Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
MBA em *Advanced Analytics & Business Optimization*

Marco Pollo Almeida

Uncertainty Quantification in Credit Risk: a case study

Curitiba
2025

Marco Pollo Almeida

Uncertainty Quantification in Credit Risk: a case study

Monografia apresentada ao MBA em *Advanced Analytics* & *Business Optimization* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Dr. Anderson Ara

Curitiba
2025

Uncertainty Quantification in Credit Risk: a case study

Marco Pollo Almeida[1], Anderson Ara[2]

[1]Statistician/Data Scientist - São Paulo/SP *
[2]Professor do Departamento de Estatística - DEST/UFPR, ara@ufpr.br

Este artigo apresenta uma aplicação de Predição Conformal em um problema clássico de risco de crédito, utilizando a base de dados *German Credit* e dois modelos de classificação base: *Naive Bayes* e *Random Forest*. A partir do esquema de *split conformal prediction*, construímos conjuntos preditivos para o rótulo de risco com garantia de cobertura marginal no nível nominal de 90% sob a suposição de permutabilidade dos dados. Avaliamos o desempenho por meio da cobertura empírica e do tamanho médio dos conjuntos no conjunto de teste, além de métricas tradicionais dos modelos base (acurácia e AUC). Os resultados indicam que a cobertura empírica permanece próxima ou acima do nível nominal, enquanto o tamanho médio dos conjuntos sugere uma proporção relevante de decisões claras ($|\hat{C}(x)| = 1$) e uma fração menor na zona cinza ($|\hat{C}(x)| = 2$). Discutimos como essa decomposição em zonas de decisão pode apoiar regras de negócio, distinguindo aprovações e recusas com maior confiança e encaminhando casos ambíguos para análise adicional. Por fim, apontamos direções para trabalhos futuros envolvendo variações mais avançadas de Predição Conformal e critérios de avaliação que considerem custos assimétricos de erro.
Palavras-chave: Predição Conformal; Inferência Conformal; risco de crédito; classificação; *Random Forest*; *Naive Bayes*; conjuntos de predição; cobertura empírica; incerteza preditiva.

This paper presents an application of Conformal Prediction to a classical credit-risk classification problem using the *German Credit* dataset and two base classifiers: *Naive Bayes* and *Random Forest*. Using *split conformal prediction*, we construct prediction sets for the risk label with a marginal coverage guarantee at the nominal 90% level under the assumption of data exchangeability. We assess performance through empirical coverage and average set size on a held-out test set, in addition to standard metrics of the base models (accuracy and AUC). The results show that empirical coverage remains close to or above the nominal level, while the average set size indicates a substantial proportion of clear decisions ($|\hat{C}(x)| = 1$) and a smaller fraction in the gray zone ($|\hat{C}(x)| = 2$). We discuss how this decision-zone decomposition can support business rules by distinguishing approvals and rejections with greater confidence and routing ambiguous cases to additional review. Finally, we outline future directions involving more advanced variants of Conformal Prediction and evaluation criteria that account for asymmetric error costs.
Keywords: Conformal Prediction; Conformal Inference; credit risk; classification; *Random Forest*; *Naive Bayes*; prediction sets; empirical coverage; predictive uncertainty.

1. Introduction

Statistical and *machine learning* (ML) models are widely used by financial institutions to support decisions across a variety of banking products. In credit-risk problems, it is common to use classification models to predict whether a new customer is a *good* or *bad* payer based on registration, behavioral, and in some cases, macroeconomic information. Typical examples include models based on logistic or multinomial regression, *Random Forest*, *gradient boosting* methods, among others.

In practice, these models provide a point estimate of the probability of default, which is then used in internal decision rules (for example, credit approval or rejection). However, they typically do not provide a measure of uncertainty associated with each individual prediction.

Standard model-evaluation tools are important for assessing the overall performance of the classifier. Nevertheless, these metrics do not provide formal gua-

*marcopollostat@gmail.com

rantees regarding prediction quality at the individual level. Moreover, the usual statistical assumptions of a model, potential sampling biases, and changes in the economic environment may affect the reliability of these models, without such additional uncertainty being clearly quantified.

In this context, the Conformal Inference approach has emerged as a complement to traditional predictive models by producing prediction sets accompanied by coverage guarantees under the assumption of data exchangeability [1]. Instead of returning only a class label or a point probability, such methods associate each new observation with a set of labels that, with a pre-specified nominal probability, contains the true label.

In recent years, the literature has documented applications of Conformal Inference in different fields, including regression, time series, and high-dimensional classification problems [2]. In finance, recent work has explored the use of this approach to quantify uncertainty in forecasts of financial variables and risk metrics [3]. Despite this, applications focused specifically on credit risk (whether retail or wholesale) remain relatively less explored.

In this work, we analyze a credit-risk classification problem using the *German Credit* dataset, which is widely used in the literature as a benchmark for evaluating default models. We consider two models: *Naive Bayes* and *Random Forest*. Based on the class probabilities provided by these models, we apply Conformal Inference techniques to construct prediction sets for classification.

The goal of this study is to present a practical application of the aforementioned methodology. In particular, we seek to answer the following questions: (i) do the theoretical coverage guarantees hold in practice? and (ii) how can this information be interpreted from the perspective of credit-risk decision-making?

The paper is organized as follows. In Section 2, we describe the dataset, the computational resources, and the Conformal Inference methodology adopted in this study. In Section 3, we present and discuss the results obtained for the *German Credit* dataset, emphasizing the relationship between coverage, set size, and decision zones. Finally, in Section 4, we summarize the main conclusions and suggest directions for future work.

2. Development

2.1. Data set

The study uses the Statlog German Credit Data dataset, available from the UCI repository [4], which consists of 1,000 customers and 20 covariates of categorical and numerical types, in addition to the (*target*) variable that indicates the customer's credit risk (response variable Y). We adopt the convention $Y = 1$ for a good payer and $Y = 0$ for a bad payer. The proportion of bad payers is 40%, and approximately 60% are good payers. From the 20 original variables, nine predictors were selected to compose the covariate vector X. These variables were chosen after a pre-processing step that included normalization of the numerical covariates, removal of predictors with zero variance, unique values, or high correlation among themselves.

Summarized descriptions of the covariates used and the response variable are presented in Table 1. To characterize the customer profile, we computed basic descriptive statistics for the selected variables, focusing on measures of central tendency for the numerical covariates, stratified by good and bad payer in Table 2. Additionally, we built a plot showing the behavior of the numerical covariates (`dur_months`, `cred_amt`, and `age`), using distinct colors for each risk class; see Figure 1.

Tabela 1: Description of the variables used from the German Credit dataset.

Variable	Brief description
`acct_bal`	Checking account status
`dur_months`	Credit contract duration, in months
`cred_hist`	Previous credit history (delinquencies, repaid loans, no credit history)
`purpose`	Main purpose of the credit (car, furniture, education, etc.)
`cred_amt`	Amount of credit granted
`savings`	Customer's savings/investment balance
`emp_dur`	Length of current employment (year ranges, including unemployment)
`property`	Type of declared property/asset (real estate, savings, others, none)
`age`	Applicant's age, in years
`target`	Credit risk (1 = good payer; 0 = bad payer)

Tabela 2: Descriptive statistics of the numerical variables, by risk class.

Variable	Mean (Good)	SD (Good)	Mean (Bad)	SD (Bad)
`dur_months`	19	11,08	25	13,28
`cred_amt`	2985,44	2401,50	3938,13	3535,82
`age`	36	11,35	34,00	11,23

Descriptive analyses were also performed for the other categorical variables included in the study, using frequency tables and bar charts stratified by risk class. However, these results are not presented here in detail due to space limitations.

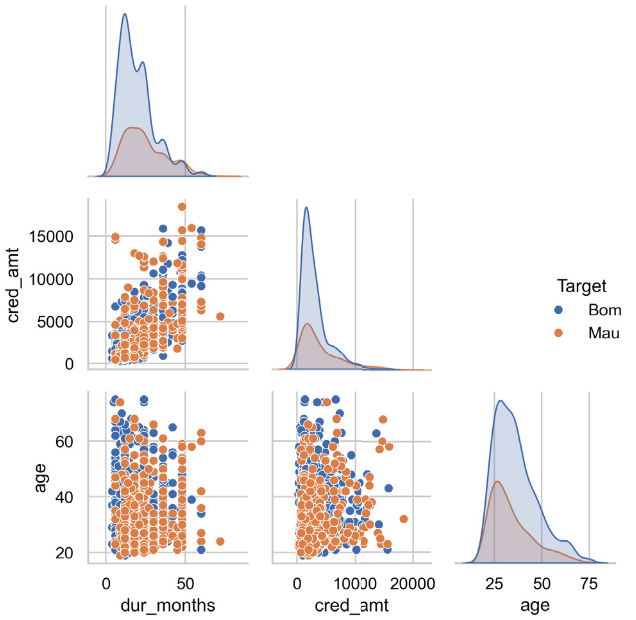

Figura 1: Distribution of the numerical variables vs. risk class (Target).

2.2. Computational resources

All data pre-processing steps, classifier fitting, and application of Conformal Inference techniques were implemented in Python (version 3.12.11), using the `pandas` library for data manipulation, `numpy` for numerical operations, and `scikit-learn` for fitting the *Naive Bayes* and *Random Forest* classifiers. The construction of conformal sets was carried out with the `MAPIE` (*Model-Agnostic Prediction Interval Estimator*) library, which provides an interface compatible with the `scikit-learn` ecosystem; see [5]. Figures were produced with the `matplotlib` and `seaborn` libraries. The hardware used was a laptop with 16 GB of RAM, and the entire project was carried out in the `VS CODE` programming environment.

2.3. Methodology

2.3.1. Classification models

We use two models that are widely applied in credit-risk problems: *Naive Bayes* [6, 7] and *Random Forest* [8, 9]. These models were chosen because they combine well- established interpretations in the literature with straightforward implementations in the `scikit-learn` library.

The *Naive Bayes* classifier is based on the assumption of conditional independence of the covariates given the class label. In practical terms, the model estimates, for each class, the distribution of the explanatory variables and combines these estimates through Bayes' rule to obtain the posterior probability of each class for a new individual.

Random Forest is an *ensemble learning* method that builds a collection of decision trees fitted on *bootstrap* samples of the training data, combining them by majority vote in the classification case. At each internal split of the trees, a random subset of covariates is considered, which induces diversity among the trees and contributes to reducing the variance of the *ensemble*.

2.4. Conformal Prediction

Let $\mathscr{D}_n = \{(X_i, Y_i)\}_{i=1}^n$ be a dataset, where X_i denotes the covariate vector of customer i and $Y_i \in \{0, 1\}$ indicates their credit-risk status (1 = good payer, 0 = bad payer). Given a new customer (X_{n+1}, Y_{n+1}), the goal of Conformal Inference is to construct a *predictive set* for the class, $\hat{C}_n(X_{n+1}) \subseteq \{0, 1\}$, such that the probability of containing the true label is at least equal to a pre-specified nominal level $1 - \alpha$:

$$\mathbb{P}\big(Y_{n+1} \in \hat{C}_n(X_{n+1})\big) \geq 1 - \alpha, \tag{1}$$

under the assumption of data exchangeability [1].

In the binary credit-risk setting considered in this work, each new customer is associated with a set $\hat{C}_n(x)$ that can be $\{1\}$ (approval case, "clear"), $\{0\}$ (rejection case, "clear") or $\{0, 1\}$ (intermediate case, in which the method indicates higher uncertainty).

2.5. Split Conformal Prediction

According to [1], *Split Conformal Prediction* is the simplest and most computationally efficient way to apply conformal prediction. Following the notation of [1], let $\mathscr{D}_n = \{(X_i, Y_i)\}_{i=1}^n$ be a set of exchangeable observations and consider a partition of the indices I_{tr} and I_{cal} such that $I_{\text{tr}} \cup I_{\text{cal}} = \{1, \ldots, n\}$ and $I_{\text{tr}} \cap I_{\text{cal}} = \emptyset$. The training set is $\mathscr{D}_{I_{\text{tr}}} = \{(X_i, Y_i)\}_{i \in I_{\text{tr}}}$, and the calibration set is $\mathscr{D}_{I_{\text{cal}}} = \{(X_i, Y_i)\}_{i \in I_{\text{cal}}}$.

First, a classification model is fitted only on $\mathscr{D}_{I_{\text{tr}}}$, yielding the probabilistic classifier

$$\hat{f}_{I_{\text{tr}}}(x) \approx \mathbb{P}(Y = k \mid X = x), \tag{2}$$

for each $k = 0, 1$.

Next, for each observation in the calibration set, absolute residuals are used to obtain the scores.

The score construction is given by

$$R_i = |Y_i - \hat{f}_{I_{\text{tr}}}(x_i)|, \quad i \in I_{\text{cal}}. \tag{3}$$

From the sample of scores $\{R_i\}_{i \in I_{\text{cal}}}$, the conformal quantile $\hat{q}_{n_2}$ at level $(1-\alpha)$ is computed according to the definition: $\hat{q}_{n_2} = \lceil (1-\alpha)(n_2+1) \rceil$, the smallest of $\{R_i\}_{i \in I_{\text{cal}}}$.

Finally, the conformal predictive set is defined using $\hat{f}_{I_{\text{tr}}}$ and $\hat{q}_{I_{\text{cal}}}$ as follows:

$$\hat{C}_n(x) = \left\{ \hat{f}_{I_{\text{tr}}}(x) - \hat{q}_{I_{\text{cal}}}, \hat{f}_{I_{\text{tr}}}(x) + \hat{q}_{I_{\text{cal}}} \right\}. \tag{4}$$

The coverage guarantee 1 is achieved because, conditional on the appropriate training set, the calibration residuals and the test residual $R_{n+1} = |Y_{n+1} - \hat{f}_{I_{\text{tr}}}(x_{n+1})|$ are all *i.i.d.*, and therefore exchangeable. Symmetry in the construction of the scores is essential to ensure the statistical property of exchangeability. The key idea is to construct residuals in a way that treats all data determining their distribution, including the test data, symmetrically. **Algorithm 1** details the complete procedure of the Split Conformal Prediction methodology presented in this section.

Algorithm 1 Split Conformal Prediction

1. **Input:** data $\mathscr{D}_n = \{(X_i, Y_i)\}_{i=1}^n$ and level $\alpha \in (0, 1)$.
2. Partition the indices into two disjoint subsets: I_{tr} (training) and I_{cal} (calibration).
3. Train a probabilistic model on $\{(X_i, Y_i)\}_{i \in I_{\text{tr}}}$, obtaining the predictor $\hat{f}_{I_{\text{tr}}}$.
4. **For** each $i \in I_{\text{cal}}$:
5. Compute the score $R_i = |Y_i - \hat{f}_{I_{\text{tr}}}(X_i)|$.
6. Compute the conformal quantile $\hat{q}_{I_{\text{cal}}}$ as the empirical quantile at level $(1-\alpha)$ of the sample $\{R_i\}_{i \in I_{\text{cal}}}$.
7. **To predict** at a new point x:
8. Compute the point prediction $\hat{f}_{I_{\text{tr}}}(x)$.
9. Define the conformal/predictive set $\hat{C}_n(x) = \{\hat{f}_{I_{\text{tr}}}(x) - \hat{q}_{I_{\text{cal}}}, \hat{f}_{I_{\text{tr}}}(x) + \hat{q}_{I_{\text{cal}}}\}$.
10. **Output:** predictive set $\hat{C}_n(x)$.

2.6. Evaluation measures

To evaluate the performance of the conformal sets, the German Credit dataset was split into three mutually exclusive subsets: training (70%), calibration (15%), and test (15%). In this split, we stratified with respect to Y in order to preserve the proportion of good and bad payers. The classification models were fitted only on the training subset; the conformal scores were constructed on the calibration subset; and all evaluation metrics were estimated on the test subset.

The main performance measure considered is the *empirical coverage* of the conformal sets. Denoting by n_{test} the number of test observations and by $\hat{C}_n(X_i)$ the predictive set associated with customer i, we define

$$\widehat{\text{cov}} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \mathbf{1}\{Y_i \in \hat{C}_n(X_i)\}, \tag{5}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. This metric summarizes the proportion of customers for whom the true label belongs to the conformal set, and it is compared directly to the nominal coverage level $1-\alpha$ used in calibration (in this study, $\alpha = 0{,}1$).

As an efficiency measure, we use the *average size* of the prediction sets, given by

$$\widehat{\text{width}} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} |\hat{C}_n(X_i)|, \tag{6}$$

where $|\hat{C}_n(X_i)|$ is the number of labels in the set associated with customer i. In the binary case, $|\hat{C}_n(X_i)| \in \{1, 2\}$, so $\widehat{\text{width}}$ can be interpreted as the relative proportion of "clear" decisions (singleton sets) versus cases in which both classes remain plausible (two-class sets). For comparison with traditional approaches, we also computed the accuracy and AUC of the classification models on the test set, without the conformal step.

2.7. Decision zones in credit risk

Based on $\hat{C}_n(x)$, we propose three decision zones of interest for the credit-risk context. The idea is to translate the set size into operational categories that support business rules. *Clear good-payer zone:* cases in which $\hat{C}_n(x) = \{1\}$, that is, the set contains only the good-payer class. Thus, the classifier and the conformal procedure point toward credit approval; *Clear bad-payer zone:* cases in which $\hat{C}_n(x) = \{0\}$, where the set contains only the bad-payer class. These customers are strong candidates for rejecting the application or for more restrictive policies; *Gray zone:* cases in which $\hat{C}_n(x) = \{0, 1\}$, indicating that, at level $1-\alpha$, both classes remain plausible. In this situation, the automatic decision is less reliable, and customers may be routed to human review.

3. Results

3.1. Coverage and set size

Table 3 summarizes, for each model, the empirical coverage estimated on the test set and the average size of the conformal sets, for a nominal confidence level of 90%. In both cases, the empirical coverage remains above the nominal level, which is consistent with the theoretical guarantees.

Regarding the average size of the prediction sets, the results suggest that both models yield a value of approximately 1.4. Since this lies between 1 and 2, it indicates that, on average, the method is operating close to the lower bound (size 1).

Tabela 3: Empirical coverage and average size of the conformal sets.

Base model	Empirical coverage	Average size
Naive Bayes	*0,920*	*1,400*
Random Forest	*0,927*	*1,427*

In practice, an average size of approximately 1.4 indicates that the classification model, together with Split Conformal Prediction, is discriminating uncertainty with reasonable efficiency: in most cases (60%), it was able to reduce the uncertainty set to a single class while maintaining the marginal coverage guarantee $1 - \alpha$.

3.2. Decision zones by model

Figure 2 shows, for each model, the proportion of test-set observations classified in the gray zone and in the clear zone. For both models, most observations fall into decision zones considered clear, although a relevant fraction remains in the gray zone, in which both classes remain plausible from the conformal perspective.

Figure 3 provides further detail by decomposing the decision zones according to the true label Y in the test set. For *Naive Bayes*, among customers who are in fact bad payers ($Y = 0$), 25 observations fall in the gray zone and 20 in the clear zone, indicating a higher concentration of ambiguous cases in this group. Among good payers ($Y = 1$), there are 35 observations in the gray zone and 70 in the clear zone, implying that about two thirds of the test-set good payers are classified in a clear decision zone. For *Random Forest*, the pattern is similar: among bad payers ($Y = 0$), 29 observations are in the gray zone and 16 in the clear zone; among good payers ($Y = 1$), 35 are in the gray zone and 70 in the clear zone. These results suggest that, for both models, Conformal Prediction tends to produce clearer decisions for the good-payer class, whereas the bad-payer class concentrates a larger proportion of cases in the gray zone, which may be relevant for defining more conservative credit policies for these profiles.

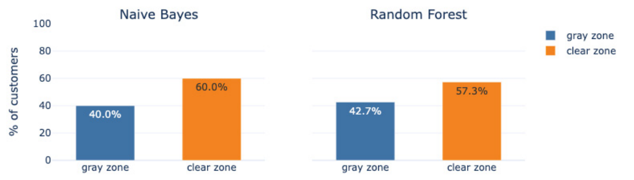

Figura 2: Proportion of test-set observations in the gray zone and in the clear zone, for the models.

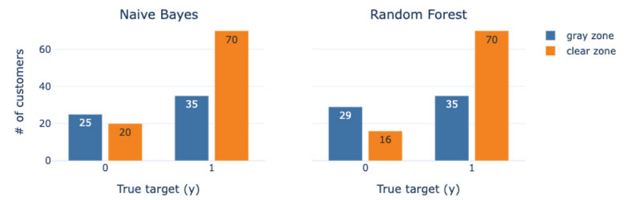

Figura 3: Number of test-set observations in the gray zone and in the clear zone, decomposed by the true label $Y \in \{0, 1\}$, for the models.

3.3. Model results

In this subsection, we analyze in greater detail the performance of the *Random Forest* model, chosen as the base model for proposing the institution's business rules. Although analogous analyzes were conducted for the *Naive Bayes* classifier, the results were very similar and, due to space limitations, are not presented here in detail. Table 4 summarizes the classic performance metrics of the *Random Forest* model in the test set, without the conformal layer.

Tabela 4: Performance of the *Random Forest* model on the test set (without Conformal Prediction).

Metric	Value
Point accuracy	0,793
AUC-ROC	0,835

When incorporating the Conformal Prediction layer with nominal level $1 - \alpha = 0{,}90$, we obtain the metrics shown in Table 5. The estimated overall coverage is approximately 0,907, very close to the nominal level, indicating that the vast majority of true labels in the test set fall within the conformal sets produced by the method. The average set size is $\widehat{\text{width}} \approx 1{,}32$, meaning that, on average, the method returns slightly more than one label per observation, while maintaining good efficiency.

Tabela 5: Conformal-layer metrics for the *Random Forest* model ($\alpha = 0{,}1$) on the test set.

Metric	Value
Overall coverage $\widehat{\text{cov}}$	0,907
Average set size $\widehat{\text{width}}$	1,32
Proportion of cases with $\lvert\hat{C}\rvert = 1$	0,68
Proportion of cases with $\lvert\hat{C}\rvert = 2$	0,32
Proportion of cases with $\lvert\hat{C}\rvert = 0$	0,00
Coverage for $Y = 1$ (good payer)	1,00
Coverage for $Y = 0$ (bad payer)	0,689

We observe that about 68% of the test observations yield singleton sets ($\lvert\hat{C}\rvert = 1$), that is, decisions considered clear from the conformal standpoint, whereas approximately 32% of cases remain in the gray zone ($\lvert\hat{C}\rvert = 2$), requiring additional review or more conservative credit policies. No cases with $\lvert\hat{C}\rvert = 0$ were observed, which is consistent with the theoretical construction of the method. Decomposing coverage by class shows that, for good payers ($Y = 1$), coverage is essentially perfect, whereas for bad payers ($Y = 0$), coverage is around 0,689. This asymmetry is consistent with the fact that Conformal Prediction guarantees marginal coverage over the mixture of classes, but not necessarily class-conditional (balanced) coverage, and it highlights the importance of analyzing the conformal-layer performance separately across distinct risk segments.

From an applied perspective, these supplementary results provide a more granular understanding of how the conformal layer interacts with the base model: on the one hand, it preserves strong overall coverage performance; on the other hand, it makes explicit the proportion of automatic decisions versus cases that should be routed to the gray zone, as well as differences in coverage between good and bad payers. This type of analysis is essential for the institution to calibrate the operational use of conformal sets, adjusting confidence levels and business rules according to the desired risk appetite.

4. Final Comments

This work presented an application of Conformal Prediction to a classical credit-risk problem, using the German Credit dataset and two base classification models, *Naive Bayes* and *Random Forest*. Using the *split conformal prediction* approach, it was possible to construct predictive sets for the probability of default with marginal coverage guarantees close to the nominal 90% level, while maintaining good predictive performance in terms of accuracy and AUC of the base model. The analysis of decision zones showed that a relevant share of customers can be classified into clear good- or bad-payer zones, while a smaller fraction remains in the gray zone, in which uncertainty should be carefully assessed.

From the perspective of business rules, the conformal layer adds an important dimension to the decision-making process: in addition to the score or point probability provided by the base model, the institution gains access to sets $\hat{C}(x)$ that make it possible to transparently distinguish customers with a strong indication of good payer ($\hat{C}(x) = \{1\}$), customers with a strong indication of bad payer ($\hat{C}(x) = \{0\}$), and borderline cases ($\hat{C}(x) = \{0, 1\}$), which are candidates for human review or more conservative policies.

For future research, we plan to explore more advanced variants of Conformal Prediction described in [1], such as *full conformal*, *jackknife+*, and *cross-conformal* schemes. In addition, it is of interest to investigate more refined evaluation criteria that account for asymmetric misclassification costs between good and bad payers, as well as to apply the methodology to other credit portfolios and banking products.

Acknowledgements

The authors gratefully acknowledge the faculty of the MBA program at the Federal University of Paraná for their support and contributions.

Referências

[1] R. J. Tibshirani. *Conformal Prediction*. Advanced Topics in Statistical Learning (lecture notes), University of California, Berkeley, Spring 2023. Available on: `https://www.stat.berkeley.edu/~ryantibs/stat239b`.

[2] Memmesheimer, Pascal and Heuveline, Vincent and Hesser, Jürgen *A Systematic Review of Conformal Inference Procedures for Treatment Effect Estimation: Methods and Challenges*. arXiv preprint arXiv:2509.21660, 2025.

[3] A. N. Angelopoulos, R. F. Barber and S. Bates. *Theoretical Foundations of Conformal Prediction*. Preprint, to appear at Cambridge University Press, arXiv:2411.11824, 2024.

[4] H. Hofmann. *Statlog (German Credit Data)*. UCI Machine Learning Repository, University of California, Irvine. Available on: `https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data`.

[5] V. Taquet, V. Blot, T. Morzadec, L. Lacombe and N. Brunel. *MAPIE: an open-source library for distribution-free uncertainty quantification*. arXiv preprint arXiv:2207.12274, 2022.

[6] P. Langley, W. Iba and K. Thompson. *An Analysis of Bayesian Classifiers*. In: Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI'92), pp. 223–228, 1992.

[7] P. Domingos and M. Pazzani. *On the Optimality of the Simple Bayesian Classifier under Zero-One Loss*. Machine Learning, 29:103–130, 1997.

[8] L. Breiman. *Bagging Predictors*. Machine Learning, 24(2):123–140, 1996.

[9] L. Breiman. *Random Forests*. Machine Learning, 45(1):5–32, 2001.