Explainable algorithm for credit risk prediction based on Patterns

PhD Research Proposal

by

Erwis Melchor Pérez

Doctoral Advisors:

Dr. José Francisco Martínez Trinidad

Dr. Jesús Ariel Carrasco Ochoa, INAOE

Instituto Nacional de Astrofísica, Óptica y Electrónica

January, 2025

Santa María de Tonantzintla, Puebla, CP 72840

Contents

Abstract

Credit risk prediction involves assigning customers a risk category, such as defaulter or non-defaulter. In the financial sector, obtaining an accurate prediction and ensuring would be comprehensible to credit experts, applicants, and regulators is essential. The challenge addressed in this Ph.D. research is to develop a pattern-based algorithm that predicts credit risk and explains the its predictions.

It is also important to note that the data utilized in credit risk prediction is frequently mixed and imbalanced, which can generate a bias towards the majority class. Therefore, a comprehensive analysis of data balancing methods is required to help obtain a suitable credit risk prediction algorithm. Thus, a comprehensive analysis of data balancing methods is included in the preliminary results to determine the most appropriate method for credit risk prediction. A study of pattern miners and pattern-based classifiers was also conducted to select the most effective algorithm for pattern mining and filtering. Finally, the proposal culminates by presenting the first algorithm for pattern-based credit risk prediction.
Keywords— Patterns, credit risk, imbalanced class, explain

1 Introduction

A credit is a cash loan that must be repaid within a defined time and under established conditions [1]. The decision to approve or reject a credit is based on analyzing the applicant's ability to repay the credit and their probable repayment behaviour.

Granting or denying credit is a risk. Therefore, credit risk is the risk associated with financing, the risk of customers defaulting on a credit, credit card, and other lending services [2]. In other words, the risk that a person fails to pay a financial institution or defaults. Credit risk prediction in a financial institution is performed by an analyst using the institution's current regulations and policies to approve or deny a credit application.

Traditionally, credit risk prediction is based on personal experiences, and uses 5C's [3] (five crucial aspects to analyze for a financial institution: character, capacity, capital, collateral, and conditions). Credit risk prediction is essential for financial institutions; thus, automated tools have been developed by combining optimization algorithms with balancing and feature selection methods and using supervised and unsupervised classifiers for credit risk prediction [4]. The performance obtained by the classifiers is reported through metrics such as ACC, F1-score, and AUC. However, these performance measures do not explain how classifiers identify defaulters and non-defaulters.

In Mexico, regulators such as the National Banking and Securities Commission (CNBV) [1] [5] requires financial institutions to explain and justify their credit decisions. Lack of explainability could lead to regulatory problems. Therefore, an explainable method will allow analysis and applicants to understand the factors contributing to their credit rating. Understanding the decision-making methods and features can help improve the creditworthiness of candidates. However, there are

[1] https://www.cnbv.gob.mx

several challenges in developing an explainable algorithm for predicting credit risk.

1. **Data complexity:** Numerical and non-numerical data are used when assessing credit risk. Preprocessing this data carefully is essential to avoid missing any important information for prediction.

2. **Interpretation of methods:** Developing an algorithm to explain credit risk prediction is challenging. This is because it's difficult to understand the results of supervised and unsupervised classifiers.

3. **Accuracy and explainability:** Classifiers with high accuracy have been reported in the-state-of-the-art. However, it is difficult to determine the reasons for the classifier´s reasons for why an application is accepted or denied by the classifier.

Explainability is crucial to ensure the confidence of credit institutions, as decisions are made using black box classifiers, and stakeholders may not understand the reasoning behind those decisions. Therefore, this research will investigate this aspect of explainability in credit risk prediction.

1.1 Justification

Several studies concentrate on credit risk prediction in financial institutions. The credit risk datasets contain information on individuals, credit cards, and companies labelled as default or non-default. These datasets contain mixed data and they are imbalanced.

In the state-of-the-art, supervised and unsupervised classifiers have been used to solve the problem of credit risk prediction. These classifiers have been implemented individually or in combination with balancing and feature selection methods for credit risk prediction.

However, these classifiers do not provide explanations for their credit risk predictions. Financial institutions must clarify their decisions to customers and regulatory bodies, such as the CNBV. Therefore, it is essential to create an algorithm that not only does it predict credit risk, but also provides a transparent explanation of what it's predicting.

Therefore, in this PhD research, we consider it important to develop an algorithm that allows working with mixed and imbalanced data to predict credit risk, which can explain its decision. Furthermore, since patterns can be used for providing explanations, we will base our algorithm on mining and using patterns for credit risk prediction.

1.2 Problem Statement

The works reported in the literature to predict credit risk are complex and operate as black boxes. These works can make accurate predictions, but their predictions can be difficult for experts and non-experts to understand. It is, therefore, essential to find ways to explain the prediction results to make them understandable to stakeholders such as financial institutions, regulators and applicants. In addition, feature selection and balancing methods are used on datasets before classifiers, where categorical data is transformed into numerical data. This pre-processing can make the classifier's prediction challenging to interpret by removing information relevant to the credit risk prediction.

Therefore, to provide information on why the algorithm's prediction will allow financial institutions to understand and explain the decision to grant or deny credit. Credit risk prediction needs to strike a balance between accuracy and explainability, such that an algorithm for credit risk prediction would have good accuracy without limiting explainability, thereby facilitating the understanding of the prediction by the parties involved in granting or denying credit.

Thus, the problem to be solved in this doctoral research is to develop a pattern-based algorithm for credit risk prediction, considering the imbalance of databases and mixed data, which obtains as good results as the state-of-the-art but allows an explanation of the prediction, since currently, the works in the literature on credit risk prediction do not explain the prediction.

1.3 Research Questions

- How can patterns be identified in mixed and imbalanced datasets for credit risk prediction?
- How can we extract a representative subset of credit risk prediction patterns?
- How can we build an efficient and explained algorithm for predicting credit risk based on patterns?
- How can we explain the results of credit risk prediction?

1.4 Objetives

1.4.1 General objective

Develop an algorithm for credit risk prediction based on patterns that allow explaining its prediction results. The proposed algorithm should be superior in quality of prediction compared to the state-of-the-art additionally, the proposed method should explain the result in terms of the experts.

1.4.2 Specific objectives

1. Develop an algorithm for mining patterns to predict credit risk in mixed and imbalanced databases.

2. Develop a filtering algorithm for selecting only patterns appropriate for credit risk prediction.

3. Develop a pattern-based algorithm for credit risk prediction that can explain the prediction result, using the algorithms obtained from (1) and (2).

1.5 Expected Contributions

The main contributions expected at the end of this doctoral research are as follows:

1. A pattern mining algorithm for mixed and imbalanced data, based on patterns for predicting credit risk.

2. An algorithm to select a subset of patterns suitable for predicting credit risk.

3. An algorithm for credit risk prediction, based on patterns, that explains its prediction results and with comparative quality than state-of-the-art algorithms.

1.6 Methodology

The methodology proposed for reaching the objectives is the following:

I.- Review and critically analyze the related works reported in the literature. In addition, to collect the most commonly used datasets for credit risk prediction. Currently, we have collected databases from: from *The UCI Machine Learning Repository* [6], *Kaggle* [2], a database from Esperanza Indigena Zapoteca (EIZ) financial institution in Oaxaca, Mexico [3], as shown in **Table 3**.

II.- Develop an algorithm for mining patterns to predict credit risk in mixed and imbalanced databases.

[2] https://www.kaggle.com/datasets

[3] http://www.cooperativaeiz.com/

i.- Review and analysis of pattern mining algorithms on mixed and imbalanced datasets.

a.- From the algorithms used for pattern extraction [7, 8, 9, 10], we will select the one that obtains the best performance measures using credit risk prediction datasets.

b.- Compare balancing methods by evaluating their impact on pattern extraction algorithms on mixed datasets. This evaluation will be done regarding the number of patterns obtained, the quality of the patterns (using the pattern quality measures) and the prediction of credit risk prediction using quality measures such as accuracy, area under the curve and others.

ii.- Based on the experience gained in i), propose a pattern mining algorithm for credit risk prediction.

a.- Considering the pattern mining algorithm selected from the study in (II.i.a). Analyze the extracted patterns to see how the balancing methods impact the mined patterns.

b.- Analysis of how patterns are extracted.

- Consider pattern extraction in mixed and imbalanced datasets.
- Using scoring measures for instances and features, taking into account, the imbalance for pattern extraction in the database.

c.- To develop a pattern mining algorithm for credit risk prediction.

- Consider non-numerical and numerical features for the supervised pattern mining algorithm.
- Identifying feature value combinations common to defaulters and non-defaulters to help mine patterns.
- Considering the imbalance of the dataset when mining patterns.

III.- To develop a filtering algorithm to select only patterns appropriate for credit risk prediction and to explain the result.

i.- A critical review of pattern filtering algorithms proposed in the literature. It is essential to ensure that the selected patterns represent the minority class, as this class includes defaulters.

ii.- Simplify redundant patterns.

a.- Explore how to filter patterns by class.

b.- Quality measures of patterns should be considered to obtain a representative subset of patterns.

c.- Define a criterion of similarity and dissimilarity for the selected patterns.

d.- Analyse the elimination of patterns that cover the same instances but with a subset or superset of features.

e.- Analyse the elimination of patterns present in the two classes.

f.- Remove patterns in each class with the lowest frequency or quality score and consider the opinion of a credit sector expert for removing some patterns.

iii.- Evaluate the quality and relevance of the filtered patterns for credit risk prediction. Compare the performance of the selected patterns for credit risk prediction.

IV.- Develop a pattern-based algorithm for credit risk prediction that can explain the prediction result.

i.- Development of a Pattern-based classifier for credit risk prediction.

a.- Review and analysis of supervised pattern-based classifiers.

b.- Patterns obtained by the algorithms proposed in points (II) and (III) will be used for classification.

c.- To classify new instances, consider, for example, the support and quality of the patterns found for each class. Higher support or quality indicates stronger membership of the new instance to that class.

ii.- Explanation of the prediction result.

a.- Find out from the experts the appropriate presentation of prediction explanations.

b.- Define the elements of the patterns useful to explain the prediction.

c.- Define how to explain the prediction using patterns following the expert opinion to provide the reasons for a decision when evaluating a credit application. The next aspects will be considered:

- Explain the prediction using several patterns.
- Sort patterns from largest to smallest by the support or quality measure of the patterns that define the predicted class. Consider the first k patterns to explain the prediction result.
- Analyze the common values of the patterns that define the class of the new instance, then formulate explanations around these recurring features.

V.- Evaluate the quality of the proposed pattern-based algorithm for credit risk prediction.

i Development of an experimental platform for evaluating credit risk prediction.

ii Evaluation of the proposed pattern-based algorithm for credit risk prediction and comparison with the state-of-the-art for credit risk prediction.

i. Make a critical study of the measures used to evaluate the performance of classifiers on problems with imbalanced classes. Select the most appropriate.

ii. Evaluate the performance of pattern-based classifiers when used for credit risk prediction with and without applying balancing methods.

iii. To use databases commonly used for credit risk prediction obtained from *The UCI Machine Learning Repository*, *Kaggle*, a database from a financial institution in Oaxaca, Mexico [4].

iv. Comparison of the proposed credit risk prediction algorithm based on the patterns obtained in [IV] with credit risk predictors of state of the art on the prediction quality and explainability of the credit risk prediction result.

1.7 Publications plan

The expected publications are outlined as follows:

International Journal of Applied Pattern Recognition (JCR journal)

- **Objetive:** Evaluate and compare the most commonly used oversampling and instance selection methods and their combination to improve credit risk prediction.
- **Status:** Submitted (under review)
- **Submision date:** 16 Dec 2024

Journal II

- **Objetive:** Publish a JCR paper presenting the preliminary results of an explainable pattern-based algorithm for credit risk prediction
- **Estimated Submission Date:** August 2025

[4] http://www.cooperativaeiz.com/

- **Options journals for submission:**
  - Finance Research Letters
  - Journal of Risk and Financial Management
  - Computational Economics
  - Accounting and finance
  - Empirical Economics
  - Expert Systems with Applications
  - Journal of Computational Science

Journal III

- **Objetive:** Publish a JCR reporting the final application of our proposed algorithm for credit risk prediction.
- **Estimated Submission Date:** December 2026
- **Options journals for submission:**
  - Finance Research Letters
  - Journal of Risk and Financial Management
  - Computational Economics
  - Accounting and finance
  - Empirical Economics
  - Expert Systems with Applications
  - Journal of Computational Science

1.8 Schedule

Table 1: Timetable of tasks to be carried out per four-month period.

Activities	Four-month period*												
	2024			2025			2026			2027			2028
	1	2	3	4	5	6	7	8	9	10	11	12	13
Review and critically analyze the related works reported in the literature	✓	✓	✓										
Drafting the proposal.	✓	✓	✓										
Evaluate the effect of balancing methods to extract patterns in problems with imbalanced classes.		✓	✓										
Evaluate the effect of feature selection methods to extract patterns in problems with imbalanced classes.		✓	✓										
Propose a pattern mining algorithm for credit risk prediction.			✓										
How to extract patterns useful for credit risk with different data types (numeric and non-numeric).			✓										
Consider the frequency of occurrence of a pattern within a class.			✓										
A critical review of pattern filtering algorithms proposed in the literature for credit risk prediction and explainability of the result.													
Measures of pattern filtering should be considered to obtain a representative subset of patterns.													
Review and analysis of pattern-based classifiers.			✓										
Develop a pattern-based algorithm for credit risk prediction that can explain the prediction result.			✓										
Define how to explain the prediction using patterns.			✓										
Experimental comparison. Evaluate the quality of the results obtained.													
Writing and submitting articles.			✓	✓									
Thesis document writing.													
Submission of the thesis document to the advisors.													
Submission of the thesis document to the committee.													
Defence of the thesis.													

* The four-month period will be [January - April], [May - August] and [September - December]. It will start from [January - April] 2024, the student's admission to the doctoral schedule.

2 Related work

Several works on credit risk prediction have been developed in the literature. Most combine optimization algorithms with balancing and feature selection methods and use supervised and unsupervised classifiers. In the works that use balancing methods oversampling and undersampling, are commonly used. Oversampling methods add instances to the minority class and the undersampling methods remove instances from the majority class [11]. Once the balancing method is applied to the dataset, a supervised classifier is employed to compute the prediction of unseen instances.

Oversampling methods as SMOTE [12], ADASYN [13], and ROS [14] has been used [15, 16, 17, 18, 19, 20, 21]. SMOTE is one of the most commonly used and best-performing oversampling methods for credit risk prediction. In addition, other variants that use SMOTE as a basis for constructing a new oversampling method and are evaluated for credit risk prediction [22, 23, 24, 25, 26, 27, 28, 29, 30].

Undersampling is used in several works [31, 32, 33]. Other undersampling methods to predict credit risk include Edited Nearest Neighbours (ENN) [34] and Instance Hardness Threshold (IHT) [35].

Credit risk prediction datasets may contain redundant or irrelevant features in their descriptions. Feature selection (FS) is the process of reducing the feature set to improve the performance of the classifier, either by maintaining or even improving it compared to using the entire feature set. FS serves as a method to reduce the dimensionality of the data, thereby facilitating visualization and understanding [36].

FS has been used for credit risk prediction in several works [37, 38, 39, 40, 41, 42]. FS method based on the Information Gain (IG) is used in [2, 43]. Other authors [44, 45] use FS and Genetic Algorithm (GA)-based wrapper methods to identify the most relevant features. In [46], Random Forest (RF) is used as a method for the FS. In [47] delete any features with a correlation above 0,7. Anticipationbased dimensionality reduction is applied to reduce the dimensionality of the data and explained in [48]. In [49], clustering algorithms for credit risk prediction using Multiple Correspondence Analysis (MCA) are proposed to analyze the relationships between categorical features and facilitate data dimensionality reduction. In [50] the Kaiser-Meyer-Olkin (KMO) statistic is used to compare simple and partial correlation coefficients between features, to reduce dimensionality.

Balancing and FS methods are used individually to improve the performance of classifiers for credit risk prediction. Therefore, some studies have considered using balancing and FS methods together to obtain better performance when evaluating classifiers [51, 52, 53, 54]. Furthermore, other authors initially perform feature selection and then apply the balancing methods to the databases used for credit risk prediction as [55, 56, 30, 57, 58, 59].

In [60], the authors employ principal component analysis (PCA) to reduce the dimensionality of the dataset. In [61] Recursive Feature Elimination (RFE) is used to recursively eliminate less important features in terms of their contribution to the prediction of the target variable.

Other authors, such as [62] compared the performance of classifiers without applying any balancing or FS methods to the databases used for credit risk prediction. Several authors also use genetic algorithms to optimize a classifier, a FS method, or a balancing method for credit risk prediction [63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74]. In addition, other work is focused on optimizing classifiers (without applying any balancing or feature selection methods) to obtain the best training and a better result in credit risk prediction. The application of genetic algorithms for the optimization of classifiers [75, 76, 77], K-Fold CV [78], random search (RS) [79, 80], grid search (GS) [61, 81], Bayesian hyper-parameter optimization [82], using stochastic optimization [83] are used to obtain better performance.

2.1 Patterns-based

This section outlines the works on credit risk prediction based on patterns. In [84], a pattern is defined as a set of items that appear frequently in a database. However, they are defined and used differently for credit risk prediction.

In [85], a pattern is defined as a recurring association observed in credit card transactions associated with an increased probability of default. The authors propose a new similarity measure between cluster instances, using k-means to find groups of defaulters and useful features to identify defaulters. Patterns are manually extracted from the clustering results for each feature by analysing the train set. However, they aren't used to predict credit risk. The classifiers Logistic Regression (LR), Decision Trees (DT), Support Vector Machine (SVM) and Artificial Neural Networks (ANN) are used to evaluate performance using Accuracy (ACC), Sensitivity, Specificity and Area Under the Curve (AUC) for credit risk prediction.

In [86], a Multi-Objective Evolutionary Fuzzy algorithm based on rules is used for credit risk prediction. Define the patterns found in the data after applying data mining techniques to gain knowledge about the data. The algorithm identifies fuzzy rules used for credit risk prediction, but these fuzzy rules do not explain the prediction result clearly. This classifier performs better than J48, Random Tree, PART and Naïve-Bayes (NB) Tree using ACC, Recall, Precision, and AUC as performance measures.

In [87], it defines patterns as groups of data elements that have similarities that are the basis for their clustering. The authors introduced a clustering algorithm and a criterion to assess the quality of the cluster and determine the appropriate number of clusters. This criterion is based on a modified solver for support vector data description (SVDD) operating in Euclidean space. The centroids represent the clusters, ensuring that instances within each cluster exhibit similarities; the authors define these similarities obtained to the clusters as patterns. Metrics such as the silhouette coefficient [88], the weighted q (WQ) and the Davies-Bouldin index (BDI) [89] are used to evaluate the clustering results. Experiments showed that the proposed algorithm accurately estimates the number of clusters and produces compact clusters with highly similar data points.

In [90], a shapelet-based method is proposed to extract behavior patterns related to defaulted and non-defaulted loans, but it is not used to predict credit risk. The features' odds achieve the interpretation of the LogR classifier. The odds ratio is used to evaluate the influence of the features on the output probability. The Behavior2Shapelets method was compared to Linear Discriminant Analysis (LDA), LR, SVM, ANN, DT, RF, eXtreme Gradient Boost (XGBoost), and Bagging classifiers. It outperformed the other classifiers based on performance measures F1 score, G mean, AUC, and Kolmogorov-Smirnov (KS), as presented in [91, 92].

2.2 Explainable

This subsection describes the works reported in the literature that explain the results of credit risk prediction.

In [93] the authors describe how, in financial services, every decision must be meaningful, with users needing to understand why an instance is classified as a defaulter. While [94] consider AI models to be 'black boxes' and thus they are not well suited for financial systems and could create new difficulties for organizations operating in the financial services industry, leading decision-makers to reject AI systems.

The purpose of explainability is to provide the reasons behind the prediction results of a classifier in human terms to reach a broader usefulness in practice [95]. Explainability is based on different eXplainable Artificial Intelligence (XAI) tools for credit risk prediction. In particular, LIME [96] explains individual predictions by approximating the black-box model with a linear model, Anchors [97] provides high-precision human-interpretable rules (if-then conditions), SHAP [98] assigns an importance value to each feature in a classifier, BEEF [99] uses a probabilistic interpretation of the classifier's predictions and LORE [100] generates local explanations using decision rules that assign weights to each feature. These tools have been used to explain the results of credit risk prediction [101, 102, 103, 104, 105, 106, 93, 107, 108, 109]. Still, they only calculate the contribution of each feature to the prediction of a new instance. None of these tools used pattern-based prediction and explanation.

In [110], the authors propose an approach to credit risk prediction using fuzzy rule-based classifiers (FRBCs). These FRBCs are constructed using multi-objective evolutionary optimization algorithms and consist of linguistic fuzzy classification rules. These authors define interpretability as the capacity to produce a compact and understandable explanation based on selecting the most important input features. The authors used a measure of interpretability that depends on the average length of the rules, the total number of features involved in each rule, and the total number of fuzzy rules considered to predict the credit risk of a new instance. FRBC uses the most important features to explain the prediction of the result of a new instance by assigning a weight to each feature.

The authors of [111] propose to select features using the Weight of Evidence (WOE) and the Information Value (IV) [112]. After, the tabular dataset (considering only the selected features, where the columns represent the selected features and the rows represent the bins obtained with WoE) is considered a greyscale image, enabling the application of 2D CNNs for credit risk prediction. To explain this in human terms, the authors use the result of applying WoE, where each bin corresponds to a pixel and is related to the segmentation performed on the feature. These superpixels represent the most important features for predicting a class in white. Explainability is defined as the ratio of these superpixels between each feature and the segmentation obtained with WoE. However, this explanation is not clear to a non-expert in deep learning, as they should understand the segmentation result of the WoE feature selection method and how the transformation is performed on a greyscale image.

The authors of [113] combined a Light Gradient Boosting Machine (LightGBM) as a classifier using SHAP to explain the prediction result. The features are dropped based on a backward feature selection procedure. To explain the prediction result for a new instance, the authors consider high feature values to contribute significantly to the prediction, while low feature values are considered less important. For the authors, these values are seen as the explanation of the prediction result, but can also be interpreted as the importance of each feature in the prediction.

In [114], the authors present a new method called CreditNetXAI for predicting credit card default using deep learning and explainable artificial intelligence (XAI) tools. The authors employed explainability-related metrics such as fidelity, defined as the mean of precision and recall; stability, where two classifiers trained on different datasets produce the same prediction; and consistency of feature importance, using SHAP to generate feature importance for each prediction. To explain the prediction of a new instance, the authors decided to provide explanations based on the contribution of each feature.

In [115], the authors present a credit default prediction model using a multilayer dynamic graph neural network. The prediction generates a probability of default for each instance based on their features and the evolution of their relationships over time. The authors have used the Shapley values and the features' importance to explain the classifiers' prediction.

2.3 Discussion

In this review of state-of-the-art credit risk prediction, many papers focus on selecting relevant features and facing the imbalance problem using public and private datasets. They show that balancing and selecting features allow for better performance for credit risk prediction when evaluated by different classifiers used in the state-of-the-art.

Table 2 summarizes the state-of-the-art algorithms, focusing exclusively on pattern-based algorithms and those that address the explainability of credit risk prediction. This table also includes the expected characteristics of the algorithm that will be developed in this research, which is included in the last row of the table. The first column presents the name of the algorithm and the reference; the second column shows the name of the algorithm used on the predictions; the third column shows the types of data used for the prediction; the fourth column shows whether it uses a feature selection methods; the fifth column shows whether it uses a balancing method; and the last column shows whether it explains the credit risk prediction.

Table 2: Comparison with previous work.

Published work	Based on	Preprocessing dataset	Feature selection	Balancing datasets	Prediction explainability
Measure of similarity [85]	Patterns	Numerical	Yes	No	No
Multi-Objective Evolutionary Fuzzy [86]	Patterns	Numerical	No	No	No
SVDD [87]	Patterns	Numerical	No	No	No
Behavior2Shapelets [90]	Patterns	Numerical	Yes	Yes	No
FRBCs [110]	Fuzzy rule-based classifiers	Mixed	No	No	No
WoE and IV using 2D CNNs [111]	Deep Learning-Based	Numerical	Yes	No	No
LightGBM and SHAP [113]	LightGBM classifier	Numerical	Yes	No	No
CreditNetXAI [114]	Deep learning	Numerical	No	No	No
GAT-LSTM-ATT [115]	LSTM	Numerical	No	No	No
Proposal	**Patterns**	**Mixed**	**No**	**Yes**	**Yes**

Although Behavior2Shapelets [90] is based on behavior patterns, it only focuses on the behavior of three features: verifying if "Previous payment history" shows a sudden decrease, "Invoice statement amount" shows a steady increase, and "Previous payment amount" shows no significant change. These three behaviors history constitute a pattern of the default behaviour, i.e. when instances show this behavior, they have a higher probability of defaulting in the following month. However, Behaviour2Shapelets only checks these patterns and does not search for others. In contrast, in [85], they define behavioural patterns as the result of analysing the values taken by the features of the defaulters and non-defaulters within each cluster. In [87], a pattern in the clusters refers to the similarity of the data structure within the clusters. In [86], fuzzy rules are used as the patterns obtained in the training set, but these patterns are only used for understanding the data and not for explaining the predicted result of new instances. As observed, the four papers claim to use patterns. However, these patterns are not employed in the prediction of credit risk or for explaining the prediction result.

Although [111] discusses the explainability of the prediction, it is based on images, making it difficult to determine which features are involved in evaluating credit applications. In contrast, studies such as [110, 61, 86] demonstrate the potential of fuzzy rules to predict credit risk. However, only [110] addresses prediction explainability by assigning weights to fuzzy rules, those with higher weights form the prediction of the result. In contrast, [61, 86] do not address the problem of explaining prediction results.

Furthermore, [116, 107, 117] highlight the prevalence of LIME and SHAP as the most commonly used tools to explain the results of credit risk prediction. Machine learning is a significant advancement in credit risk prediction, improving real-time prediction and processing, but also requiring good explanations of the results to allow a fair and transparent application of classifiers.

As we can see, the reviewed algorithms do not provide a clear or easily understandable explanation of the results to those involved in credit risk prediction. Moreover, although some claim to use patterns, they only analyze the values of the features and the frequency of some values among the instances of defaulters and non-defaulters. Consequently, none of the reviewed works provide an understandable explanation of the prediction. Moreover, none of them use patterns to predict and explain prediction results. Therefore, it is important to propose new algorithms to solve these problems in the financial sector, particularly in credit risk prediction.

3 Preliminary Results

In the literature, see **Section 2**, most credit risk prediction works do not explain the prediction. Those that do often provide a global prediction explanation based on feature weighting, but not for a specific prediction.

Since pattern-based classifiers have reported good results, they also benefit from being explainable through their patterns. As a preliminary result, we propose an explicable algorithm for credit risk prediction based on patterns. As a first approach, we propose to address this in two stages. First, the credit risk prediction will be based on patterns mined from defaulters and non-defaulters instances data. Then, the prediction of a new instance will be explained based on the patterns covering it so that the explanation will be fitted to the instance instead of a general explanation as usual by the state-of-the-art works.

3.1 A new explainable algorithm for credit risk prediction based on patterns

The proposed algorithm for credit risk prediction and explanation first applies a pattern-based classifier for mixed and imbalanced data to make a prediction; and then uses patterns to explain the prediction. As mentioned earlier, credit risk prediction datasets are imbalanced. Thus, a previous balancing was performed. To define which balancing method to apply, we evaluate different oversampling and undersampling methods using traditional classifiers commonly used for credit risk prediction. These experiments are shown in **Section 3.2.1**. For the prediction, we evaluated the performance of several pattern-based classifiers on imbalanced and mixed datasets in the context of credit risk prediction, to identify the best classifier. The experiments for selecting the classifier used for prediction are shown in **Section 3.2.2**. It involves training a classifier, based on patterns designed to handle imbalanced and mixed datasets and classify new instances as defaulters or non-defaulters.

In our proposal, we will use patterns to explain the prediction. However, the number of patterns that can be mined using the entire dataset would be too high to make clear explanations. Therefore, we propose to create clusters within each class (for now we use K-modes [118, 119]; in the future, we will study the use of other clustering algorithms that allow the use of mixed data), followed by a pattern mining and filtering process applied separately to each cluster, to reduce the number of patterns used in constructing an explanation for the prediction.

So, first, a new instance is predicted through the pattern-based classifier. To explain the prediction, we determine which cluster in the class predicted is most similar to the predicted instance, and the patterns of this cluster that cover the new instance will be used to explain the prediction

Figure 1 shows the training process of the explainable algorithm we propose as a preliminary result of this PhD research.

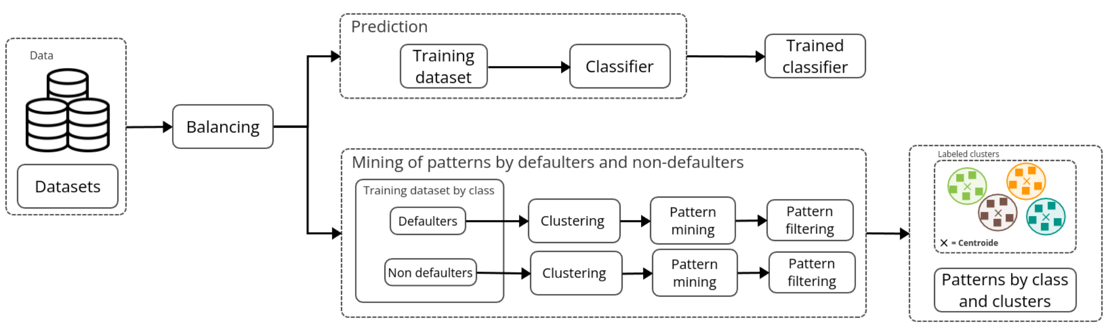


Figure 1: Proposed algorithm training

Figure 2 shows the process of prediction and explanation. The classifier predicts the new instance, and the similarity of the new instance to each centroid of the identified clusters of the predicted class is determined. The k patterns extracted from the most similar cluster that covers the new instance are used to explain the prediction.

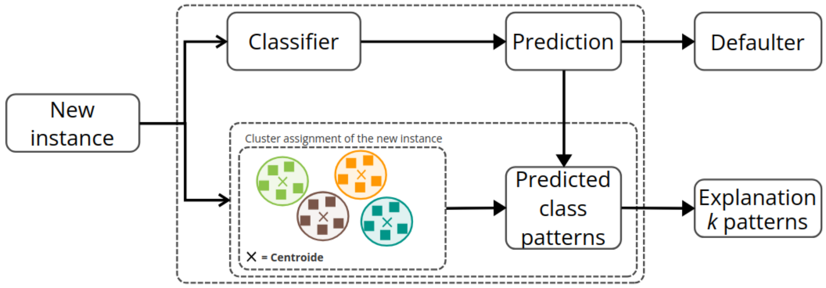


Figure 2: Prediction and explanation of a new instance using our proposed algorithm.

3.2 Experiments

To evaluate our explainable algorithm for pattern-based credit risk prediction, we first assessed the impact of different balancing methods on the dataset in **Section 3.2.1**, as described in **Section 1.6**, item **II.i.b**. This evaluation focused on mitigating the class imbalance and selecting the best method to balance a dataset for credit risk. In **Section 3.2.2** we compared pattern-based classifiers to choose the most suitable one for prediction. In **Section 3.2.3**, we evaluated the prediction quality of our algorithm by comparing its results with state-of-the-art credit risk predictors. Finally, in **Section 3.2.4**, we contrasted the pattern-based explanations generated by our algorithm with the explanations of other credit risk predictors reported in the literature.

3.2.1 Evaluation of balancing methods

The datasets associated with credit risk prediction are often imbalanced (as shown in **Table 3**), thus, balancing and instance selection methods are used to balance the datasets.

Table 3: Description of the imbalanced and mixed datasets used to predict credit risk.

Dataset	Instances	Features	Numerical/Categorical	No-defaulter	Defaulter	IR	Domain
German [120]	1,000	20	7/13	700	300	2.33	Public
Australia [121]	690	14	6/8	307	383	1.24	Public
Japan [122]	690	15	6/9	307	383	1.24	Public
Credit Card Econometrics [123]	1,319	11	7/4	1,023	296	3.45	Public
The Home Equity dataset [124]	5,960	12	10/2	4,771	1,189	4.01	Public
EIZ [5]	5,510	22	13/9	2,295	3,215	1.40	Private

We compared state-of-the-art balancing and instance selection methods commonly employed in credit risk prediction to determine which method to use before applying our proposed algorithm. **Figures 3, 4** and **5** show the results of this comparison in terms of ACC, AUC, and type II error. The reported data represent the average performance across all datasets evaluated with different classifiers. Based on these results, the balancing method achieved a similar average ACC and AUC in this experiment, while SMOTENC [12] obtained the highest average ACC and AUC. Additional results for other quality measures are available at: https://ccc.inaoep.mx/~ariel/riskPrediction/.

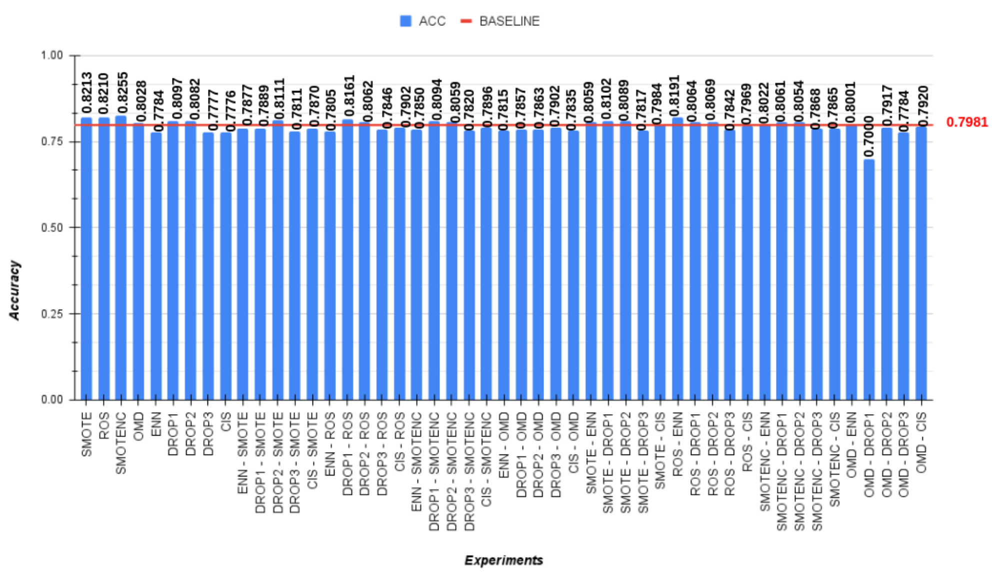


Figure 3: Accuracy of all the methods evaluated.

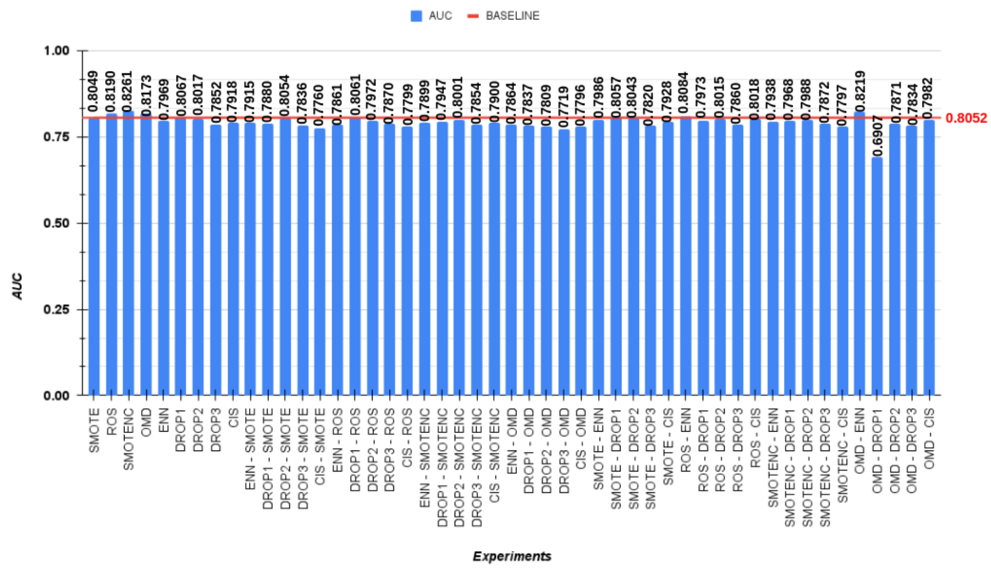


Figure 4: AUC of all the methods evaluated.

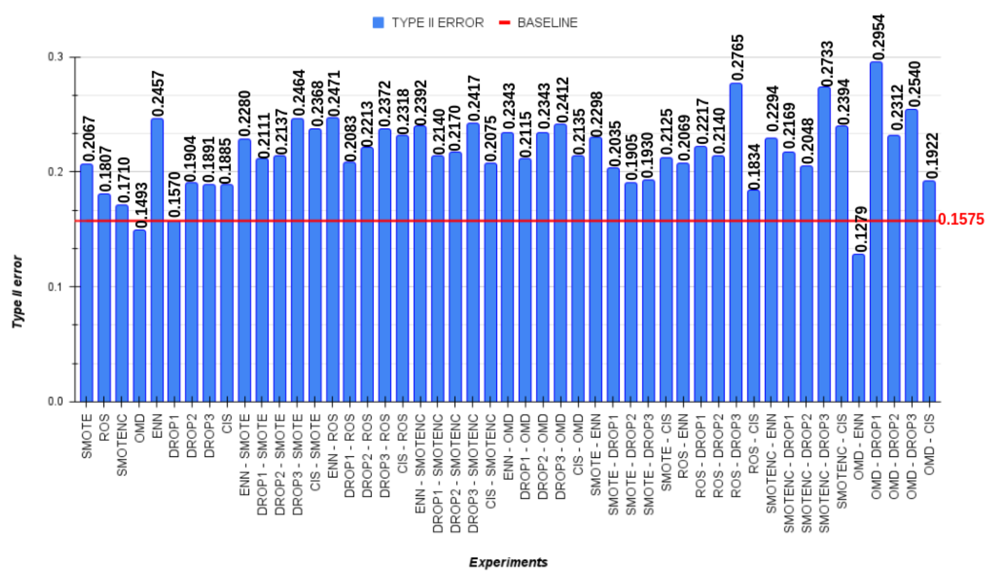


Figure 5: Type II error of all the methods evaluated.

Regarding type II error (see **Figure 5**), the lowest error was obtained by applying the OMD oversampling method and the ENN instance selection method.

3.2.2 Evaluation of pattern based classifiers on credit risk prediction

Several pattern-based classifiers have been proposed in the literature. Although these algorithms are for imbalanced databases, they tend to produce results biased towards the majority class when evaluated on credit risk prediction databases. Therefore, we apply SMOTENC, the best balancing option according to the results shown in **Section 3.2.1**.

This experiment aims to assess the performance of pattern-based algorithms on credit risk databases. As it is common in the literature, ACC, AUC, Type II error and 10-fold cross-validation were used in this experiment. Four pattern-based classifiers were employed in the experimental process, PbC4cip [7], PbCA [8], CSPmCASCP [10] and FT4cip [9]. These classifiers were used because they report good results compared to other pattern-based classifiers.

To select the classifier to be used in our proposed algorithm, Friedman's test was used to compare the results, as suggested by [125], and allowed us to select the classifier for our proposal. Post-hoc results are presented using Critical Difference (CD) plots, which provide a concise visualization of the comparative performance of the evaluated classifiers. In a CD plot, the classifier on the right is the best-performing classifier. **Figures 6, 7**, and **8** show the ranking of the pattern-based classifiers evaluated without applying the SMOTENC balancing method, using ACC, AUC and Type II error as the performance measures on the datasets in **Table 3**.

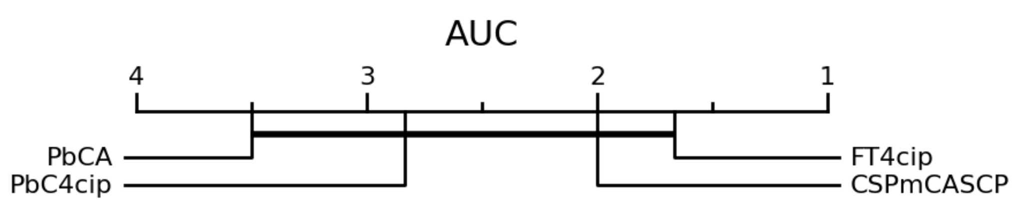


Figure 6: CD diagram of the AUC of the evaluated pattern-based classifiers.

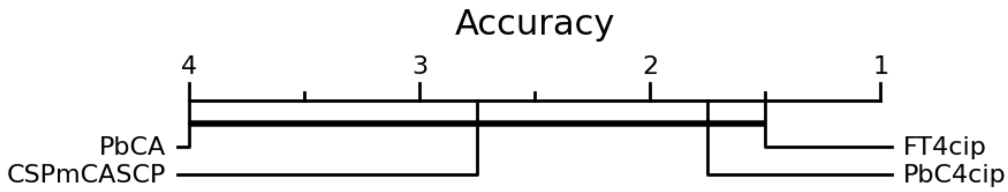


Figure 7: CD diagram of the ACC of the evaluated pattern-based classifiers.

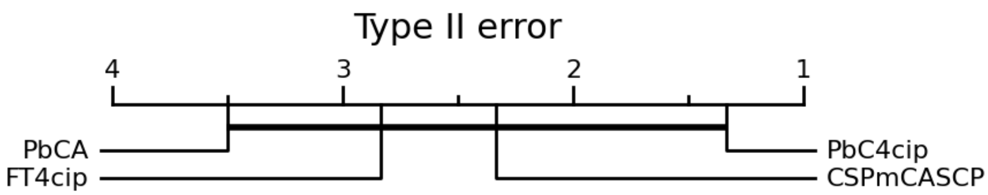


Figure 8: CD diagram of the Type II error of the evaluated pattern-based classifiers.

As it can be seen in **Figures 6** and **7**, although there is no statistical difference, the algorithm with the highest rank is FT4cip, which is used in our proposed credit risk prediction algorithm. **Figure 8**, the FT4cip and PbCA classifiers have the highest error rates, indicating poorer performance in this context. In contrast, the PbC4cip and CSPmCASCP classifiers had the lowest type II error; thus, PbC4cip patterns will be used to explain the prediction in our proposal. On the other hand, **Figures 9, 10**, and **11** show the performance of the classifiers with and without the SMOTENC balancing method. Although these classifiers consider the imbalances of the datasets, they perform better when the databases are balanced using performance measures such as ACC, AUC and Type II error.

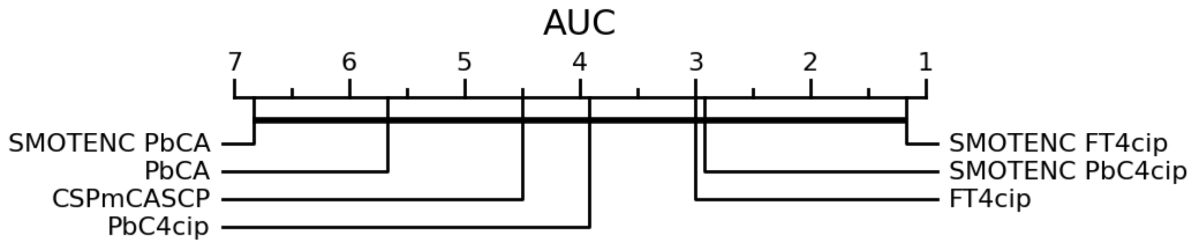


Figure 9: CD diagram of the AUC of the evaluated pattern-based classifiers with and without using SMOTENC.

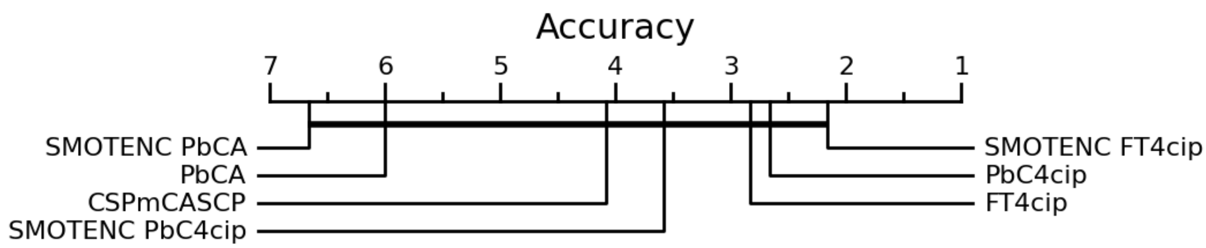


Figure 10: CD diagram of the ACC of the evaluated pattern-based classifiers with and without using SMOTENC.

As seen in **Figures 9** and **10**, balancing using the SMOTENC method with the FT4cip classifier gives the best performance in terms of AUC and ACC. This indicates that FT4cip has a good discriminative ability even without balancing, but it can benefit from handling imbalanced datasets using SMOTENC.

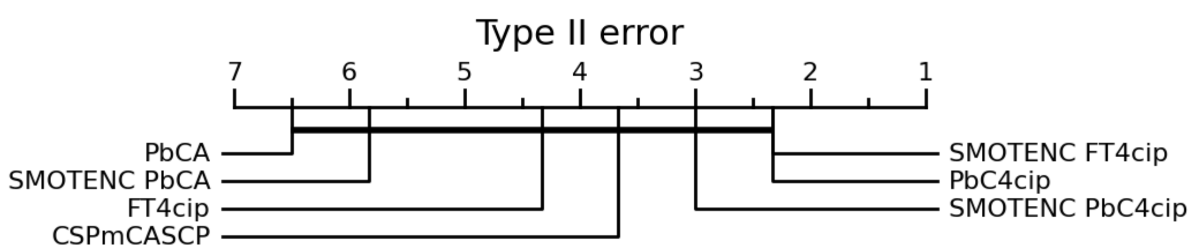


Figure 11: CD diagram of the Type II error of the evaluated pattern-based classifiers with and without using SMOTENC.

Figure 11 shows that the two classifiers with the lowest error are PBC4cip and FT4cip with SMOTENC, indicating that they are reliable and to reduce the type II error in classifying defaulters and non-defaulters.

3.2.3 Evaluation of the proposed algorithm

To assess the prediction results of our proposed algorithm, we employed the datasets shown in **Table 3**, which are commonly used in the literature for evaluating credit risk prediction. The quality prediction was evaluated using the ACC, AUC, and Type II error, the quality measures most used in the literature for credit risk prediction.

The prediction results of our proposed algorithm were compared against a DT, SVM, RL, XGBoost, ANN, RF, Fuzzy [86] and CNNs [111], all of them widely used for credit risk prediction reported in the literature. The average results for all datasets are shown in **Table 4**. This table shows a comparative analysis of the performance of various state-of-the-art classifiers, including the proposed algorithm, using ACC, AUC and Type II error as quality measures. The process has been carried out under the same considerations as the reported works. Meanwhile, [86] and [111] were implemented according to the paper's authors.

Table 4: Proposed algorithm in comparison with the state of the art.

State of the art	Accuracy	AUC	Type II error
Fuzzy [86]	0.7849	0.7582	0.3462
CNNs [111]	0.8536	0.8468	0.1967
DT [15, 53]	0.8342	0.8405	0.1722
ANN [21, 68, 126, 127]	0.8206	0.8751	**0.0716**
SVM [56, 30, 127]	0.8662	0.8601	0.1328
RL [56, 43, 127]	0.8566	0.8496	0.1501
XGBoost [2, 20, 68, 127]	0.8690	0.8752	0.1061
RF [2, 128, 129, 43, 130]	**0.8833**	**0.8912**	0.1067
Our proposal	0.8739	0.8364	0.2170

Table 4 shows that the classifiers RF (0.8833) and our proposal (0.8739) have the higher ACC. According to AUC, RF (0.8912) is the best in this metric. The proposed algorithm (0.8364) has a lower AUC and a higher type II error. On the other hand, ANN has the lowest type II error, meaning that it misclassifies a few defaulters as non-defaulters.

3.2.4 Contrasting explanations

To explain the prediction of our proposal, we use patterns as combinations of values that appear often in the data. This approach makes the prediction results easily understandable for experts, customers, and regulatory entities.

For example, in the German dataset, our proposed algorithm identifies the following instances as belonging to cluster number 2: [A11, 36, A32, A42, 3959, A61, A71, 4, A93, A101, 3, A122, 30, A143, A152, 1, A174, 1, A192, A201]; provides the following three patterns (P1, P2, and P3) to explain the prediction as a defaulter. See the repository for more database details, such as descriptions of features and additional experiments https://mega.nz/folder/1plFAQzR#SlPiHGDHSkXffskEVwcahw.

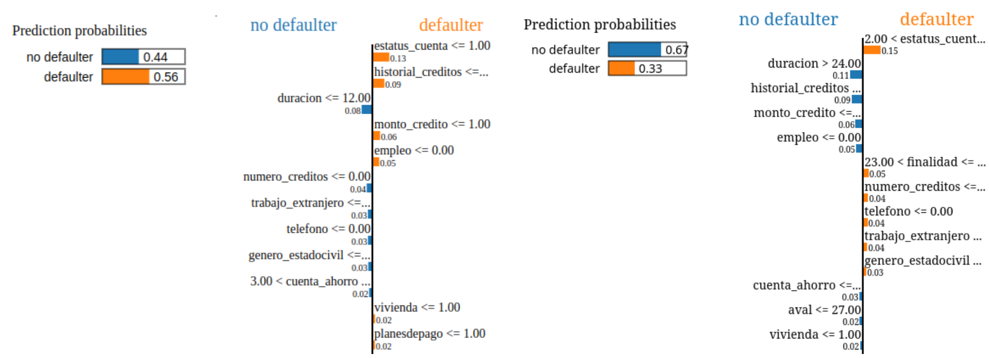
- P1 = historical credits ! = A34 AND credit amount <= 7,824.00 AND duration > 28.00 AND interest rate > 2.00 AND number of loans <= 1.00
- P2 = historical credits ! = A34 AND employment ! = A75 AND duration > 28.00 AND account status = A11
- P3 = credit amount <= 7,824.00 AND historical credits ! = A34 AND age <= 36.000 AND duration > 33.00 AND gender and marital status = A93 AND account status = A11

Considering the pattern with the highest support we can build an explanation of the prediction based on this pattern as follows:

- Given the pattern P1 = historical credits ! = A34 AND credit amount <= 7,824.00 AND duration > 28.00 AND interest rate > 2.00 AND number of loans <= 1.00
- The explanation would be as follows: The algorithm has classified the instance as a **'Defaulter'** because the person has a credit history other than A34; the amount requested is less than or equal to $7,824.00; the term is greater than 28 months; the credit has an interest rate greater than 2; and the person has one active credit with the institution.

As shown, patterns can be used to build a specific explanation for new instances, providing interpretable explanations for the specific prediction. Using patterns to explain predictions allows more specific and easy-to-understand explanations for end users. On the other hand, SHAP and LIME help to understand the internal behaviour of classifiers in determining the weighting of features. However, they do not help explain a decision for customers, as credit institutions must explain their decisions to their customers.

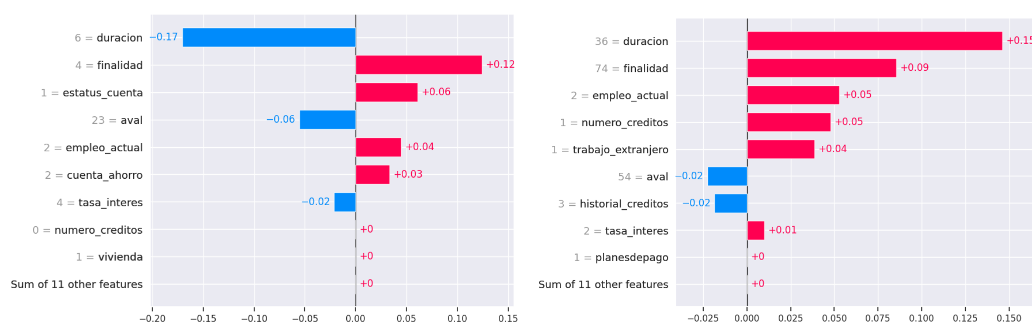
In contrast, tools such as SHAP and LIME focus on explaining the relative importance of individual features (see **Figures** 12 and 13), providing information on which features contribute most to the prediction of each class. The explanations for these tools can be difficult to understand for non-technical audiences. Furthermore, they describe the importance of features at the class level but do not explain a particular decision.



(a) Defaulter values

(b) Non-defaulter values

Figure 12: Explanation of prediction by feature weighting using LIME.



(a) Defaulter values

(b) Non-defaulter values

Figure 13: Explanation of prediction by feature weighting using SHAP.

The LIME and SHAP tools provide valuable information about the contribution of each feature to the classifier predictions. By assigning interpretable weights to each feature, they help to identify the most influential features in the classifier prediction. Therefore, these tools analyze the contributions of the features and generate a vector of weights in the form of an explanation.

4 Final Remarks

In this proposal, our objective is to develop a pattern-based credit risk prediction algorithm that can explain its predictions. We propose a first explainable credit risk prediction algorithm as a preliminary result. Experimental results show that the proposed algorithm is similar in prediction quality compared to current state-of-the-art credit risk predictors. Moreover, as we have shown, our proposal provides an interpretable explanation for specific predictions. These results motivate and demonstrate that achieving the proposed objectives of this PhD research is feasible in the times expected by the Coordination of Computer Science.

References

[1] A. M. Castro and J. A. M. Castro, *Crédito y cobranza*. Grupo Editorial Patria, 2014.

[2] I. Emmanuel, Y. Sun, and Z. Wang, "A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method," *Journal of Big Data*, vol. 11, no. 1, p. 23, 2024.

[3] J. E. Baiden, "The 5 c's of credit in the lending industry," *Available at SSRN 1872804*, 2011.

[4] A. Bhattacharya, S. K. Biswas, and A. Mandal, "Credit risk evaluation: a comprehensive study," *Multimedia Tools and Applications*, vol. 82, no. 12, pp. 18217–18267, 2023.

[5] C. N. B. y de Valores, "Disposiciones de carácter general aplicables a las instituciones de crédito." https://www.cnbv.gob.mx/Normatividad/Disposiciones%20de%20car%C3%A1cter%20general%20aplicables%20a%20las%20instituciones%20de%20cr%C3%A9dito.pdf, 2024.

[6] Markelle Kelly, Rachel Longjohn, Kolby Nottingham,, "The UCI Machine Learning Repository, https://archive.ics.uci.edu," 25 April 2024.

[7] O. Loyola-González, M. A. Medina-Pérez, J. F. Martínez-Trinidad, J. A. Carrasco-Ochoa, R. Monroy, and M. García-Borroto, "Pbc4cip: A new contrast pattern-based classifier for class imbalance problems," *Knowledge-Based Systems*, vol. 115, pp. 100–109, 2017.

[8] A. E. Gutierrez-Rodríguez, J. F. Martínez-Trinidad, M. García-Borroto, and J. A. Carrasco-Ochoa, "Mining patterns for clustering using unsupervised decision trees," *Intelligent Data Analysis*, vol. 19, no. 6, pp. 1297–1310, 2015.

[9] L. Canete-Sifuentes, R. Monroy, and M. A. Medina-Pérez, "Ft4cip: A new functional tree for classification in class imbalance problems," *Knowledge-Based Systems*, vol. 252, p. 109294, 2022.

[10] O. Loyola-González, J. F. Martínez-Trinidad, J. A. Carrasco-Ochoa, and M. Garcia-Borroto, "Cost-sensitive pattern-based classification for class imbalance problems," *IEEE Access*, vol. 7, pp. 60411–60427, 2019.

[11] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: overview study and experimental results," in *2020 11th international conference on information and communication systems (ICICS)*, pp. 243–248, IEEE, 2020.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[13] H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, pp. 1322–1328, Ieee, 2008.

[14] G. Menardi and N. Torelli, "Training and assessing classification rules with imbalanced data," *Data mining and knowledge discovery*, vol. 28, pp. 92–122, 2014.

[15] M. Abdul Salam, K. M. Fouad, D. L. Elbably, and S. M. Elsayed, "Federated learning model for credit card fraud detection with data balancing techniques," *Neural Computing and Applications*, pp. 1–26, 2024.

[16] R. Qaddoura and M. M. Biltawi, "Improving fraud detection in an imbalanced class distribution using different oversampling techniques," in *2022 Interna-*

tional Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI), pp. 1–5, IEEE, 2022.

[17] A. Rangel-Díaz-de-la Vega, Y. Villuendas-Rey, C. Yáñez-Márquez, O. Camacho-Nieto, and I. López-Yáñez, "Impact of imbalanced datasets preprocessing in the performance of associative classifiers," *Applied Sciences*, vol. 10, no. 8, p. 2779, 2020.

[18] D. Yang, W. Zhang, X. Wu, J. H. Ablanedo-Rosas, L. Yang, and W. Yu, "A novel multi-stage ensemble model with fuzzy clustering and optimized classifier composition for corporate bankruptcy prediction," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 3, pp. 4169–4185, 2021.

[19] H. He, W. Zhang, and S. Zhang, "A novel ensemble method for credit scoring: Adaption of different imbalance ratios," *Expert Systems with Applications*, vol. 98, pp. 105–117, 2018.

[20] H. Xia, W. An, and Z. J. Zhang, "Credit risk models for financial fraud detection: A new outlier feature analysis method of xgboost with smote," *Journal of Database Management (JDM)*, vol. 34, no. 1, pp. 1–20, 2023.

[21] N. Ayed and K. Bougatef, "Performance assessment of logistic regression (lr), artificial neural network (ann), fuzzy inference system (fis) and adaptive neuro-fuzzy system (anfis) in predicting default probability: The case of a tunisian islamic bank," *Computational Economics*, pp. 1–33, 2023.

[22] K. Zhou, C. Zhang, Y. Yu, S. Cong, and X. Yue, "Improving smote technology for credit card fraud detection category imbalance issues," *Engineering Letters*, vol. 31, no. 4, 2023.

[23] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-smote: a new over-sampling method in imbalanced data sets learning," in *International conference on intelligent computing*, pp. 878–887, Springer, 2005.

[24] H. M. Nguyen, E. W. Cooper, and K. Kamei, "Borderline over-sampling for imbalanced data classification," *International Journal of Knowledge Engineering and Soft Data Paradigms*, vol. 3, no. 1, pp. 4–21, 2011.

[25] G. Douzas, F. Bacao, and F. Last, "Improving imbalanced learning through a heuristic oversampling method based on k-means and smote," *Information Sciences*, vol. 465, p. 1–20, 2018.

[26] G. E. Batista, A. L. Bazzan, M. C. Monard, *et al.*, "Balancing training data for automated annotation of keywords: a case study.," *Wob*, vol. 3, pp. 10–8, 2003.

[27] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD explorations newsletter*, vol. 6, no. 1, pp. 20–29, 2004.

[28] F. Sağlam and M. A. Cengiz, "A novel smote-based resampling technique trough noise detection and the boosting procedure," *Expert Systems with Applications*, vol. 200, p. 117023, 2022.

[29] T. Berhane, T. Melese, and A. M. Seid, "Performance evaluation of hybrid machine learning algorithms for online lending credit risk prediction," *Applied Artificial Intelligence*, vol. 38, no. 1, p. 2358661, 2024.

[30] M. Mahbobi, S. Kimiagari, and M. Vasudevan, "Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks," *Annals of Operations Research*, vol. 330, no. 1, pp. 609–637, 2023.

[31] K. K. Pandey and D. Shukla, "Stratified linear systematic sampling based clustering approach for detection of financial risk group by mining of big data," *International Journal of System Assurance Engineering and Management*, vol. 13, no. 3, pp. 1239–1253, 2022.

[32] M. Bakoben, T. Bellotti, and N. Adams, "Identification of credit risk based on cluster analysis of account behaviours," *Journal of the Operational Research Society*, vol. 71, no. 5, pp. 775–783, 2020.

[33] H. Zhu, M. Zhou, G. Liu, Y. Xie, S. Liu, and C. Guo, "Nus: Noisy-sample-removed undersampling scheme for imbalanced classification and application to credit card fraud detection," *IEEE Transactions on Computational Social Systems*, 2023.

[34] D. L. Wilson, "Asymptotic properties of nearest neighbor rules using edited data," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 3, pp. 408–421, 1972.

[35] N. A. Verdikha, T. B. Adji, and A. E. Permanasari, "Study of undersampling method: Instance hardness threshold with various estimators for hate speech classification," *IJITEE (International Journal of Information Technology and Electrical Engineering)*, vol. 2, no. 2, pp. 39–44, 2018.

[36] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A review of unsupervised feature selection methods," *Artificial Intelligence Review*, vol. 53, no. 2, pp. 907–948, 2020.

[37] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.

[38] Y. Xia, C. Liu, B. Da, and F. Xie, "A novel heterogeneous ensemble credit scoring model based on stacking approach," *Expert Systems with Applications*, vol. 93, pp. 182–199, 2018.

[39] J. Nalić and G. Martinovic, "Building a credit scoring model based on data mining approaches," *International journal of software engineering and knowledge engineering*, vol. 30, no. 02, pp. 147–169, 2020.

[40] I. Gasmi, S. Neji, S. Smiti, and M. Soui, "Features selection for credit risk prediction problem," *Information Systems Frontiers*, pp. 1–20, 2025.

[41] W. Hu, Y. Wu, and Z. Yang, "An analysis of credit risk prediction for small and micro enterprises," *Journal of Artificial Intelligence Research*, vol. 1, no. 2, pp. 1–14, 2024.

[42] T. Li, G. Kou, Y. Peng, and P. S. Yu, "Feature selection and grouping effect analysis for credit evaluation via regularized diagonal distance metric learning," *INFORMS Journal on Computing*, 2024.

[43] E. Duman, M. S. Aktas, and E. Yahsi, "Credit risk prediction based on psychometric data," *Computers*, vol. 12, no. 12, p. 248, 2023.

[44] I. D. Mienye and Y. Sun, "A machine learning method with hybrid feature selection for improved credit card fraud detection," *Applied Sciences*, vol. 13, no. 12, p. 7254, 2023.

[45] M. Shetabi, "Evolutionary-based ensemble feature selection technique for dynamic application-specific credit risk optimization in fintech lending," *Annals of Operations Research*, pp. 1–43, 2024.

[46] W. Bao, N. Lianju, and K. Yue, "Integration of unsupervised and supervised machine learning algorithms for credit risk assessment," *Expert Systems with Applications*, vol. 128, pp. 301–315, 2019.

[47] H. Rizkyanto and L. G. Ford, "Customer Segmentation of Personal Credit using Recency, Frequency, Monetary (RFM) and K-means on Financial Industry," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 4, 2023.

[48] I. Singh, N. Kumar, K. Srinivasa, S. Maini, U. Ahuja, and S. Jain, "A multilevel classification and modified PSO clustering based ensemble approach for credit scoring," *Applied Soft Computing*, vol. 111, p. 107687, 2021.

[49] P. K. Jadwal, S. Pathak, and S. Jain, "Analysis of clustering algorithms for credit risk evaluation using multiple correspondence analysis," *Microsystem Technologies*, vol. 28, no. 12, pp. 2715–2721, 2022.

[50] R. Chen, S. Wang, Z. Zhu, J. Yu, and C. Dang, "Credit ratings of chinese online loan platforms based on factor scores and k-means clustering algorithm," *Journal of Management Science and Engineering*, vol. 8, no. 3, pp. 287–304, 2023.

[51] S. R. Lenka, S. K. Bisoy, R. Priyadarshini, and M. Sain, "Empirical analysis of ensemble learning for imbalanced credit scoring datasets: A systematic review," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.

[52] G. Kou, H. Chen, and M. A. Hefni, "Improved hybrid resampling and ensemble model for imbalance learning and credit evaluation," *Journal of Management Science and Engineering*, vol. 7, no. 4, pp. 511–529, 2022.

[53] L. Ni, J. Li, H. Xu, X. Wang, and J. Zhang, "Fraud feature boosting mechanism and spiral oversampling balancing technique for credit card fraud detection," *IEEE Transactions on Computational Social Systems*, 2023.

[54] G. Yao, X. Hu, P. Song, T. Zhou, Y. Zhang, A. Yasir, and S. Luo, "Adafndfs: An adaboost ensemble model with fast nondominated feature selection for predicting enterprise credit risk in the supply chain," *International Journal of Intelligent Systems*, vol. 2024, no. 1, p. 5529847, 2024.

[55] Y.-C. Chang, K.-H. Chang, H.-H. Chu, and L.-I. Tong, "Establishing decision tree-based short-term default credit risk assessment models," *Communications in Statistics-Theory and Methods*, vol. 45, no. 23, pp. 6803–6815, 2016.

[56] A. R. Khalid, N. Owoh, O. Uthmani, M. Ashawa, J. Osamor, and J. Adejoh, "Enhancing credit card fraud detection: An ensemble machine learning approach," *Big Data and Cognitive Computing*, vol. 8, no. 1, p. 6, 2024.

[57] Z. Hassani, M. Alambardar Meybodi, and V. Hajihashemi, "Credit risk assessment using learning algorithms for feature selection," *Fuzzy Information and Engineering*, vol. 12, no. 4, pp. 529–544, 2020.

[58] Z. Salekshahrezaee, J. L. Leevy, and T. M. Khoshgoftaar, "The effect of feature extraction and data sampling on credit card fraud detection," *Journal of Big Data*, vol. 10, no. 1, p. 6, 2023.

[59] Y. Mou, Z. Pu, D. Feng, Y. Luo, Y. Lai, J. Huang, Y. Tian, and F. Xiao, "Cost-aware credit-scoring framework based on resampling and feature selection," *Computational Economics*, pp. 1–26, 2024.

[60] B. Li, R. Tao, M. Li, *et al.*, "Identification of Enterprise Financial Risk Based on Clustering Algorithm," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[61] F. Baser, O. Koc, and A. S. Selcuk-Kestel, "Credit risk evaluation using clustering based fuzzy classification method," *Expert Systems with Applications*, vol. 223, p. 119882, 2023.

[62] T. N. Pandey, A. K. Jagadev, S. K. Mohapatra, and S. Dehuri, "Credit risk analysis using machine learning classifiers," in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp. 1850–1854, IEEE, 2017.

[63] S. Oreski and G. Oreski, "Genetic algorithm-based heuristic for feature selection in credit risk assessment," *Expert systems with applications*, vol. 41, no. 4, pp. 2052–2064, 2014.

[64] N. Metawa, M. K. Hassan, and M. Elhoseny, "Genetic algorithm based model for optimizing bank lending decisions," *Expert Systems with Applications*, vol. 80, pp. 75–82, 2017.

[65] D. Martens, T. Van Gestel, M. De Backer, R. Haesen, J. Vanthienen, and B. Baesens, "Credit rating prediction using ant colony optimization," *Journal of the Operational Research Society*, vol. 61, pp. 561–573, 2010.

[66] M. Jisha and D. V. Kumar, "Population based optimized and condensed fuzzy deep belief network for credit card fraudulent detection," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 9, 2020.

[67] P. Danenas and G. Garsva, "Selection of support vector machines based classifiers for credit risk domain," *Expert systems with applications*, vol. 42, no. 6, pp. 3194–3204, 2015.

[68] C. Qin, Y. Zhang, F. Bao, C. Zhang, P. Liu, and P. Liu, "Xgboost optimized by adaptive particle swarm optimization for credit scoring," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–18, 2021.

[69] Z. Wang, X. Zhang, Z. Zhang, and D. Sheng, "Credit portfolio optimization: a multi-objective genetic algorithm approach," *Borsa Istanbul Review*, vol. 22, no. 1, pp. 69–76, 2022.

[70] Y. Guo, J. He, L. Xu, and W. Liu, "A novel multi-objective particle swarm optimization for comprehensible credit scoring," *Soft Computing*, vol. 23, pp. 9009–9023, 2019.

[71] Z. Zhang, J. He, G. Gao, and Y. Tian, "Sparse multi-criteria optimization classifier for credit risk evaluation," *Soft Computing*, vol. 23, pp. 3053–3066, 2019.

[72] M. Soui, I. Gasmi, S. Smiti, and K. Ghédira, "Rule-based credit risk assessment model using multi-objective evolutionary algorithms," *Expert systems with applications*, vol. 126, pp. 144–157, 2019.

[73] W. Xue, "Application of multiobjective particle swarm optimization in rural credit system," *Journal of Sensors*, vol. 2021, pp. 1–9, 2021.

[74] M. B. Seitshiro and S. Govender, "Credit risk prediction with and without weights of evidence using quantitative learning models," *Cogent Economics & Finance*, vol. 12, no. 1, p. 2338971, 2024.

[75] X. Ye, L.-a. Dong, and D. Ma, "Loan evaluation in p2p lending based on random forest optimized by genetic algorithm with profit score," *Electronic Commerce Research and Applications*, vol. 32, pp. 23–36, 2018.

[76] Y. Bai and D. Zha, "Commercial Bank Credit Grading Model Using Genetic Optimization Neural Network and Cluster Analysis," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[77] E. S. Kamimura, A. R. F. Pinto, and M. S. Nagano, "A recent review on optimisation methods applied to credit scoring models," *Journal of Economics, Finance and Administrative Science*, no. ahead-of-print, 2023.

[78] E. Dumitrescu, S. Hué, C. Hurlin, and S. Tokpavi, "Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects," *European Journal of Operational Research*, vol. 297, no. 3, pp. 1178–1192, 2022.

[79] Y. Zou, C. Gao, and H. Gao, "Business failure prediction based on a cost-sensitive extreme gradient boosting machine," *IEEE Access*, vol. 10, pp. 42623–42639, 2022.

[80] R. Y. Goh, L. S. Lee, H.-V. Seow, and K. Gopal, "Hybrid harmony search–artificial intelligence models in credit scoring," *Entropy*, vol. 22, no. 9, p. 989, 2020.

[81] T. M. Alam, K. Shaukat, I. A. Hameed, S. Luo, M. U. Sarwar, S. Shabbir, J. Li, and M. Khushi, "An investigation of credit card default prediction in the imbalanced datasets," *Ieee Access*, vol. 8, pp. 201173–201198, 2020.

[82] Y. Xia, C. Liu, Y. Li, and N. Liu, "A boosted decision tree approach using bayesian hyper-parameter optimization for credit scoring," *Expert systems with applications*, vol. 78, pp. 225–241, 2017.

[83] D. Joseph, "Estimating credit default probabilities using stochastic optimization," *Data Science in Finance and Economics*, vol. 1, no. 3, pp. 253–271, 2021.

[84] K. S. Lim, L. H. Lee, and Y.-W. Sim, "A review of machine learning algorithms for fraud detection in credit card transaction," *International Journal of Computer Science & Network Security*, vol. 21, no. 9, pp. 31–40, 2021.

[85] T. Xiong, S. Wang, A. Mayers, and E. Monga, "Personal bankruptcy prediction by mining credit card data," *Expert systems with applications*, vol. 40, no. 2, pp. 665–676, 2013.

[86] N. Gulsoy and S. Kulluk, "A data mining application in credit scoring processes of small and medium enterprises commercial corporate customers," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 3, p. e1299, 2019.

[87] T. Li, G. Kou, Y. Peng, and S. Y. Philip, "An integrated cluster detection, optimization, and interpretation approach for financial data," *IEEE transactions on cybernetics*, vol. 52, no. 12, pp. 13848–13861, 2021.

[88] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[89] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *Journal of intelligent information systems*, vol. 17, pp. 107–145, 2001.

[90] L. Yu and C. He, "A shapelet-based behavioral pattern extraction method for credit risk classification with behavior sparsity," *Advanced Engineering Informatics*, vol. 58, p. 102227, 2023.

[91] H. Zhao, C. Zhao, X. Zhang, N. Liu, H. Zhu, Q. Liu, and H. Xiong, "An ensemble learning approach with gradient resampling for class-imbalance problems," *INFORMS Journal on Computing*, vol. 35, no. 4, pp. 747–763, 2023.

[92] F. Shen, X. Zhao, G. Kou, and F. E. Alsaadi, "A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique," *Applied Soft Computing*, vol. 98, p. 106852, 2021.

[93] V. Chang, Q. A. Xu, S. H. Akinloye, V. Benson, and K. Hall, "Prediction of bank credit worthiness through credit risk analysis: an explainable machine learning study," *Annals of Operations Research*, pp. 1–25, 2024.

[94] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable machine learning in credit risk management," *Computational Economics*, vol. 57, no. 1, pp. 203–216, 2021.

[95] A. Thampi, *Interpretable AI: Building explainable machine learning systems*. Simon and Schuster, 2022.

[96] M. T. Ribeiro, S. Singh, and C. Guestrin, "" 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*, pp. 1135–1144, 2016.

[97] M. T. Ribeiro, S. Singh, and C. Guestrin, "Anchors: High-precision model-agnostic explanations," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.

[98] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[99] S. Grover, C. Pulice, G. I. Simari, and V. Subrahmanian, “Beef: Balanced english explanations of forecasts,” *IEEE Transactions on Computational Social Systems*, vol. 6, no. 2, pp. 350–364, 2019.

[100] R. Guidotti, A. Monreale, S. Ruggieri, D. Pedreschi, F. Turini, and F. Giannotti, “Local rule-based explanations of black box decision systems,” *arXiv preprint arXiv:1805.10820*, 2018.

[101] H. H. Nguyen, J.-L. Viviani, and S. Ben Jabeur, “Bankruptcy prediction using machine learning and shapley additive explanations,” *Review of Quantitative Finance and Accounting*, pp. 1–42, 2023.

[102] S. Bahoo, M. Cucculelli, X. Goga, and J. Mondolo, “Artificial intelligence in finance: a comprehensive review through bibliometric and content analysis,” *SN Business & Economics*, vol. 4, no. 2, p. 23, 2024.

[103] J. Zacharias, M. von Zahn, J. Chen, and O. Hinz, “Designing a feature selection method based on explainable artificial intelligence,” *Electronic Markets*, vol. 32, no. 4, pp. 2159–2184, 2022.

[104] G. Babaei, P. Giudici, and E. Raffinetti, “Explainable fintech lending,” *Journal of Economics and Business*, vol. 125, p. 106126, 2023.

[105] M. Ariza-Garzon, J. Arroyo, A. Caparrini, and M. Segovia-Vargas, “Explainability of a machine learning granting scoring model in peer-to-peer lending.,” 2020.

[106] G. Szepannek and K. Lübke, “How much do we see? on the explainability of partial dependence plots for credit risk scoring,” *Argumenta Oeconomica*, vol. 50, no. 1, 2023.

[107] J. Černevičienė and A. Kabašinskas, “Explainable artificial intelligence (xai) in finance: a systematic literature review,” *Artificial Intelligence Review*, vol. 57, no. 8, p. 216, 2024.

[108] M. J. Ariza-Garzón, J. Arroyo, A. Caparrini, and M.-J. Segovia-Vargas, "Explainability of a machine learning granting scoring model in peer-to-peer lending," *Ieee Access*, vol. 8, pp. 64873–64890, 2020.

[109] V. Moscato, A. Picariello, and G. Sperlí, "A benchmark of machine learning approaches for credit score prediction," *Expert Systems with Applications*, vol. 165, p. 113986, 2021.

[110] M. B. Gorzałczany and F. Rudziński, "A multi-objective genetic optimization for fast, fuzzy rule-based credit classification with balanced accuracy and interpretability," *Applied Soft Computing*, vol. 40, pp. 206–220, 2016.

[111] X. Dastile and T. Celik, "Making deep learning-based predictions for credit scoring explainable," *IEEE Access*, vol. 9, pp. 50426–50440, 2021.

[112] N. Siddiqi, *Credit risk scorecards: developing and implementing intelligent credit scoring*, vol. 3. John Wiley & Sons, 2012.

[113] P. E. De Lange, B. Melsom, C. B. Vennerød, and S. Westgaard, "Explainable ai for credit assessment in banks," *Journal of Risk and Financial Management*, vol. 15, no. 12, p. 556, 2022.

[114] F. M. Talaat, A. Aljadani, M. Badawy, and M. Elhosseini, "Toward interpretable credit scoring: integrating explainable artificial intelligence with deep learning for credit card default prediction," *Neural Computing and Applications*, vol. 36, no. 9, pp. 4847–4865, 2024.

[115] S. Zandi, K. Korangi, M. Óskarsdóttir, C. Mues, and C. Bravo, "Attention-based dynamic multilayer graph neural networks for loan default prediction," *European Journal of Operational Research*, vol. 321, no. 2, pp. 586–599, 2025.

[116] A. A. Montevechi, R. de Carvalho Miranda, A. L. Medeiros, and J. A. B. Montevechi, "Advancing credit risk modelling with machine learning: A compre-

hensive review of the state-of-the-art," *Engineering Applications of Artificial Intelligence*, vol. 137, p. 109082, 2024.

[117] Y. Chen, R. Calabrese, and B. Martin-Barragan, "Interpretable machine learning for imbalanced credit scoring datasets," *European Journal of Operational Research*, vol. 312, no. 1, pp. 357–372, 2024.

[118] Z. Huang, "Clustering large data sets with mixed numeric and categorical values," proceedings of 1st pacific-asia conference on knowledge discouvery and data mining," 1997.

[119] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data mining and knowledge discovery*, vol. 2, no. 3, pp. 283–304, 1998.

[120] H. Hofmann, "Statlog (German Credit Data)." UCI Machine Learning Repository, 1994. DOI: https://doi.org/10.24432/C5NC77.

[121] R. Quinlan, "Statlog (Australian Credit Approval)." UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C59012.

[122] C. Sano, "Japanese Credit Screening." UCI Machine Learning Repository, 1992. DOI: https://doi.org/10.24432/C5259N.

[123] W. Greene, "Available online: `https://www.kaggle.com/dansbecker/aer-credit-card-data`," 30 March 2024.

[124] O. Kinda and A. Achonu, "Building a credit scoring model for the savings and credit mutual of the potou zone (meczop)/senegal," *Consilience*, no. 7, pp. 17–31, 2012.

[125] D. G. Pereira, A. Afonso, and F. M. Medeiros, "Overview of friedman's test and post-hoc analysis," *Communications in Statistics-Simulation and Computation*, vol. 44, no. 10, pp. 2636–2653, 2015.

[126] Y. Tang, J. Ji, S. Gao, H. Dai, Y. Yu, and Y. Todo, "A pruning neural network model in credit classification analysis," *Computational intelligence and neuroscience*, vol. 2018, no. 1, p. 9390410, 2018.

[127] W. Liu, H. Fan, and M. Xia, "Credit scoring based on tree-enhanced gradient boosting decision trees," *Expert Systems with Applications*, vol. 189, p. 116034, 2022.

[128] Y. Zou and C. Gao, "Extreme learning machine enhanced gradient boosting for credit scoring," *Algorithms*, vol. 15, no. 5, p. 149, 2022.

[129] J. K. Afriyie, K. Tawiah, W. A. Pels, S. Addai-Henne, H. A. Dwamena, E. O. Owiredu, S. A. Ayeh, and J. Eshun, "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions," *Decision Analytics Journal*, vol. 6, p. 100163, 2023.

[130] A. Szwabe and P. Misiorek, "Decision trees as interpretable bank credit scoring models," in *Beyond Databases, Architectures and Structures. Facing the Challenges of Data Proliferation and Growing Variety: 14th International Conference, BDAS 2018, Held at the 24th IFIP World Computer Congress, WCC 2018, Poznan, Poland, September 18-20, 2018, Proceedings 14*, pp. 207–219, Springer, 2018.