Predictive Modeling in Marketing: Ensemble Methods for Response Modeling

Gabriela Alves Werb and Martin Schmidberger

Abstract: Ensemble methods have received a great deal of attention in the past years in several disciplines. One reason for their popularity is their ability to model complex relationships in large volumes of data, providing performance improvements compared to traditional methods. In this article, we implement and assess ensemble methods' performance on a critical predictive modeling problem in marketing: predicting cross-buying behavior. The best performing model, a random forest, manages to identify 73.3 % of the cross-buyers in the holdout data while maintaining an accuracy of 72.5 %. Despite its superior performance, researchers and practitioners frequently mention the difficulty in interpreting a random forest model's results as a substantial barrier to its implementation. We address this problem by demonstrating the usage of interpretability methods to: (i) outline the most influential variables in the model; (ii) investigate the average size and direction of their marginal effects; (iii) investigate the heterogeneity of their marginal effects; and (iv) understand predictions for individual customers. This approach enables researchers and practitioners to leverage the superior performance of ensemble methods to support data-driven decisions without sacrificing the interpretability of their results.

Keywords: machine learning, predictive modeling, interpretable machine learning, ensemble methods, customer response, response modeling

Prädiktive Modellierung im Marketing: Ensemble-Methoden zur Modellierung von Kundenreaktionen

Zusammenfassung: Ensemble-Methoden haben in vielen Disziplinen große Popularität erlangt. Sie zeichnen sich durch ihre Fähigkeit aus, komplexe Beziehungen in großen Datenmengen zu modellieren. Dies führt typischerweise dazu, dass sie im Vergleich zu herkömmlichen Methoden genauere Prognosen erzielen. Dieser Beitrag implementiert Ensemble-Methoden und bewertet ihre Prognosefähigkeit in Bezug auf eine wichtige Marketingfragestellung: die Vorhersage der Kundennachfrage nach zusätzlichen Produkten („Cross-Buying"). Das Modell mit der besten Prognoseleistung, ein „Random-Forest", identifiziert 73,3% der Cross-Buyer in den Holdout-Daten und erreicht eine Genauigkeit von 72,5%. Trotz seiner überlegenen Prognoseleistung stellen seine schwer interpretierbaren Ergebnisse noch eine hohe Hürde für seine Einführung in die Marketingforschung und Marketingpraxis dar. Um dies anzugehen, verwenden wir Interpretierbarkeitsmethoden um: (i) die einflussreichsten Variablen im Modell darzustellen; (ii) die durchschnittliche

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Größe und Richtung dieser Effekte zu visualisieren; (iii) die Heterogenität dieser Effekte zu untersuchen; und (iv) die Vorhersagen für einzelne Kunden zu verstehen. Dieser Ansatz ermöglicht Forschern und Praktikern, die überlegene Prognoseleistung von Ensemble-Methoden zu nutzen, um datenbasierte Entscheidungen zu unterstützen, ohne die Interpretierbarkeit ihrer Ergebnisse zu beeinträchtigen.

Stichwörter: maschinelles Lernen, prädiktive Modellierung, interpretierbares maschinelles Lernen, Ensemble-Methoden, Kundenreaktion, Response-Modellierung

1. Introduction

Predictive modeling has established itself as a cornerstone of business decision-making. With the increasing demand for data-driven decisions and the widespread collection of fine-grained customer and transaction data, predictive modeling has acquired a central role in marketing. Possible applications include sales forecasting, churn prevention, or customer response modeling, among several others.

In some cases, one might be only interested in obtaining accurate predictions. However, most applications in marketing also require understanding the resulting model. Besides generating new insights, understanding the model's inner workings is crucial to assess their face validity and detect problems early on. This aspect is also essential if the model will support managerial decisions, as its underlying logic often needs to be explained to stakeholders. Consequently, choosing a suitable predictive modeling method usually requires a balance between the resulting model's performance and interpretability (*Shmueli/Koppius*, 2011).

Even though practitioners could benefit from using more complex machine learning methods, they frequently mention the difficulty in explaining results and the risk of bias as current barriers to implementing them (*Bughin et al.*, 2018). As a result, many of them still rely on regression-based methods, as one expert in the field vividly summarized on Twitter: "When you're fundraising, it's AI [artificial intelligence]. When you're hiring, it's ML [machine learning]; when you're implementing, it's linear regression." (*Schwarz*, 2019). The *posthoc* interpretation of models resulting from ensemble methods, a class of machine learning methods, is particularly cumbersome because they consist of several weak models (typically hundreds).

Against this background, the goal of this article is to demonstrate how researchers and practitioners can leverage the superior performance of ensemble methods without needing to sacrifice the interpretability of their results. We do so by tackling the problem of identifying customers who are more likely to cross-buy. Because cross-buying behavior has a substantial impact on firm performance (*Reinartz/Kumar*, 2002), this is an important response modeling problem in marketing.

We find that ensemble methods consistently improve the predictive performance of predictive models. The best performing model, a random forest, improved all performance measures in the holdout data compared to a logistic regression or a single classification tree. But true to its reputation of being a "black box", this model is difficult to interpret on its own, as ensembles do not indicate the size nor direction of the effects.

To address this issue, we demonstrate how researchers and practitioners can use four different interpretability methods to perform a thorough *posthoc* analysis of the model.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Thereby, we provide valuable insights about the relationships uncovered by a highly complex model and explain the predictions it generates for selected customers.

The remaining of this article is organized as follows. To help put our contributions into perspective, we first outline previous literature's main findings on cross-buying prediction and customer response modeling with ensemble methods. Then, we describe our methodological approach. Next, we present our empirical analysis of the demographic and transaction data of 100,000 customers from a collaborating bank in Germany and offer implications for researchers and practitioners.

2. Related Literature

Cross-buying refers to customers buying additional products from the same firm – for example, when a customer starts a relationship with a bank, opening a savings account and one year later opens a checking account. Some studies on cross-buying behavior focused on predicting which customers are more likely to cross-buy (*Kamakura et al.*, 1991; *Verhoef et al.*, 2001), while others addressed the question of which product firms should offer them next (*Knott et al.*, 2002; *Li et al.*, 2005). Interestingly, *Knott et al.* (2002) and *Li et al.* (2005) also show that customers in the financial services industry often purchase products in a natural sequence as they progress in different financial maturity stages.

Relevant Literature	Data	Modeling Method	Main Findings
Kamakura et al. (1991)	Financial panel, 1,517 individuals	Multinomial regressions	Positive effect of income, age, education, and employment status on financial maturity.
Verhoef et al. (2001)	Insurance company, 2,018 customers	Ordered probit regression	Positive effect of direct mail, loyalty program, and price fairness relative to competitors. No main effect of satisfaction.
Knott et al. (2002)	Retail bank, 270,842 customers	Logistic and multinomial regression, neural networks, among others	Current product ownership followed by customer value and demographics as most important predictors of next-product-to-buy.
Li et al. (2005)	Retail bank, 1,201 households	Multivariate probit regression (hierarchical Bayesian framework)	Positive effect of cumulative purchases, cumulative balance, and satisfaction. Stronger effect for higher education level, gender (male), and higher income.
Kumar et al. (2008)	Retailer, 3,000 observations from customer cohort	Random coefficient regression	Positive effect of mailing and cross-selling efforts (up to a threshold). Inverted U-shaped relationship for age and income.
Mende et al. (2013)	Survey, 1,199 insurance customers	Multinomial logistic regression	Positive effect of income and preference for closeness. Negative effect of age and attachment anxiety.

Table 1: Comparison of Empirical Studies on Cross-Buying Behavior

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Most studies find a significant effect of a firm's marketing efforts, customer demographics, perception of price fairness, and previous transaction behavior on cross-buying behavior. Customer attachment styles, preference for closeness, and customer acquisition channels are also important cross-buy drivers (*Mende et al.*, 2013). As Table 1 shows, most studies explaining cross-buying behavior use regression-based methods; (multinomial) logistic regression is the most popular one.

Nevertheless, the increasing popularity of machine learning methods prompted further studies investigating their applicability and predictive performance in marketing. Most of them find that models built with ensemble methods consistently outperform single models (whether regression, neural networks, or trees).

As a result, ensembles have been applied to predict customer churn and the profitability of customer retention campaigns (e.g., *Larivière/Van den Poel*, 2005; *De Bock/Poel*, 2011; *Lemmens/Gupta*, 2020). However, somewhat surprisingly, the same tendency was not observed in studies that address cross-buying behavior. As Table 2 shows, a limited number of studies rely on ensemble methods to predict cross-buy, or more generally, customer response.

Selected Literature	Data	Target Variable	Methods	Interpretation Methods
Kim/Street (2004)	Insurance industry, 9,822 households	Customer response	Ensemble of neural networks with genetic algorithms for feature selection	List of important variables
Ha et al. (2005)	Retailer, 20,300 customers	Customer response	Ensemble of neural networks	None
Larivière/ Van den Poel (2005)	Financial services, 100,000 customers	Next-buy, churn, profit drop, profit evolution	Random forests, logistic and linear regression	Variable importance, descriptive statistics
Prinzie/ Van den Poel (2008)	Retailer, 74,386 customers	Next-product-to-buy	Random forests and ensembles of multinomial logistic regressions	Variable importance
Lessmann et al. (2021)	25 data sets from different sources	Churn, customer response, and profitability	Random forest and stochastic gradient boosting (among others)	None

Table 2: Comparison of Empirical Studies Using Ensemble Methods to Predict Customer Response

We believe that this sparse adoption is related to the subsequent difficulty in interpreting the aggregated model. For example, *Ha et al.* (2005) conclude their study by outlining the impossibility of understanding how each variable impacts the predicted response probabilities or how they interact with each other in their model. In a recent study, *Lessmann et al.* (2021) also stress the importance of clarifying how variables influence predictions in "black-box" models.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Some studies attempt to address this shortcoming by reporting variable importance measures or descriptive statistics. However, to the best of our knowledge, no predictive modeling study in marketing has yet explained the size and direction of effects or provided insights on individual predictions from ensemble models.

Therefore, we contribute to the literature on predictive models in marketing by: (i) implementing and assessing the performance of ensemble methods to predict cross-buying behavior, (ii) explaining the size and direction of the underlying effects captured by the ensemble model, and (iii) explaining individual predictions of the ensemble model.

3. Methodology

This section discusses the methods used in this study to model cross-buying behavior. We start by describing classification trees, which will be the base model for the ensembles. Then, we briefly introduce ensemble methods and describe those used in this study. Next, we present the evaluation criteria we use to assess their predictive performance and the interpretability methods used to interpret their results.

3.1. Classification Trees

Classification trees recursively partition the data into smaller subsets to predict a categorical dependent variable. They start with a *root node* containing all the observations in the data. The trees then split the observations into two or more subsets (*child nodes*) so that the proportion of the dependent variable becomes more homogeneous (*Breiman et al.*, 1984). Nodes that split observations are *decision nodes*, and nodes that do not lead to further splits are *terminal nodes* or *leaves*.

Classification trees can base their split decisions on significance tests or node *purity* measures, such as the Gini Index (*Gini*, 1912). In the context of splitting decisions, the Gini Index measures the homogeneity (or *purity*) of the observations in a node concerning the dependent variable. The tree looks for a split at each decision node that maximizes the decrease in node impurity, which implies minimizing the impurity in the child nodes. Due to their top-down "*divide and conquer*" approach, classification trees are relatively fast, even with large datasets. However, single classification trees are prone to overfitting, so it is crucial to avoid that the tree grows to its maximum, a procedure known as *pruning*.

In our empirical application, we perform a random search over three *pruning* parameters: the complexity parameter, the minimum number of observations in a node for a split to be pursued, and the maximum tree depth. Another way of avoiding overfitting while keeping most of the advantages inherent to classification trees is to use ensemble methods, as we discuss in the following paragraphs.

3.2. Ensemble Methods

The idea behind ensemble models is to combine many "weak" models to produce a more potent and stable model. For example, if we were to predict a game score, our own opinion might only be a rough approximation. However, combining the opinion of hundreds of different people will likely result in a much more accurate prediction.

Most ensembles found in the machine learning literature are based on *bagging* and *boosting*. The term *bagging* stands for *bootstrap aggregating* and was proposed by *Breiman* (1996) to improve accuracy and reduce instability in tree predictions. *Bagging*

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

artificially generates "new" training datasets by generating bootstrap replicates of the original dataset. The aggregated predictions over all the replicate datasets typically result in a higher predictive performance, as the model learns from several "perturbed versions" of the dataset.

Boosting was proposed by *Freund/Schapire* (1997) to make the models *adapt* and focus on observations that are more difficult to predict. *Boosting* also combines a series of "weak" models, but it builds the models sequentially. At first, all observations receive the same weight. After building the first model, misclassified observations receive a higher weight, so each subsequent model progressively focuses on predicting the observations with higher weights. For the ensemble prediction, each model's vote depends on its accuracy on the weighted training dataset.

A significant advantage of ensemble methods is their ability to reduce the variance of a model. Results obtained by *Bauer/Kohavi* (1999) suggest that *boosting* methods are effective in reducing both variance and bias of a tree model. In the following, we will discuss the two ensemble methods used in this study, random forests and gradient boosting, which build upon *bagging* and *boosting*, respectively.

3.2.1. Random Forest

As proposed by *Breiman* (2001), random forests are ensembles of trees that introduce randomization in two stages. The first one occurs at the observation level by training each tree on a bootstrap replicate of the training data, such as in *bagging*. In contrast, the second randomization occurs at the variable level, when only a random subsample of the explanatory variables is considered to search for the best split at each node. Thereby, random forests reduce the trees' correlation by forcing them to split on different explanatory variables. Such as in *bagging*, the final prediction is aggregated for all trees through averaging or majority voting. *Breiman* (2001) suggests setting the random subsample m of the explanatory variables considered at each node as the square root of the number of explanatory variables in the data, expressed by p. Nevertheless, the optimal value can vary in each setting.

A nice aspect of random forests is that they estimate the generalization or out-of-bag (OOB) error, which indicates how the model would perform on previously unseen data. They do so using the observations that are part of the original training dataset but were not part of the bootstrapped subsample used to train a given tree, the OOB observations. These typically amount to one-third of the observations in the original training dataset.

In our empirical application, we perform a random search over the following random forest parameters: the number of trees, the number of explanatory variables randomly selected for consideration at each split (m), the minimum number of observations in a terminal node, and the maximum number of terminal nodes.

3.2.2. Gradient Boosting

Freund/Schapire (1997) originally described *boosting* as a method that optimizes a loss function. Moving further in this direction, *Friedman* (2001) proposes to use gradient descent to optimize the loss function of a boosted model, the gradient boosting machine (GBM). As gradient boosting fits many trees consecutively, overfitting can become a problem.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Therefore, one of its essential inputs is the learning rate, which determines how slow the aggregated model learns from each tree. A low learning rate reduces a single tree's impact in the aggregated model, so the GBM takes many small steps towards the final prediction. However, there is a trade-off between the learning rate and the convergence speed because a slowly learning model requires a larger number of trees to achieve a good predictive performance. Moreover, unlike in random forests, trees in gradient boosting are usually small and often are simply stumps, i.e., trees with only one split.

In our empirical application, we perform a random search over the following gradient boosting parameters: the number of trees, the learning rate, the maximal tree depth, the minimum sum of weights in a terminal node, and lambda, a parameter for L2-regularization on the leaf weights.

3.3. Performance Evaluation Criteria

To evaluate the models' performance on the validation data, we use four criteria: recall, the area under the precision-recall curve (AUC-PR), F1-score, and accuracy. To ensure comparability with previous studies, we also compare their performance with that of a logistic regression.

Before we explain the motivation for these criteria, let us first consider the confusion matrix for the binary prediction problem of cross-buy prediction (Table 3). The cells of the confusion matrix provide the basis to compute the following measures: accuracy ($(TN + TP)/Total$), recall or sensitivity ($TP/(TP + FN)$), precision ($TP/(TP + FP)$), and specificity ($TN/(TN + FP)$).

Prediction \ Actual	No Cross-Buy	Cross-Buy	Sum
No Cross-Buy	True negative (TN)	False negative (FN)	TN + FN
Cross-Buy	False positive (FP)	True positive (TP)	FP + TP
Sum	TN + FP	FN + TP	Total

Table 3: Confusion Matrix for Cross-Buy Prediction

While accuracy is a widespread evaluation criterion, it has the shortcoming of giving equal importance to predicting positive and negative cases, which is problematic in imbalanced data sets. For example, if 90 % of the customers do not cross-buy, a model that guesses that no customer cross-buys has an accuracy of 90 %, but it would still be useless to identify cross-buyers. Similarly, analyses based on the receiver operating characteristics (ROC) curve are misleading for imbalanced data sets.

Therefore, we focus on evaluation criteria that are adequate for handling imbalanced data sets and reflect our primary interest in distinguishing cross-buyers. We start with recall, which denotes the share of true cross-buyers that the model correctly identifies. We also rely on the precision-recall curve, which plots precision against recall for all potential decision thresholds, outlining the performance trade-offs involved in selecting an appropriate threshold. The decision threshold translates predicted probabilities into a class prediction. Therefore, changing it can substantially affect the model's performance and implies a trade-off between decreasing one type of error at the other's expense.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Our third measure, the F1-score, is the harmonic mean of precision and recall. It is expressed as $(2*(precision*recall)/(precision + recall))$, so it considers the compromise between false positives and false negatives. For example, a model that predicts all customers to be cross-buyers will have a perfect recall but a poor precision, which will influence the F1-score. We also report accuracy to enable a comparison with the predictive performance reported in previous cross-buy studies.

3.4. Interpretability Methods

Despite their superior performance, models generated using ensemble methods are challenging to interpret. They do not provide effect sizes as in a regression nor a simple graphical representation of their structure, as in a classification tree. However, the machine learning literature has flourished over the past years with several proposed methods to shed light on "black-box" models. Table 4 provides an overview of the most popular interpretability methods[1] described in more detail in this section.

Interpretation Question	Method	Literature
What are the most influential explanatory variables in the model?	Variable Importance Measures (VIM)	*Breiman et al.* (1984), *Breiman* (2001)
What is the average size and direction of the effect of an explanatory variable?	Partial Dependence Plots (PDP)	*Friedman* (2001)
How heterogeneous is the effect of an explanatory variable?	Individual Conditional Expectation Plots (ICE)	*Goldstein et al.* (2015)
Why was a particular customer predicted as a (non)cross-buyer?	Local Interpretable Model-Agnostic Explanations (LIME)	*Ribeiro et al.* (2016)

Table 4: Overview of the Interpretability Methods Used in This Study

Variable Importance Measures (VIM) are among the first methods proposed to interpret highly complex models. They compute a numerical measure that indicates how much each explanatory variable influences the model predictions. One way to calculate them is to consider the decrease in node impurity that follows from a split on the variable and sum it within a tree and across all trees for ensembles. These measures are known as impurity-based VIM.

Another possibility is to consider the decrease in predictive accuracy resulting from a permutation of the explanatory variable. If the variable is essential in the model, permuting its values should increase the prediction error. Such measures are known as permutation-based VIM. In random forests, this approach is implemented using the out-of-bag (OOB) observations in each tree (*Breiman*, 2001).

Even though variable importance measures are a starting point to identify influential variables, they provide no insight into the effects' size or direction. *Friedman* (2001) closes this gap with Partial Dependence Plots (PDP), which show how the model predictions

1 SHapley Additive exPlanations (SHAP) (*Lundberg/Lee*, 2017) are also a popular interpretability method. For the sake of brevity, we do not discuss it in our study.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

change for different values of a given variable. Because PDP keep all other variables at their mean values in the sample, we can interpret them as a depiction of the average marginal effect of the respective variable in the model. Consequently, PDP can be misleading whenever the effects are highly heterogeneous across the observations.

To address this problem, *Goldstein et al.* (2015) propose Individual Conditional Expectation Plots (ICE). ICE are rooted in a similar principle as PDP, but instead of estimating the average marginal effect, they estimate an explanatory variable's effect for each observation. To do so, ICE vary the value of the given explanatory variable to compute the model's predictions for each observation but keeping the other variables at their actual values. Therefore, ICE plots have as many paths as the number of observations in the data set.

The ICE paths' average corresponds precisely to the PDP, so visualizing both methods' results indicates whether the average marginal effects provided by the PDP are a good approximation for the entire data set. Nevertheless, both PDP and ICE may give misleading results in the presence of strong interaction effects because they assume that the variable of interest is independent of the other ones.

Overall, these methods provide insights into the size and influence of different explanatory variables in the model. However, we might also want to know why the model predicts specific customers to have a high or low cross-buying probability. Local Interpretable Model-Agnostic Explanations (LIME) (*Ribeiro et al.*, 2016) tackle this challenge by translating the complex decisions of a "black-box" model into a set of interpretable rules.

To do so, LIME fit a surrogate model to explain the predictions of the complex model. They create a synthetic version of the training data set through small changes (e.g., add one year to a customer's age) and generate predictions for these data using the complex model. Based on these data, LIME use a simple model to interpret the complex model's decisions locally. The complex model's predictions on the perturbed data become the dependent variable for this simpler model, while the distances between the actual and the perturbed observations become weights in the model.

Our empirical application relies on these interpretability methods to uncover the model's underlying relationships and understand individual predictions.

4. Empirical Application

We conduct our empirical study in collaboration with one of the largest financial services providers in Germany. As in other industries, customers usually exhibit sequential buying patterns when they progressively enhance their relationship with their financial services provider. Therefore, cross-buying is an important step to strengthen the customers' relationship with the firm.

Li et al. (2005) show that checking accounts play a central role in the relationship between customers and their financial services providers since they tend to "open the door" to acquire more advanced financial services. Against this background, our collaborating firm would like to run a cross-sell campaign and incentivize customers who do not yet have a checking account to open one. Thereby, the firm must concentrate its marketing efforts on customers who are more likely to cross-buy and open a checking account.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

4.1. Data Sampling and Description

Because our goal is to predict customers' likelihood of opening a checking account in the future, we first identify all customers of the collaborating firm who do not yet own a checking account. Based on this cross-section, we obtain a random sample of 10,000 customers who opened a checking account within the follow-up period of six months and 90,000 who did not. Figure 1 outlines this sampling strategy.

The empirical data set with 100,000 customers contains 34 explanatory variables that describe customer demographics, transaction behavior, and the firm's marketing efforts. We collect these at the initial period (T_0), so the customer's cross-buy decision in the follow-up period cannot influence them. The dependent variable cross-buy is binary and receives the value one for customers who opened a checking account within six months and zero otherwise. A description of all the variables in the data set is available in the Online Appendix A.

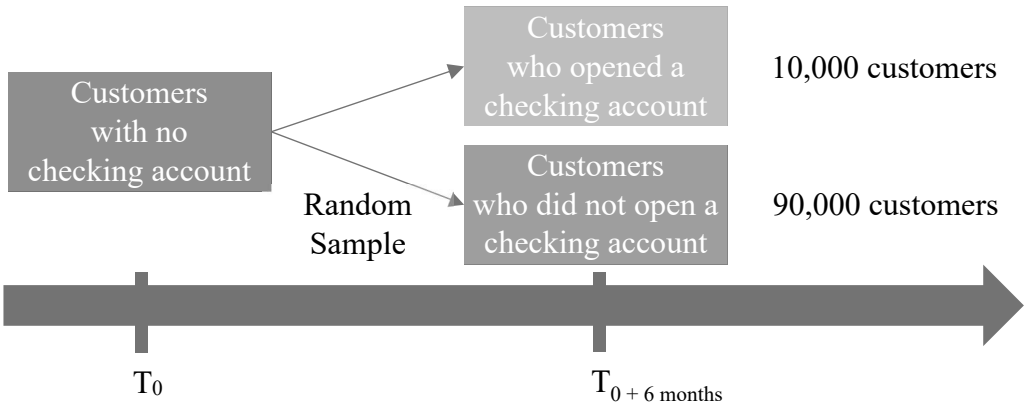

Figure 1: Data Sampling Strategy

We randomly split this data set into two parts and use 80 % of the observations as training data and 20 % as test data. Because many studies find that class imbalance affects predictive models' performance (*Prinzie/Van den Poel*, 2008), we follow previous literature and undersample the non-cross-buyers in the training data until we achieve a 50 %-50 % split. However, we do not change the class imbalance in the test data to avoid overly optimistic model performance measures[2].

Because the random forest implementation used does not support missing values, we imputed those using random forest proximities, as described by *Breiman/Cutler* (2004). We implement this procedure using five runs, each with 500 trees. We use the imputed train and test datasets for the random forest models and the logistic regression. As a robustness check, we also impute the train and dataset using the median values for numeric variables and the most frequent class for categorical variables.

2 We thank an anonymous reviewer for this suggestion.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

4.2. Model Building

We build the models using the discussed methods in the open-source software R. For the classification trees, we use the package *rpart* (*Therneau et al.*, 2019), which implements the CART algorithm (*Breiman et al.*, 1984) with the Gini Index as the splitting criterion. Surrogate variables handle the missing values. We tune the following hyperparameters: complexity parameter (cp), the minimum number of observations in a node (minsplit), and the maximum tree depth (maxdepth).

For the random forests, we use the package developed by *Liaw/Wiener* (2002) and tune the following hyperparameters: the number of trees (ntree), the number of explanatory variables randomly selected at each split (mtry), the minimum number of observations in a terminal node (nodesize), and the maximum number of terminal nodes (maxnodes).

Finally, we implement the boosting framework proposed by *Chen/Guestrin* (2016) and implemented in R by *Chen et al.* (2020). We tune the following hyperparameters: the number of trees (nrounds), the tree's maximal depth (max_depth), the minimum sum of weights in a terminal node (min_child_weight), the learning rate (eta), and the parameters for L2-regularization on the leaf weights (lambda).

For all methods, we perform a simultaneous[3] parameter optimization with 10-fold cross-validation. We define a grid of values for each parameter, as outlined in Table 5. To speed up the tuning process, we use a random search algorithm that randomly chooses 50 parameter combinations from those defined in our hyperparameter space.

Classification Tree	Random Forest	Extreme Gradient Boosting
cp: [0.0005, 0.015]	*mtry:* [6, 30]	*lambda:* [0.1, 1]
minsplit: [10, 50]	*ntree:* [100, 1000]	*eta:* [0.05, 0.3]
maxdepth: [5, 30]	*nodesize:* [5, 30]	*max.depth:* [1, 6]
-	*maxnodes:* [5, 500]	*nrounds* [100, 1000]
-	-	*min_child_weight:* [1, 10]

Table 5: Intervals of the Hyperparameters for Tuning the Classification Tree, Random Forest, and Extreme Gradient Boosting Models

As the cross-buy response is highly imbalanced, we use the AUC-PR to select the best hyperparameter combination for each method based on the estimates cross-validated during the model training. Then, we evaluate these models' performance in the test fold during the cross-validation and in the test dataset using four criteria: recall, the area under the precision-recall curve (AUC-PR), F1-score, and accuracy. Finally, we estimate a logistic regression using the imputed training dataset to make our results comparable with previous cross-buying behavior studies.

3 An alternative is to optimize each parameter separately. However, as there is a trade-off among different parameters, doing so does not guarantee that their combination will deliver an optimal performance.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

5. Results

5.1. Model Performance Evaluation

For the classification tree, the model with the highest average AUC-PR (71.1 %) in the test folds during the cross-validation has a complexity parameter of approximately 0.001, a minimum of 28 observations in a node, and 12 as maximum tree depth.

In contrast, the random forest model with the highest average AUC-PR (74.6 %) in the test folds shows a slight but relevant improvement relative to a single classification tree. This model has 556 trees, 11 explanatory variables randomly selected at each split, a minimum of 29 observations in a terminal node, and a maximum of 464 terminal nodes. Figure 2 shows that a higher number of trees generally leads to marginally better performance, as long as the tree complexity remains conservative.

Furthermore, a gradient boosting model with the following parameters had the highest AUC-PR (75.4 %) in the test folds: 160 trees, a learning rate of 0.07, a maximal depth of 3, a minimum sum of weights in a terminal node of 3.31, and a L2-regularization parameter of 0.53. Despite the surprisingly low number of trees, this model performs well because the other parameters effectively regularize it and avoid unnecessary splits.

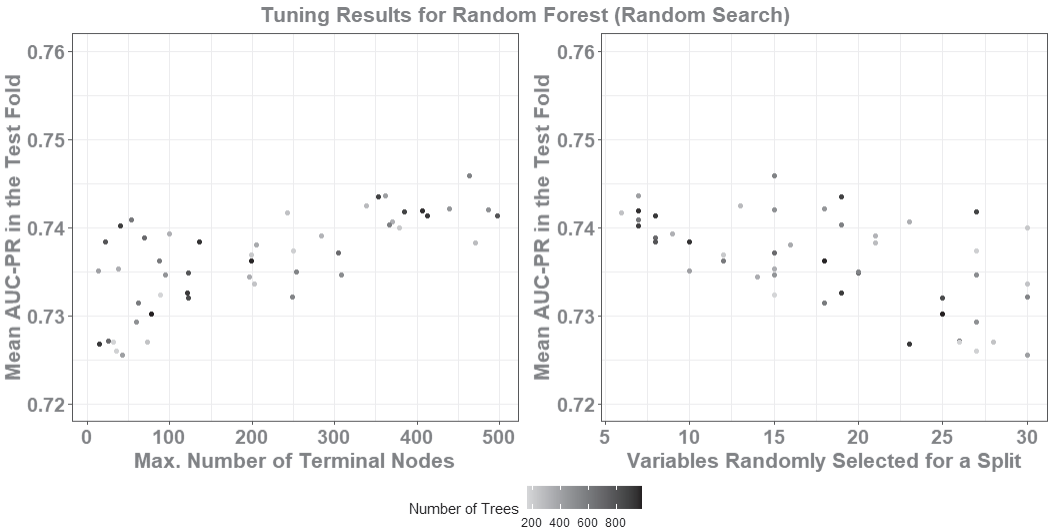

Figure 2: Mean Area Under the Precision-Recall Curve (AUC-PR) in the Test Folds for Selected Hyperparameter Combinations of the Random Forest

Our predictions consider a cutoff threshold of 0.5. Therefore, we classify all customers with a predicted cross-buy probability of 0.5 or higher as cross-buyers. While a different threshold could deliver better results, we find that this improvement does not generalize beyond the training data set. Table 6 shows the average performance of the highest AUC-PR models in the test folds (cross-validation in the training phase), which should provide a fair indication of their expected performance on new data. We also benchmark these models with a logistic regression, including all explanatory variables.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Average Cross-Validated Measures (Test Fold in the Training Phase)				
Model	**AUC-PR**	**Recall**	**F1 Score**	**Accuracy**
Logistic Regression	71.4 %	65.3 %	66.3 %	66.8 %
Classification Tree	71.1 %	65.5 %	67.0 %	67.8 %
Random Forest	74.6 %	65.2 %	67.5 %	68.6 %
Gradient Boosting	75.4 %	66.9 %	68.1 %	68.7 %

Table 6: Comparison of the Average Performance Measures of the Best Performing Models in the Test Folds

Despite having a slight advantage on the AUC-PR, the random forest and the gradient boosting model seem to have a similar performance across the other criteria in the test folds when compared to a single classification tree or a logistic regression. However, Table 7 shows that the performance differences become more substantial in the holdout data (test dataset with 20 % of the observations). As the test data set has a more realistic distribution of the dependent variable (10 % cross-buyers – 90 % non-cross-buyers), there is a substantial decrease in some performance measures[4].

	Test Data Set Measures			
Model	**AUC-PR**	**Recall**	**F1 Score**	**Accuracy**
Logistic Regression	25.1 %	66.2 %	29.0 %	68.1 %
Best Classification Tree	23.6 %	67.4 %	31.1 %	70.7 %
Best Random Forest	**38.9 %**	**73.3 %**	**34.4 %**	**72.5 %**
Best Gradient Boosting	30.4 %	69.0 %	31.8 %	70.9 %

Table 7: Comparison of the Performance of the Best Performing Models in the Test Data Set

Overall, the models can identify a significant share of true cross-buyers. The performance of the logistic regression is similar to previous cross-buy and next-buy studies. For example, *Knott et al.* (2002) find an accuracy ranging from 38.3 % to 55.1 % for different model specifications, whereas *Kumar et al.* (2008) find a 71 % accuracy in the holdout sample. On the other hand, *Larivière/Van den Poel* (2005) obtain a 74.5 % accuracy with the logistic regression. In their case, random forests also show a slight but consistently higher performance.

4 We note that the choice of imputation method does not influence the qualitative results in Table 6 and Table 7. In all cases, the ordering of the models in all performance measures remains unaffected.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

5.2. Model Interpretation

5.2.1. Most Influential Explanatory Variables in the Model

We compute the permutation variable importance measures (VIM) for our tuned random forest using 50 Monte Carlo iterations and resampling with replacement. Figure 3 displays the ten most important variables.

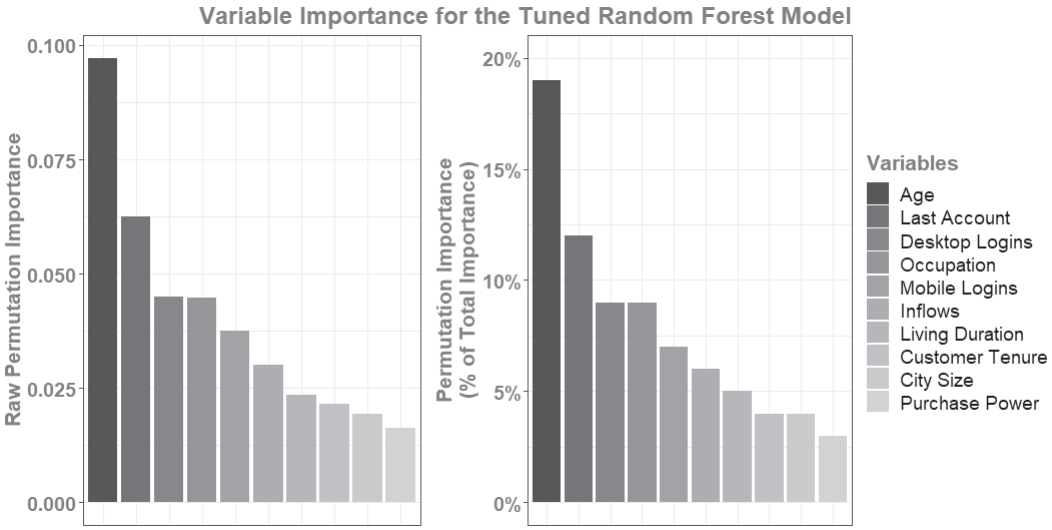

Figure 3: Raw and Percentage Permutation Variable Importance Measures for the Tuned Random Forest Model

Age and occupation are among the most important variables in the model. The purchase power in a customer's neighborhood, a proxy for income, also appears among the ten most important variables. These results corroborate previous findings that demographic characteristics are important predictors of cross-buying behavior (e.g., *Kamakura et al.*, 1991; *Knott et al.*, 2002; *Kumar et al.*, 2008; *Mende et al.*, 2013).

Furthermore, the number of days since the customer opened another type of account, the number of logins from a desktop and a mobile device in the previous six months, and the volume of inflows into saving accounts also contribute to the model. Previous studies also show that recency, frequency, and monetary value (RFM) variables influence customers' propensity to cross-buy (*Knott et al.*, 2002; *Li et al.*, 2005).

Interestingly, we find additional characteristics associated with cross-buying behavior: the city's size in which the customer lives and the average number of years people live in the customer's residential building. To the best of our knowledge, these effects do not appear in previous literature. They provide novel insights, as well as an interesting avenue for future research.

5.2.2. Average Size, Direction, and Heterogeneity of the Effects

Even though the VIM provide a first glance at the most influential variables in the random forest model, they do not indicate the direction or the size of the effects. Therefore, we

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

resort to Partial Dependence Plots (PDP) and Individual Conditional Expectation Plots (ICE) to visualize the average marginal effect and the marginal effect for each observation, respectively. For brevity, we illustrate this approach for the two most important explanatory variables identified by the VIM and display the results in Figure 4. A representation from the third to the tenth most important variable is available in the Online Appendix B.

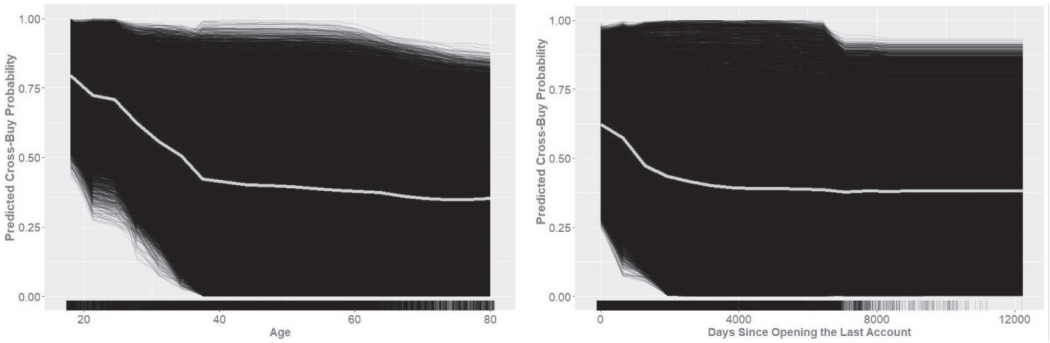

Figure 4: Effect of Age and Days Since Opening the Last Account on the Predicted Cross-Buy Probability

The black lines represent the effects estimated by the ICE for each observation in the data set, while the grey lines represent the average effect estimated by the PDP. The rugged bars below the graphs indicate the number of observations in each region. We find that customers approaching 30 years old experience a substantial decrease in their predicted cross-buying probability. The negative effect stabilizes around 40 years old. These results support previous findings from *Mende et al.* (2013) in the financial industry. Furthermore, an average negative marginal effect of age is also consistent with the framework outlined by *Kamakura et al.* (1991). Interestingly, the effect of age is heterogeneous across different customers, as demonstrated by the black lines.

The number of days since opening the last account has an average negative effect on cross-buy propensity, but it is heterogeneous. For customers with a high cross-buy propensity, it is modest and only present after about 6,000 days. However, for customers with a low predicted cross-buy probability, the first 2,000 days induce an abrupt negative effect, after which the predicted cross-buy probability approaches zero.

The PDP in Figure 5 displays the interaction effect between both variables – young new customers have a higher predicted cross-buy likelihood. Interestingly, we see that after customers older than 35 years old pass a threshold around 2,000 days since opening their last account, they reach and stay at a very low cross-buy probability.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

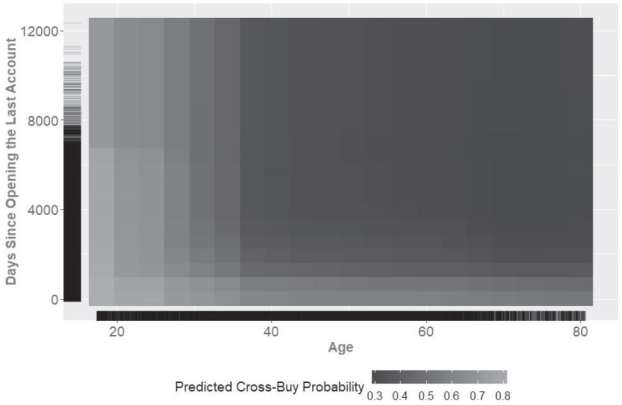

Figure 5: Average Interaction Effect between Age and Days since Opening the Last Account

5.2.3. Explaining Individual Predictions

Understanding individual predictions is a crucial step to assess the model's face validity and uncover possible biases. To illustrate how Local Interpretable Model-agnostic Explanations (LIME) can be used to achieve this goal, we randomly select two actual cross-buyers in the test data, one correctly and the other incorrectly predicted.

Figure 6 indicates that the random forest correctly classified customer 11,739 as a cross-buyer because he is a male with a positive balance between €6,899 and €25,810 across different products and an inflow to his savings accounts between €150 and €3,000 in the previous six months. However, the model incorrectly classified customer 12,421 as a non-cross-buyer because she has a joint bank account, had less than €150 inflows to the savings accounts in the previous six months and opened another type of account between 3,550 and 5,588 days ago.

We perform this analysis for all the cross-buyers in the test data set and find a similar pattern. Figure 7 shows that age, gender, and the number of days since opening another type of account play a crucial role in correctly identifying cross-buyers. In particular, the random forest attributed these customers a high cross-buying likelihood because they are 32 years old or younger, are male, or have opened another type of account less than 570 days ago.

These results indicate that demographic characteristics and RFM variables previously identified in the literature explain the customers' random forest model predictions, suggesting the face validity of the relationships captured by the "black-box" model.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

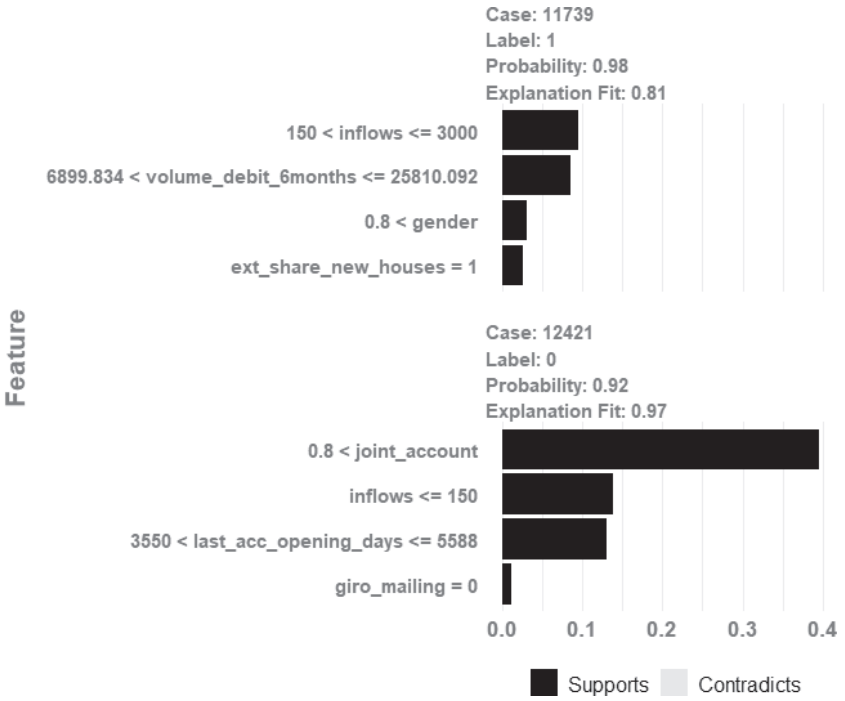

Figure 6: Explanation of the Predictions for Two Customers using Local Interpretable Model-Agnostic Explanations

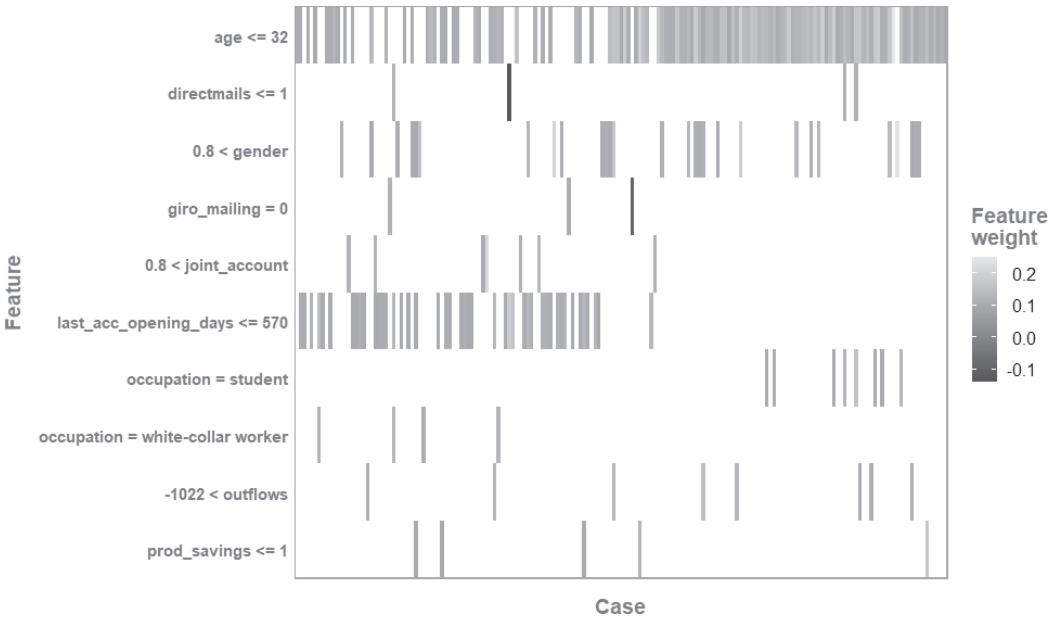

Figure 7: Explanation of the Predictions for Cross-Buyers Correctly Identified by the Model in the Test Data

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Note: For better visualization, this figure includes only the ten most relevant explanations that explain the predictions for at least three cross-buyers and have an absolute minimum weight of 0.1 in the simple model.

6. Discussion

In line with previous studies on customer response (e.g., *Kim/Street*, 2004; *Larivière/Van den Poel*, 2005), our study shows that ensemble methods consistently improve the predictions of cross-buying behavior. A random forest performs better in all evaluation measures in the holdout data than a logistic regression or a classification tree. However, true to its "black box" reputation, this model is difficult to interpret.

Using permutation-based variable importance measures (VIM), we find several characteristics that influence customers' predicted cross-buy propensity in the random forest model and reflect previous findings in the cross-buy literature. Furthermore, the random forest model seems to have uncovered novel relationships. We find that the city's size where the customers live and the average number of years people live in the customers' residential building are also important predictors. Even though their effects are not very substantial in our data (less than 3 % higher or lower propensity to cross-buy), they provide an exciting avenue for future investigations.

The ICE show that the effects of the two most important variables (according to the VIM) are highly heterogeneous across the customers investigated. For some customers, age's negative impact on the propensity to cross-buy is modest and remains mostly stable once they approach 40 years old. However, for most customers, there is a steep decrease in cross-buying probability until they reach the end of their thirties. These results suggest that managers should prioritize customers of the critical age groups in their cross-selling efforts.

This decreasing propensity to cross-buy and open a checking account is consistent with the idea that customers purchase financial services in a natural sequence (*Knott et al.*, 2002; *Li et al.*, 2005). As customers grow older and progress in different financial maturity stages, they become less likely to open a checking account. However, for customers with a high propensity to cross-buy, the ICE uncover that age has an inverted U-shaped effect in the model. *Kumar et al.* (2008) also find an inverted U-shaped effect of age on the propensity to cross-buy in a retail setting.

Furthermore, we find that the predicted propensity to cross-buy decreases substantially with the number of days since customers opened their last account. This effect is in line with previous studies that find a significant effect of recency on cross-buying behavior (*Knott et al.*, 2002; *Li et al.*, 2005). However, we find that this effect is heterogeneous among the analyzed customers. There is a steep negative effect for customers with a low propensity cross-buy: after 2,000 days since opening the last account, their propensity to cross-buy approaches zero. For customers with a high propensity to cross-buy, the effect is less pronounced and emerges much later, after about 6,000 days since opening the last account.

Interestingly, the PDP also show a significant interaction effect between both variables. This finding suggests that the firm managers should not waste resources by targeting customers who are both in their late thirties or older and have opened their last account more than 5,5 years ago, as they have a low probability of cross-buying (on average, only 30 %).

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

In addition, we find that customers who use online banking more often are more likely to cross-buy. The positive relationship is present for both desktop and mobile devices but is much stronger for mobile device logins. Customers with more than 200 logins from a mobile device within six months (slightly more than once a day) have a 30 % higher propensity to cross-buy than customers who do not use online banking. Previous studies also find that mobile channel usage positively affects customers' subsequent purchases in other settings (*Gensler et al.*, 2012; *Steinhoff et al.*, 2019).

Finally, we zoom into the predictions for selected customers to understand why the model predicted them to be cross-buyers. This investigation helps us assess the results' face validity and detect potential biases or inconsistencies in the model.

7. Conclusion

This study demonstrates how managers and researchers can leverage ensemble methods for cross-buying predictions. The best performing model, a random forest, manages to identify 73.3 % of the cross-buyers in the holdout data while maintaining an accuracy of 72.5 %. Therefore, it identifies cross-buyers without mistakenly targeting too many non-cross-buyers, which is crucial for an efficient cross-selling strategy.

We contribute to the marketing literature by demonstrating the benefits of using ensemble methods for predictive models of cross-buying behavior. Furthermore, we address an important shortcoming of these models: the difficulty in interpreting their results. We employ four interpretability methods that enable us to: identify the most influential explanatory variables in the model; assess the average size, the direction, and the heterogeneity of their effects; and explain predictions for individual customers.

Our results should encourage practitioners and researchers to leverage ensemble methods to support their data-driven decisions without the fear of sacrificing the interpretability of their results. While we implement ensemble and interpretability methods to tackle cross-buy prediction, the presented approach applies to any predictive modeling setting.

8. References

Bauer, E./Kohavi, R. (1999): An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants, in: Machine Learning, Vol. 36, No. 1, S. 105-139.

Breiman, L. (1996): Bagging Predictors, in: Machine Learning, Vol. 24, No. 2, S. 123–140.

Breiman, L. (2001): Random Forests, in: Machine Learning, Vol. 45, No. 1, S. 5–32.

Breiman, L./Cutler, A. (2004): Random Forests. www.stat.berkeley.edu/~breiman/RandomForests/ (accessed 2020/06/07).

Breiman, L./Friedman, J. H./Stone, C. J./Olshen, R. A. (1984): Classification and Regression Trees. Belmont, CA.

Bughin, J./Seong, J./Manyika, J./Chui, M./Joshi, R. (2018): Notes from the AI Frontier: Modeling the Impact of AI on the World Economy. www.mckinsey.com (accessed 2020/01/02).

Chen, T./Guestrin, C. (2016): XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, Association for Computing Machinery, S. 785–794.

Chen, T./He, T./Benesty, M./Khotilovich, V./Tang, Y. (2020): xgboost: eXtreme Gradient Boosting, in: R Package Version 0.4 – 2.

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

De Bock, K. W./Poel, D. V. d. (2011): An Empirical Evaluation of Rotation-Based Ensemble Classifiers for Customer Churn Prediction, in: Expert Systems with Applications, Vol. 38, No. 10, S. 12293–12301.

Freund, Y./Schapire, R. E. (1997): A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, in: Journal of Computer and System Sciences, Vol. 55, No. 1, S. 119–139.

Friedman, J. H. (2001): Greedy Function Approximation: A Gradient Boosting Machine, in: Annals of Statistics, Vol. 29, No. 5, S. 1189–1232.

Gensler, S./Leeflang, P./Skiera, B. (2012): Impact of Online Channel Use on Customer Revenues and Costs to Serve: Considering Product Portfolios and Self-Selection, in: International Journal of Research in Marketing, Vol. 29, No. 2, S. 192–201.

Gini, C. (1912): Variabilità e Mutabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche. Bologna.

Goldstein, A./Kapelner, A./Bleich, J./Pitkin, E. (2015): Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation, in: Journal of Computational and Graphical Statistics, Vol. 24, No. 1, S. 44–65.

Ha, K./Cho, S./MacLachlan, D. (2005): Response Models Based on Bagging Neural Networks, in: Journal of Interactive Marketing, Vol. 19, No. 1, S. 17–30.

Kamakura, W. A./Ramaswami, S. N./Srivastava, R. K. (1991): Applying Latent Trait Analysis in the Evaluation of Prospects for Cross-Selling of Financial Services, in: International Journal of Research in Marketing, Vol. 8, No. 4, S. 329–349.

Kim, Y./Street, W. N. (2004): An Intelligent System for Customer Targeting: a Data Mining Approach, in: Decision Support Systems, Vol. 37, No. 2, S. 215–228.

Knott, A./Hayes, A./Neslin, S. A. (2002): Next-Product-to-Buy Models for Cross-Selling Applications, in: Journal of Interactive Marketing, Vol. 16, No. 3, S. 59–75.

Kumar, V./George, M./Pancras, J. (2008): Cross-Buying in Retailing: Drivers and Consequences, in: Journal of Retailing, Vol. 84, No. 1, S. 15–27.

Larivière, B./Van den Poel, D. (2005): Predicting Customer Retention and Profitability by Using Random Forests and Regression Forests Techniques, in: Expert Systems with Applications, Vol. 29, No. 2, S. 472–484.

Lemmens, A./Gupta, S. (2020): Managing Churn to Maximize Profits, in: Marketing Science, Vol. 39, No. 5, S. 956–973.

Lessmann, S./Haupt, J./Coussement, K./De Bock, K. W. (2021): Targeting Customers for Profit: An Ensemble Learning Framework to Support Marketing Decision-Making, in: Information Sciences, Vol. 557, No. 2021, S. 286–301.

Li, S./Sun, B./Wilcox, R. T. (2005): Cross-Selling Sequentially Ordered Products: An Application to Consumer Banking Services, in: Journal of Marketing Research, Vol. 42, No. 2, S. 233–239.

Liaw, A./Wiener, M. (2002): Classification and Regression by randomForest, in: R News, Vol. 2, No. 3, S. 18–22.

Lundberg, S. M./Lee, S.-I. (2017): A Unified Approach to Interpreting Model Predictions, in: Proceedings of the 30th Conference on Advances in Neural Information Processing Systems, S. 4765–4774.

Mende, M./Bolton, R. N./Bitner, M. J. (2013): Decoding Customer–Firm Relationships: How Attachment Styles Help Explain Customers' Preferences for Closeness, Repurchase Intentions, and

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.

Changes in Relationship Breadth, in: Journal of Marketing Research, Vol. 50, No. 1, S. 125–142.

Prinzie, A./Van den Poel, D. (2008): Random Forests for Multiclass Classification: Random MultiNomial Logit, in: Expert Systems with Applications, Vol. 34, No. 3, S. 1721–1732.

Reinartz, W./Kumar, V. (2002): The Mismanagement of Customer Loyalty, in: Harvard Business Review, Vol. 80, No. 7, S. 86–94, 125.

Ribeiro, M. T./Singh, S./Guestrin, C. (2016): Why Should I Trust You? Explaining the Predictions of Any Classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, S. 1135–1144.

Schwarz, B., [@xaprb]. (2019, February 19). When you're fundraising, it's AI. Twitter. https://twitter.com/xaprb.

Shmueli, G./Koppius, O. R. (2011): Predictive Analytics in Information Systems Research, in: MIS Quarterly, Vol. 35, No. 3, S. 553–572.

Steinhoff, L./Arli, D./Weaven, S./Kozlenkova, I. V. (2019): Online Relationship Marketing, in: Journal of the Academy of Marketing Science, Vol. 47, No. 3, S. 369–393.

Therneau, T./Atkinson, B./Ripley, B. (2019): rpart: Recursive Partitioning and Regression Trees, in: R Package Version 4.1 – 15.

Verhoef, P. C./Franses, P. H./Hoekstra, J. C. (2001): The Impact of Satisfaction and Payment Equity on Cross-Buying: A Dynamic Model for a Multi-Service Provider, in: Journal of Retailing, Vol. 77, No. 3, S. 359–378.

Gabriela Alves Werb, Ph.D, ist Professorin für Betriebliche Informationssysteme an der Frankfurt University of Applied Science und Data Science Expertin bei der Deutschen Bank.

Anschrift: Frankfurt University of Applied Sciences, Nibelungenplatz 1, 60318 Frankfurt, Germany, Tel.: +49 (0) 69/1533-2042, E-Mail: gabriela.alveswerb@fb2.fra-uas.de

Martin Schmidberger, Dr., ist Honorarprofessor an der Goethe Universität Frankfurt und Leiter des Bereichs Customer Interactions bei der ING-DiBa AG.

Anschrift: Theodor-Heuss-Allee 2, 60486 Frankfurt am Main, Germany, Tel.: +49 (0) 69/5050-0105, E-Mail: martin.schmidberger@ing-diba.de

https://doi.org/10.5771/0042-059X-2021-3-376 - Generiert durch IP 216.73.217.60, am 09.05.2026, 14:18:05. © Urheberrechtlich geschützter Inhalt. Ohne gesonderte Erlaubnis ist jede urheberrechtliche Nutzung untersagt, insbesondere die Nutzung des Inhalts im Zusammenhang mit, für oder in KI-Systemen, KI-Modellen oder Generativen Sprachmodellen.