R., 2018. Influence of consumer reviews on online purchasing decisions in older and younger adults. Decis. Support Syst. 113, 1–10. https://doi.org/10.1016/j.dss.2018.05.006

Wen, Y.T., Yeh, P.W., Tsai, T.H., Peng, W.C., Shuai, H.H., 2018. Customer purchase behavior prediction from payment datasets. WSDM 2018 - Proc. 11th ACM Int. Conf. Web Search Data Min. 2018-Febua, 628–636. https://doi.org/10.1145/3159652.3159707

Wu, J., Padgett, D., 2004. A direct comparative framework of customer satisfaction: An application to Internet search engines. J. Interact. Mark. 18, 32–50. https://doi.org/10.1002/dir.20004

Xia, Y., Zhao, J., He, L., Li, Y., Niu, M., 2020. A novel tree-based dynamic heterogeneous ensemble method for credit scoring. Expert Syst. Appl. https://doi.org/10.1016/j.eswa.2020.113615

Xu, Z., Dang, Y., Wang, Q., 2022. Potential buyer identification and purchase likelihood quantification by mining user-generated content on social media. Expert Syst. Appl. 187, 115899. https://doi.org/10.1016/j.eswa.2021.115899

YAN, S., 2019. Building a Customer Identification Model with SPSS. DEStech Trans. Soc. Sci. Educ. Hum. Sci. 117–121. https://doi.org/10.12783/dtssehs/ermis2019/29907

Yeo, J., Hwang, S.W., Kim, S., Koh, E., Lipka, N., 2020. Conversion Prediction from Clickstream: Modeling Market Prediction and Customer Predictability. IEEE Trans. Knowl. Data Eng. 32, 246–259. https://doi.org/10.1109/TKDE.2018.2884467

Zhang, Y., 2021. Prediction of Customer Propensity Based on Machine Learning. Proc. - 2021 Asia-Pacific Conf. Commun. Technol. Comput. Sci. ACCTCS 2021 5–9. https://doi.org/10.1109/ACCTCS52002.2021.00009

Zhou, J., Wang, C., Ren, F., Chen, G., 2021. Inferring multi-stage risk for online consumer credit services: An integrated scheme using data augmentation and model enhancement. Decis. Support Syst. 149, 113611. https://doi.org/10.1016/j.dss.2021.113611