information

Article

Feature Selection Engineering for Credit Risk Assessment in Retail Banking

Jaber Jemai [1,*,†] and Anis Zarrad [2,†]

[1] Computer and Information Systems Division, Higher Colleges of Technology, Abu Dhabi 32092, United Arab Emirates
[2] School of Computer Science, University of Birmingham Dubai, Dubai 73000, United Arab Emirates
* Correspondence: jjemai@hct.ac.ae
† These authors contributed equally to this work.

Abstract: In classification, feature selection engineering helps in choosing the most relevant data attributes to learn from. It determines the set of features to be rejected, supposing their low contribution in discriminating the labels. The effectiveness of a classifier passes mainly through the set of selected features. In this paper, we identify the best features to learn from in the context of credit risk assessment in the financial industry. Financial institutions concur with the risk of approving the loan request of a customer who may default later, or rejecting the request of a customer who can abide by their debt without default. We propose a feature selection engineering approach to identify the main features to refer to in assessing the risk of a loan request. We use different feature selection methods including univariate feature selection (UFS), recursive feature elimination (RFE), feature importance using decision trees (FIDT), and the information value (IV). We implement two variants of the XGBoost classifier on the open data set provided by the Lending Club platform to evaluate and compare the performance of different feature selection methods. The research shows that the most relevant features are found by the four feature selection techniques.

Keywords: feature selection engineering; credit risk assessment; machine learning; classification

Citation: Jemai, J.; Zarrad, A. Feature Selection Engineering for Credit Risk Assessment in Retail Banking. *Information* **2023**, *14*, 200. https://doi.org/10.3390/info14030200

Academic Editors: Rim Moussa, Jihene Rezgui, Tarek Bejaoui and Soror Sahri

Received: 24 December 2022
Revised: 12 March 2023
Accepted: 20 March 2023
Published: 22 March 2023

1. Introduction

Machine learning and data mining techniques are widely used in business, science, and technology [1]. Financial services are deeply disrupted by the emergent applications of advanced information technology tools and techniques, such as machine learning. This trend has led to the birth of Fintech where the scope is the development of artificial intelligence solutions to improve the quality and efficiency of retail or corporate banking. Among arising subfields under the umbrella of Fintech, we can cite Blockchain, e-payment, and machine learning applications. In retail banking personal loans, mortgages, and credit cards represent the main services and consist of offering credit facilities or stage payments to suitable creditworthy customers [2]. Banks and financial institutions face the risk of borrower defaults when the payment of a part of the debt is delayed or definitely unpaid. The decision of approving a loan request is generally based on a credit score derived from the financial records of the customer. A customer's financial record shows the data of the borrower, including actual and past engagements, and current and historical financial transactions, among others. Such data can be seen as the customer's features used in order to approve or deny a loan or a credit card request [3]. In this context, we propose in this paper studying the main features driving the decision of a financial analyst when replying to a loan inquiry. To do so, we design and implement a features engineering approach based on the comparison of different feature selection techniques.

A feature in data science represents the learning variable in a dataset [4]. The type of data used in machine learning can be structured, semi-structured, and unstructured. In the context of the credit scoring problem, the dataset representing the customers' attributes is

structured where the features describe current and historical information about the financial behavior of the client. The dataset encloses also the type of customer as a data feature (defaulter or not). The objective of the classification problem is to draw the relationship between the type of customer (label) and other features (descriptors) of the client as reported in the dataset [5]. Therefore, the question is how to map the feature values to a type of customer. Such a fact means that each feature used in the mapping contributes to separating good customers from bad customers (defaulters). However, the participation of features in fitting the final machine learning model varies. Some features give more information about the customer than others. An irrelevant selected feature will lead to overfitting in the learning process without a significant contribution to the performance of the model. The decision maker, i.e., the financial analyst, needs to know the most relevant features to help in approving or disapproving a request.

The literature offers a variety of feature selection approaches and techniques including univariate selection [6], recursive feature elimination [7], feature importance determination using a classification technique, such as decision trees [8], and information value [9], among others. The objective of this paper is to find the most relevant/important features to refer to in the credit risk assessment problem. We apply, firstly, different feature selection techniques on a real dataset [10], namely the UFS, RFE, FIDT, and the IV to generate their respective training and test sets. Then, we fit an XGBoost model [11] for each dataset. The performance of the obtained classifiers is compared using different metrics, such as F1, accuracy, recall, precision, and the ROC (receiver operating characteristics)–AUC (area under the curve) curves. As the aim of the paper is to find the most relevant features, we build the wordcloud and list the most important nine features used by each selector. It is worth noting that prior knowledge of the determinant of the type of customer will help the financial analyst in taking and justifying the decision on whether to approve or decline a loan request. Additionally, if a classification model has to be developed to classify the customers of the bank, then the use of the minimum relevant features in fitting the model will reduce the curse of dimensionality of the model, the training time, and the overfitting from one side. It will also improve the model's accuracy and effectiveness by reducing the noise induced by the use of irrelevant features from another side.

In this paper, we start by conducting a critical analysis of the feature engineering methods used in machine learning in general and for credit risk analysis in particular. We then implement a data science procedure to find the main features to use in order to discriminate bad from good customers. After preprocessing and cleaning the data set provided by the Lending Club platform, we call the four selected feature selection techniques to build their respective training sets. On each data set, we fit an XGBoost model. In order to evaluate the performance of the fitted XGBoost classifiers, we use different classification metrics, namely the F1 score, roc_auc_score, accuracy, precision, and recall score. The final step is to derive the main features used in different models and examine their performance to build the list of relevant features for the credit risk analysis problem. This paper is organized as follows: the next section defines the credit risk assessment problem in the context of retailing finance. Section 3 details the feature engineering process and details the used feature selection techniques named also selectors. The next section presents the implemented approach by describing the main algorithm, the description of the dataset, the preprocessing phase, and the classification and comparison of the models' performance. In Section 5, we report the research findings by presenting the features found to be relevant in the discrimination of defaulters from good customers. The last section contains the conclusions and the potential perspectives of this work.

2. Credit Risk Assessment in Retail Banking

Financial institutions, including banks, lending companies, insurances, etc., offer different services to their customers. Providing loans and credit cards, among other services, are highly risky operations, and their success is measured later by the ability of the customer to pay back their debt [12,13]. The credit scoring problem measures a customer's creditwor-

thiness before approving the loan demand. Its solution is an estimation of the customer's ability to cope with the requirement of paying back their installments [14]. The credit score is obtained based on customer behavior when dealing with their past and current engagements. Customer engagements can be the payment of different bills (electricity, water, internet, etc.), credit card balances, and previous loan installments, among others [15]. Generally, a customer with less financial engagement and timely fulfillment of their financial engagements has a higher likelihood of being a good customer who will pay future credits on time and without delay. The opposite side is more important to handle because rejecting a customer's demand is equivalent to losing them. Therefore, the risk of approving a loan should be calculated and measured using previous customer records and converted into an evaluation of the customer's default and credit worthiness. Given the fact that banks and money-lending institutions concur on the risk of borrowers defaulting in paying back their debt, the expected loss (EL) needs to be estimated/calculated. The EL is the product of the following three terms ($EL = PD * EAD * LGD$):

- Probability of default (PD): the likelihood that the borrower will fail to pay the loan.
- Exposure at default (EAD): the expected value of the loan at the time of default (how much remains to be paid?).
- Loss given default (LGD): the amount of loss if there is a default expressed as a percentage of the EAD.

Different credit scoring methods and models can be found in the literature.

1. Statistical analysis methods [16–18]. This class of techniques refers to a variety of statistical credit scoring models, including the following:
    - Linear discriminant analysis [19];
    - Logistic regression [17];
    - Naive Bayes networks [20].
2. Multicriteria methods [21,22].
3. Machine learning methods:
    - Clustering using k-nearest neighbor algorithms [23].
    - Classification algorithms [23,24].
        - Decision trees [25].
        - Genetic algorithms [26].
        - Support vector machines [27].
        - Perceptrons and neural networks based techniques [20,28–30].
        - Bagging [31,32].
        - Random forests [33,34].
        - Boosting and extreme gradient boost [35,36].

The above-mentioned machine learning methods used different feature selection techniques mainly to reduce the dimensionality of the training sets. We can see that no agreement is made on the appropriate technique to use in order to find the best features to learn from in the context of credit scoring problems. Therefore, we consider that the study of the performance of different feature selection methods is relevant. The objective is to recommend the best features to learn from when deciding on the approval or denial of loan requests.

3. Features Selection Engineering

The objective of machine learning is to fit learning models from training sets able to classify correctly seen and unseen data [13]. Nowadays, data are abundant [37] and big datasets can be found and generated from different sources. Given this curse of dimensionality, it is required to filter and select the sample of a dataset that seems to contribute more to discriminating the target values. Therefore, an feature selection step is mandatory to reduce the size of the training sets and consequently minimize the noise

induced by the use of nonsignificant features. A variety of feature selection types and techniques are available in the literature. In this paper, we use the following four techniques:

- The univariate feature selection [6] determines a vector of weight showing the strength of the relationships between each feature and the label. These techniques are based on statistical tests to calculate the weights vector. Therefore, many variants of univariate feature selections can be implemented depending on the statistics used, such as the ANOVA F-value method, chi-squared (chi^2) tests, etc. The selected features are the K features with the highest weights. In this paper, we implement a (chi^2) based univariate feature selection algorithm named *chi2UFS*.
- The recursive feature elimination [7] consists of removing the features showing less importance. The recursive selection process is based on ranking the features according to their importance in a defined model. Recursively, the model is rebuilt using the remaining features (initially all features), and the least important attributes are removed. The number of features to return is defined by the parameter K. However, it is possible to use the cross-validation technique to calculate the optimal number of features. In this paper, we implement an RFE algorithm named *lrRFECV* based on the logistics regression model, using cross-validation techniques. As a recursive algorithm, the *lrRFECV* is time consuming depending on the number of features being evaluated. Therefore, the *lrRFECV* is preceded in this paper by the calculation of the features correlation matrix and the removal of highly correlated features.
- The feature importance determination [8] named also the model-based feature importance consists of building a classifier using all available features. Then, the K most important features are selected based on the ranking of the model. The output of the feature importance technique depends mainly on the used classifier, which can be a logistics regression, Bayesian linear classifier, decision tree, random forest, support vector machine (SVM), or an XGBoost classifier. In this paper, we implement the feature importance determination technique using a decision tree classifier named *FIDT*.
- The information value [9] is a statistic that shows how well a feature (predictor) will separate between a binary target variable, such as the type of the borrowers in the credit scoring problem (good or bad). It is based on the weight of evidence of the feature. The WoE of a feature is a simple statistic showing the strength of a predictor in separating the values of a binary target. The target in the credit scoring problem is to identify good and bad borrowers, denoted respectively as g and b. The weight of evidence of feature x is given below:

$$WoE_x = \ln(g_i/b_i) \tag{1}$$

where g_i is the percent of good customers and b_i is the percent of bad customers. The information value of a feature x is calculated as follows [9]:

$$IV_x = \sum_{i=1}^{n} (g_i - b_i) * \ln(g_i/b_i) \tag{2}$$

which is equivalent to

$$IV_x = \sum_{i=1}^{n} (g_i - b_i) * WoE_x \tag{3}$$

There is no clear threshold identified to use a cut-off to select a feature or reject it. However, in the context of credit scoring, it is commonly used that a feature x is selected if $IV_x \geq 0.3$.

4. Implementation and Computational Results

The main Algorithm 1 of the features selection engineering approach is depicted below.

Algorithm 1 Main algorithm

```
dataset ← readData()
selectorsList ← ['UFS', 'RFE', 'IV', 'FIDT']
classifiersrsList ← [XGBoost]
comparisionFrame ← initComparisionFrame()
dataPreprocessing()
for slct in selectorsList do
    data[slct] ← applySelector(dataset, slct)        ▷ apply the features selection technique
    sltFeatures[slct] ← getFeatures(data[slct])      ▷ read the selected features
    data[slct] ← SMOTE(data[slct])                   ▷ apply Smote to balance the training sets
    Xtrain, ytrain, Xtest, ytest, Xval, yval ← splitData(data[slct])   ▷ split the learning set
    for clf in classifiersList do
        model ← fitModel(clf, data[sclt])
        metrics ← testValidate(model, Xtrain, ytrain, Xtest, ytest, Xval, yval)
        comparisionFrame ← updateComparisionFrame(metrics, slct)
    end for
end for
buildWordCloud(slctdFeatures)                        ▷ plot the wordcloud of the selected features
return comparisonFrame
```

4.1. Data Set

The empirical study in this paper uses a dataset provided by the repository of the Lending Club platform (http://www.lendingclub.com) accessed on 10 September 2021. The Lending Club platform is an online crowd-sourcing and lending platform in the United States. The Lending Club platform provides quarterly-updated data sets on the status of approved loans. We choose the dataset compiled from 2007 to December 2018, composed of 1,408,575 records and 152 data columns. Seven different values give the status of a loan: current, charged off, fully paid, default, in the grace period, late (16–30 days), and late (31–120 days). We consider that current and grace period loans do not inform about the final default of a borrower, as they may default later. Therefore, all current and grace period loans are excluded from the data set, reducing the number of samples to 440,151. A customer is considered a good borrower if the status of their loan is fully paid. A bad customer has a loan status of charged-off, default, late (16–30), or late (31–120) . The ratio of good samples to bad samples in the data set is approximately 4:1. The structure of the original data set is shown in Table 1.

Table 1. Structure of the original data set.

Category	Number of Samples	Proportion
Good	345,844	78.57%
Bad	94,307	21.43%
Total	440,151	100%
	Probability of default PD = 0.2143	

4.2. Data Pre-Processing

1. Manual screening:
   According to the expert's opinion, some attributes are irrelevant and may not be used to determine the customer's creditworthiness. In total, 53 features are identified to be removed from the initial data set. Then, 99 elements are kept.
2. Null values handling:
   The phase of data acquisition always comes with some missing values. Null values mean that some attributes cannot be used in the prediction of the credit scoring system without adjusting their empty records. In this paper, we consider that features with

more than 50% values missing are ignored. The average of the column values will replace the remaining empty cells if they are numeric. Null values of discrete columns are replaced by their mode. The number of removed features for null values cause is 66, and the empty cells of 22 features are adjusted using their mean/mode value. After handling null values, 33 features are kept.

3. Standardization:
   Standardization is a data transformation method that scales the data to the standard normal distribution (mean equals 0 and standard deviation equals 1).
4. Encoding categorical variables:
   Categorical variables cannot be handled, as they are by machine learning algorithms. Therefore, a transformation step is required to convert categorical data into numerical data. For instance, we used the label encoding method to transform categorical features into numeric values.
5. Features correlation The data set was initially composed of 152 variables used to describe whether a customer may default or not in the payment of their loan. However, some of the features may not be important to assess a customer's default. Moreover, some features may also appear to be much more significant in separating good customers from bad customers than other features. The financial consultant review helped to find insignificant features and removed 53 columns. After the null handling process, we finished with 33 features.
   The second step in feature selection is based on the correlation between variables. For instance, highly positive or negative correlated variables seem to inform about the target variable in the same way, which means that it is enough to keep one of them. Given the correlation matrix between the 33 variables and the used threshold of 0.8, eight (8) variables were dropped.
6. Correlation to the label Given the 25 remaining features, we conducted a correlation to the label (Status) evaluation. With a threshold of 0.97, we found that some features are highly correlated with the type of borrower. Therefore, six other variables were also dropped. By the end, we found a clean dataset composed of 217,320 customers described by 19 variables.

4.3. Computational Results

After cleansing the raw lending club dataset and checking the correlation between the variables and the label, we implemented the four feature selection techniques in order to find the relevant features to assess the credit risk in retail banking. For each selector, we kept only the nine most important features and we used them to build their respective dataset for training and testing. The built datasets were then used to fit four XGBoost classifiers. It is worth noting that the XGBoost supports out-of-sample validation using k-fold cross-validation. This fact will avoid overfitting as stated in [38]. The performance of the classifiers is reported in Table 2 and Figure 1. Table 2 shows the classification performance metrics on the train and test sets. We can see that the F1 scores are slightly different for the UFS, RFE, and IV but lower for FIDT with 0.708. Furthermore, we can see that there is no clear difference between the metrics values on the training and test sets (the mean difference is 0.01). This fact illustrates the absence of overfitting or underfitting, even if the accuracy of all models is less than 0.8. The area under the curve scores are close to 0.8 for all models, except for the FIDT selector with 0.736. This fact shows that there is excellent discrimination between bad and good customers by the UFS-, RFE-, and IV-based models. It is important to mention also that the FIDT feature selection technique returns the lowest performance. The remaining techniques are quite similar in terms of accuracy, recall, and precision. We can explain this behavior by the fact that the UFS, RFE, and IV share mostly the same main features.

Table 2. Performance of 4 XGBoost models.

Model	F1	Roc_Auc_Score	Accuracy		Recall		Precision	
			Train	Test	Train	Test	Train	Test
XGBoost								
UFS	0.780	0.780	0.784	0.777	0.790	0.783	0.781	0.770
RFE	0.773	0.776	0.784	0.780	0.787	0.786	0.782	0.779
FIDT	0.708	0.736	0.718	0.708	0.718	0.709	0.717	0.711
IV	0.777	0.780	0.792	0.779	0.788	0.776	0.795	0.779

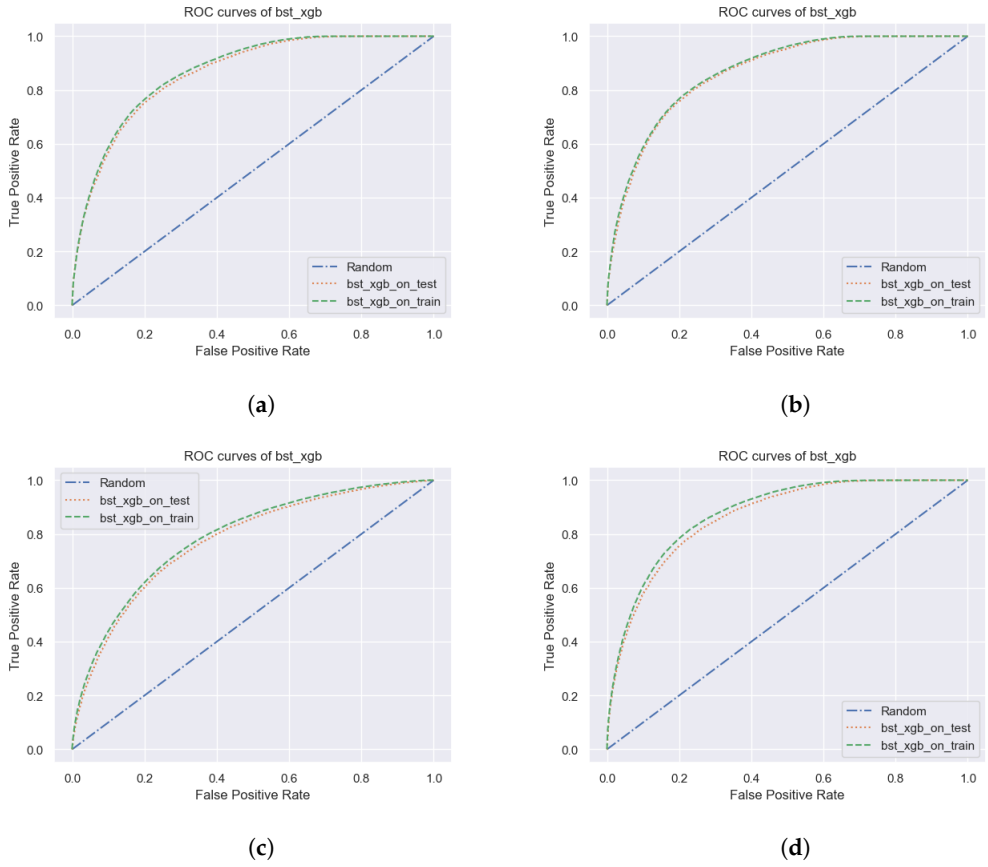

Figure 1. ROC Curves of the XGBoost models based on different features selection techniques. (**a**) UFS-XGBoost. (**b**) FIDT-XGBoost. (**c**) RFE-XGBoost. (**d**) IV-XGBoost.

As for the ROC curves, we can see that all four curves share the same shape. The training and testing curves are very close for all models; this is due to the fact that there is no big difference between the accuracy of the training and testing sets. The training and test curves of the FIDT are clearly overlapping.

5. Research Findings

The objective of this paper was to find the most relevant features to use in order to classify the type of customers and then decide on their creditworthiness. The most important features are plotted in the word cloud, Figure 2 and detailed in Table 3. The asterisk (*) in Table 3 shows that the feature was selected by the feature selection technique. Below, we detail the main four features:

1. Number of open trades in last 6 months: This feature states that if a customer tends to open more accounts (trades) in the last 6 months, then there is a high chance that they will default on their loan.

2. Interest rate on the loan: Higher interest rates increase the amount of the installments, which may push the borrower to default.
3. Balance to the credit limit on all trades: This feature informs on how much debt the customer has and how much credit they are using. A customer with a low ratio has a much greater chance of being a good borrower.
4. Number of personal finance inquiries: The number of accesses to the customer credit records made by authorized entities to check the customer's score. Such accesses are generally made based on a request made by the customer for a loan or credit card from the accessing entity. Generally, more access means that the customer is seeking more credits from different banks, increasing the probability of future defaults.

Table 3. List of selected features.

Feature Name	Feature Description	UFS	FIDT	RFE	IV
'open_acc_6m'	Number of open trades in last 6 months	*	*	*	*
'int_rate'	Interest Rate on the loan	*	*	*	*
'all_util'	Balance to the credit limit on all trades	*	*	*	
'inq_fi'	Number of personal finance inquiries	*	*	*	
'loan_amnt'	The listed amount of the loan applied for by the borrower	*	*		*
'funded_amnt'	The total amount committed to that loan at that point in time	*			*
'sub_grade'	Assigned loan subgrade	*	*		
'term'	The number of payments on the loan	*	*		
'total_cu_tl'	Number of finance trades	*	*		
'home_ownership'	The home ownership status provided by the borrower		*		
'num_op_rev_tl'	Number of open revolving accounts			*	
'num_sats'	Number of satisfactory accounts			*	
'open_acc'	The number of open credit lines in the borrower's credit file			*	
'total_bal_ex_mort'	Total credit balance excluding mortgage			*	
'total_il_high _credit_limit'	Total installment high credit/credit limit			*	
'installment'	The monthly payment owed by the borrower if the loan originates				*
'annual_inc'	The self-reported annual income provided by the borrower				*
'debt_to_income'	Percentage of the gross monthly income used to pay monthly debt				*
'revol_bal'	Total credit revolving balance				*
'application_type'	Indicates whether the loan is an individual or a joint application				*

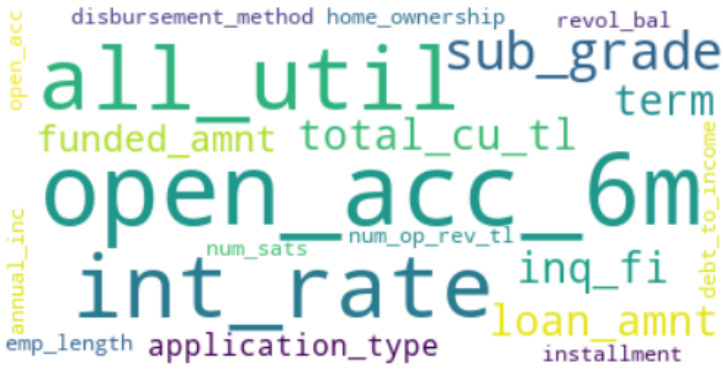

Figure 2. Wordcloud of the selected features.

As for the performance of the fitted XGBoost model for each selector, we can conclude that all feature selection techniques built slightly similar models in terms of accuracy. Although the results (predictions) of the machine learning models, such as the XGBoost classifier used in this paper, are generally unexplainable, we were able to derive the most important features shared by different selectors [39]. This finding is due to the fact the four selectors share common initial tree-splitting features as mentioned above.

Although this research was able to recommend the relevant features and variables to rely on when deciding on the loan request status, it is important to investigate other types of feature selection techniques on different data sets rather than the Lending Club data. The XGBoost classifier may be also replaced by another more evolved technique, which may improve the performance. It is worth noting that the the definition of default used in this paper is too conservative. We might classify a customer as a defaulter if they passed 90 days without paying back the installment as defined by the regulations of the Bank for International Settlements (BIS).

6. Conclusions and Future Work

For financial institutions, credit risk assessment is crucial in approving or rejecting a loan request. Banks and money-lending companies aim to discriminate between good and bad borrowers. Good borrowers are customers who are able to abide by their liabilities as scheduled. However, bad customers will default during the reimbursement of their debt. In this paper, we studied the main features that help to distinguish between good and bad customers. We implemented different feature selection techniques followed by building their respective XGBoost classifier. The first objective was to find the relevant features that the financial analyst shall observe in order to determine the possible type of borrower. Secondly, the found features can be used to fit an artificial classifier with the reduced curse of dimensionality, lesser training time, no overfitting, and higher performance. We found that the number of trades/accounts opened in the last 6 months, the interest rate, the balance of all trades to the credit limit, the number of personal finance inquiries, and the loan amount are the most important determinants of customer creditworthiness. Furthermore, we found that the performance of the XGBoost classifiers are slightly different for all selectors due to the fact that they share the same main features used to perform the first splits in their fitted decision trees.

Author Contributions: Conceptualization, J.J. and A.Z.; methodology, J.J.; software, J.J.; validation, A.Z.; data curation, J.J.; writing original draft preparation, J.J.; writing—review and editing, A.Z.; visualization, J.J.; project administration, J.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The lending club dataset can be found in Kaggle site www.kaggle.com/datasets/wordsforthewise/lending-club.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hull, J. *Machine Learning in Business: An Introduction to the World of Data Science*; Amazon Fulfillment Poland Sp. zoo: Sady, Poland, 2021.
2. Dumitrescu, E.I.; Hué, S.; Hurlin, C.; Tokpavi, S. Machine Learning or Econometrics for Credit Scoring: Let's Get the Best of Both Worlds (15 January 2021). *SSRN* **2020**. https://doi.org/10.2139/ssrn.3553781.
3. Chen, W.; Xiang, G.; Liu, Y.; Wang, K. Credit risk Evaluation by hybrid data mining technique. *Syst. Eng. Procedia* **2012**, *3*, 194–200.
4. Das, S.R. The future of fintech. *Financ. Manag.* **2019**, *48*, 981–1007.
5. Tang, J.; Alelyani, S.; Liu, H. Feature selection for classification: A review. In *Data Classification Algorithms and Application*; Chapman & Hall: London, UK, 2014; Volume 37.
6. Kar, M.; Dewangan, L. Univariate feature selection techniques for classification of epileptic EEG Signals. In *Advances in Biomedical Engineering and Technology*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 345–365.
7. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422.
8. Zien, A.; Krämer, N.; Sonnenburg, S.; Rätsch, G. The feature importance ranking measure. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Bled, Slovenia, 7–11 September 2009; Springer: Berlin/Heidelberg, Germany, 2009; pp. 694–709.
9. Lund, B.; Brotherton, D. Information Value Statistic. In Proceedings of the MWSUG 2013, Columbus, OH, USA, 22–24 September 2013.
10. Lending Club Platform. 2021. Available online: www.kaggle.com/datasets/wordsforthewise/lending-club (accessed on 10 September 2021).
11. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
12. Berg, T.; Burg, V.; Gombović, A.; Puri, M. On the rise of fintechs: Credit scoring using digital footprints. *Rev. Financ. Stud.* **2020**, *33*, 2845–2897.
13. Bumacov, V.; Ashta, A. The conceptual framework of credit scoring from its origins to microfinance. In Proceedings of the Second European Research Conference on Microfinance, Groningen, The Netherlands, 16–18 June 2011.
14. Boyes, W.J.; Hoffman, D.L.; Low, S.A. An econometric analysis of the bank credit scoring problem. *J. Econom.* **1989**, *40*, 3–14.
15. Thomas, L.; Crook, J.; Edelman, D. *Credit Scoring and Its Applications*; SIAM: Bangkok, Thailand, 2017.
16. Avery, R.B.; Bostic, R.W.; Calem, P.S.; Canner, G.B. Credit scoring: Statistical issues and evidence from credit-bureau files. *Real Estate Econ.* **2000**, *28*, 523–547.
17. Amaro, M.M. Credit Scoring: Comparison of Non-Parametric Techniques against Logistic Regression. Master's Thesis, Lisbon, Portugal, February 2020.
18. Dastile, X.; Celik, T.; Potsane, M. Statistical and machine learning models in credit scoring: A systematic literature survey. *Appl. Soft Comput.* **2020**, *91*, 106263.
19. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification and Scene Analysis*; Wiley: New York, NY, USA, 1973; Volume 3.
20. Lessmann, S.; Baesens, B.; Seow, H.V.; Thomas, L.C. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *Eur. J. Oper. Res.* **2015**, *247*, 124–136.
21. Liu, W. Enterprise Credit Risk Management Using Multicriteria Decision-Making. *Math. Probl. Eng.* **2021**, *2021*, *6191167*.
22. Pla-Santamaria, D.; Bravo, M.; Reig-Mullor, J.; Salas-Molina, F. A multicriteria approach to manage credit risk under strict uncertainty. *Top* **2021**, *29*, 494–523.
23. Baesens, B.; Van Gestel, T.; Viaene, S.; Stepanova, M.; Suykens, J.; Vanthienen, J. Benchmarking state-of-the-art classification algorithms for credit scoring. *J. Oper. Res. Soc.* **2003**, *54*, 627–635.
24. Louzada, F.; Ara, A.; Fernandes, G.B. Classification methods applied to credit scoring: Systematic review and overall comparison. *Surv. Oper. Res. Manag. Sci.* **2016**, *21*, 117–134.
25. Teles, G.; Rodrigues, J.J.; Saleem, K.; Kozlov, S.; Rabêlo, R.A. Machine learning and decision support system on credit scoring. *Neural Comput. Appl.* **2020**, *32*, 9809–9826.
26. Oreski, S.; Oreski, G. Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Syst. Appl.* **2014**, *41*, 2052–2064.
27. Yu, L.; Yue, W.; Wang, S.; Lai, K.K. Support vector machine based multiagent ensemble learning for credit risk evaluation. *Expert Syst. Appl.* **2010**, *37*, 1351–1360.
28. Desai, V.S.; Crook, J.N.; Overstreet, G.A., Jr. A comparison of neural networks and linear scoring models in the credit union environment. *Eur. J. Oper. Res.* **1996**, *95*, 24–37.
29. West, D. Neural network credit scoring models. *Comput. Oper. Res.* **2000**, *27*, 1131–1152.

30. Lai, K.K.; Yu, L.; Wang, S.; Zhou, L. Credit risk analysis using a reliability-based neural network ensemble model. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4132, p. 682.
31. Zhang, D.; Zhou, X.; Leung, S.C.; Zheng, J. Vertical bagging decision trees model for credit scoring. *Expert Syst. Appl.* **2010**, *37*, 7838–7843.
32. Louzada, F.; Anacleto-Junior, O.; Candolo, C.; Mazucheli, J. Poly-bagging predictors for classification modelling for credit scoring. *Expert Syst. Appl.* **2011**, *38*, 12717–12720.
33. Wang, G.; Ma, J.; Huang, L.; Xu, K. Two credit scoring models based on dual strategy ensemble trees. *Knowl.-Based Syst.* **2012**, *26*, 61–68.
34. Zhang, X.; Yang, Y.; Zhou, Z. A novel credit scoring model based on optimized random forest. In Proceedings of the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 8–10 January 2018.
35. Qin, C.; Zhang, Y.; Bao, F.; Zhang, C.; Liu, P.; Liu, P. XGBoost Optimized by Adaptive Particle Swarm Optimization for Credit Scoring. *Math. Probl. Eng.* **2021**, *2021*, *6655510*.
36. Li, H.; Cao, Y.; Li, S.; Zhao, J.; Sun, Y. XGBoost model and its application to personal credit evaluation. *IEEE Intell. Syst.* **2020**, *35*, 52–61.
37. Hurley, M.; Adebayo, J. Credit scoring in the era of big data. *Yale J. Law Technol.* **2016**, *18*, 148.
38. Andreeva, G.; Altman, E.I. The Value of Personal Credit History in Risk Screening of Entrepreneurs: Evidence from Marketplace Lending. *J. Financ. Manag. Mark. Inst.* **2021**, *9*, 2150004.
39. Bastos, J.A.; Matos, S.M. Explainable models of credit losses. *Eur. J. Oper. Res.* **2022**, *301*, 386–394.