2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG) | 979-8-3503-5815-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/SEB4SDG60871.2024.10630264

Credit Scoring Prediction Using Boruta Feature Selection with Different Sampling Techniques

Mutiu Ganiyu
Dept. of Computer Science
Federal Polytechnic, Ile-Oluji
Ondo State, Nigeria
mutganiyu@fedpolel.edu.ng

Olabisi E Johnson
Dept. of Computer Science
Federal Polytechnic, Ile-Oluji
Ondo State, Nigeria
estjohnson@fedpolel.edu.ng

Olanrewaju V. Johnson
School of Computer Sciences
Universiti Sains Malaysia
Penang, Malaysia
olajohnson@student.usm.my

***Abstract*— The Credit Scoring prediction model is crucial for effective financial risk management. Mishandling financial risk management can lead to significant losses for the organization. The risks associated with granting credit to customers with a poor credit history have consistently driven the demand for an improved prediction system compared to traditional scorecards. Machine learning (ML) techniques have been successfully utilized in credit scoring predictions. However, existing approaches do not consider an agnostic method to feature selection (important debtor's information as predictors) capable of improving credit scoring prediction. This paper, therefore, presents a Boruta-based technique for feature selection for building a robust agnostic credit scoring model. In addition, different sampling techniques, including Random Oversampling (RO), Synthetic Minority Over-sampling Technique (SMOTE), and SMOTEtomek, were applied to address data imbalance. A conditional inference tree (ctree) technique is used to train the German credit data consisting of 1000 instances with 21 features, including the target label. Performance metrics such as F1-score, specificity, and Matthews Correlation Coefficient (MCC) were primarily considered instead of accuracy to evaluate the models. The proposed model was found to outperform other models considered in this study. As more advanced models are developed, credit scoring prediction models will continue to contribute to reducing losses.**

Keywords*— *Boruta, Credit Scoring, Data imbalance, Feature Selection, Machine Learning

I. Introduction

Financial risk management is a complex issue that requires careful consideration of several vital factors [1]. It is particularly relevant for financial institutions such as banks and lending agencies, as they are exposed to various risks that can impact both their operations and their clients. One of the essential services provided by these institutions is the extension of credit to consumers. However, it is crucial to assess the creditworthiness of potential borrowers to avoid substantial losses resulting from lending to individuals who are not financially fit. To address this challenge, credit scoring (CS) has emerged as a valuable tool for banks to analyze borrowing and lending activities based on customer historical data (demographic and financial information) [2]. By this consistent and continuous information gathering from bank customers and other financial institutions, the financial institution can effectively manage financial risks and make informed decisions regarding the provision of loans [3].

Financial firms use credit scoring as a decision support system to differentiate between good and bad borrowers. According to the majority of these businesses, a good borrower is simply someone who consistently repays their loans, while a bad borrower is someone who does not. Additionally, financial firms consider good borrowers to have clean and positive records, whereas those without such records are classified as "bad borrowers." It is important to note that classifying clients as "excellent borrowers" or "bad borrowers" has been in place since the 1950s. During that time, credit assessment was determined using a 5C's judgmental method: (1) character (knowledge about the individual or family history), (2) capital (loan amount requested by the individual), (3) collateral (what the borrower is willing to offer as security), (4) capacity (repayment ability and plan), and (5) condition (market circumstances at the time) [4].

Due to the difficulty of evaluating a large number of applications daily, it is clear that this basic selection strategy does not ensure correct classification. As a result, new and accurate automated systems with lower prediction errors are employed to handle large and complex credit scoring data. Early successful credit scoring models include probability of default and scorecards. Scorecards ensure consistent and non-judgmental treatment of all borrowers. They generate a score that quantifies the risk associated with lending money to borrowers. The scorecard makes it easier to collect customer information and classify them as "good" or "bad" based on a cut-off score determined by the Kolmogorov-Smirnov statistic [4]. Meanwhile, the cut-off score strategy has led to remarkable success in Machine Learning (ML) applications for credit scoring in the literature.

However, existing studies do not consider an agnostic method for feature selection like the Boruta technique. Additionally, the CS data is characterized by imbalanced data, where the number of observations belonging to the two classes is often unequal. Therefore, this paper proposes the use of the Boruta technique for robust feature extraction, along with three different data-sampling methods: Random Oversampling (ROS), Synthetic Minority Oversampling Technique (SMOTE) [5], and SMOTEtomek. To train the German dataset, a conditional inference tree (*ctree*) is then proposed. Hence, this study provides answers to the following research questions: (1) how agonistic is the Boruta technique to feature extraction? (2) Does class distribution modification significantly improve classifier performance? (3) Which combination of data sampling methods with Boruta leads to better identification of non-credit-worthy customers?

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

The rest of this paper is structured as follows. Section 2 discusses the ML models used. In Section 3, the experimental results were discussed in detail, including exploratory analysis and empirical results, while the last section concludes the paper.

II. Literature Review

In recent times, ML techniques have been predominantly applied to CS. These techniques include Logistic Regression (LR), Random Forest (RF), Naïve Bayes (NB), Decision Trees (DT), Neural Network (NN), Support Vector Machines (SVM), and Deep Neural Networks (DNN) [4, 6-9]. Zonneveldt et al. [8] presented Bayesian network classifiers for CS prediction using German credit data to train the model. Additionally, tree-augmented naive bayes (TANs) and forest-augmented networks (FANs) were introduced to compare with the NB model. The models were evaluated based on accuracy and mean cost. It was found that NB significantly outperformed the other networks, having obtained an accuracy score of 75.4667 and a mean cost of 0.8733, respectively.

Furthermore, in their study, Zhao et al. [10] examined a neural network (NN) technique for CS. The NN model was based on multi-layer perception (MLP). The authors trained 34 models 20 times, each with different initial weights and training instances. The models included 6 to 39 hidden units and one hidden layer. They utilized a German credit dataset for their analysis. The test results and model comparison revealed a classification accuracy of 87%.

Additionally, Trivedi [11] presented a CS model that employed various feature selection and ML approaches to determine the optimal combination of techniques. The results indicated that the combination of RF-Chi-Square performed the best in terms of accuracy, F-measure, and low False Positive (FP) and False Negative Rates (FNR) compared to other combinations. This study offers financial institutions the opportunity to develop an automated model for credit scoring that is more effective. Similarly, a study in [1] explored the use of five classifiers in combination with different feature selection techniques and data-balancing methods. The study analyzed a retail credit bank dataset and found that the best combination for this dataset was RF combined with RF-recursive-feature elimination and random oversampling. Previously, Jadhav et al. [12] identified the “curse of dimensionality” as a significant challenge in ML, specifically in the field of CS. The authors introduced the Information Gain Directed-Feature Selection (IGDFS) method, which employs a genetic algorithm to rank features and select the most important ones. The authors conducted experiments using three CS datasets and three classifiers. The results showed that utilizing the extracted features from the datasets significantly enhanced the performance of the classifiers.

III. ML Models

A. Conditional Inference Tree (ctree)

An approach used in this study encompasses the integration of a *ctree* model, the Boruta technique, and three distinct sampling methods. The Boruta technique is employed for robust feature selection, enhancing the model’s ability to discern relevant variables. As a result, a comparative analysis is conducted, evaluating the performance of the *ctree* model in conjunction with the Boruta technique across the three diverse sampling methods. The aim is to investigate the effect of imbalance class distribution and determine the optimal combination of sampling methods with Boruta on the *ctree* model for CS model performance.

The *ctree* model is characterized as a regression model that articulates the conditional distribution of a response variable denoted as $\boldsymbol{Y}$, given the states of m covariates. The response variable $\boldsymbol{Y}$, derived from a sample space $\mathcal{Y}$, manifests as a multivariate entity. Meanwhile, the m-dimensional covariate vector $\boldsymbol{X} = (X_1, \dots, X_m)$ is drawn from a sample space $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$. Assuming the conditional distribution, denoted as $D(\boldsymbol{Y}|\boldsymbol{X})$, then the response variable $\boldsymbol{Y}$ depending on $\boldsymbol{X}$ by a function g is given as:

$$D(\boldsymbol{Y}|\boldsymbol{X}) = \mathrm{D}(\mathrm{Y}|(X_1, \dots, X_m) = D(\boldsymbol{Y}|g(X_1, \dots, X_m)) \quad (1)$$

Assuming data at node d has D_d with N_d samples and each candidate split $\alpha = (j, t_d)$ on feature j and cut-off t_m, then the partition subsets $D_d^{left}(\alpha)$ *and* $D_d^{right}(\alpha)$ as

$$D_d^{left}(\alpha) = \{(X, Y) | X_j \le t_m\} \quad (2)$$

$$D_d^{right}(\alpha) = \frac{D_d}{D_d^{left}(\alpha)} \quad (3)$$

The quality of a candidate split at the same node is therefore calculated using impurity function H(), for either regression or classification problem as

$$G(D_{d,}\alpha) = \frac{N_d^{left}}{N_d} H\left(D_d^{left}(\alpha)\right) + \frac{N_d^{right}}{N_d} H\left(D_d^{right}(\alpha)\right) \quad (4)$$

A stopping criterion $N_d < min_{samples}$ or $N_d = 1$ is the maximum allowable depth for recursively splitting the subsets tree further.

B. Boruta Technique

Feature selection is an important feature engineering step in ML [13]. It helps provide training models with relevant features that enhance performance in addition to understanding data, reduce computation requirements, and ease the curse of dimensionality effects. Many methods are provided in the literature, including filters, wrappers, and embedded. Each of these methods comes with its pros and cons. At times, the methods are either hybridized or ensemble to gain potentials inherent in each method.

This paper proposed a Boruta technique, which is designed as an all-relevant feature selection wrapper algorithm. It is capable of agnostically working with ample feature space, provided the categorical variables are transformed to output their feature importance measure. Its strength is based on the state-of-the-art RF model. It utilizes a top-down search for pertinent features by comparing the significance of the original characteristics with the significance that may be randomly determined using their permuted duplicates. It then gradually removes extraneous features to stabilize the test.

C. Data Sampling Methods

Imbalanced datasets are common in real-world applications such as customer retention [14], employee attrition [15], and ECG heartbeat [16]. The handling of imbalanced data significantly impacts ML model performance. Available methods to solve data-imbalanced problems in literature are classified into data-level, algorithmic-level, and cost-sensitive learning. The techniques in the data-level category used in this study are described as follows:

1) Random oversampling-ROS: This method is among the most effortless approaches. It increases the minority class by employing a random repetition of the minority class. The idea is to be or relatively be at par with the majority class instances. However, a potential drawback is a substantial increase in computing costs in the event of a huge dataset. In addition, the likelihood of overfitting may rise since it produces identical duplicates of the minority class.

2) SMOTE: This technique was proposed to create new instances, interpolating two instances of the original data set, the first one chosen randomly and the second one chosen randomly from among its closest neighbors. The method intends to address some of the problems of ROS; however, the cause of overfitting is yet challenging.

3) SMOTEtomek: This method uses the principle of hybridizing SMOTE and Tomek techniques. The dataset is first up-sampled by using SMOTE. The output is fit into Tomek after that to downsample the majority of instances. In this approach, a percentage threshold is for both the up-sampling and down-sampling pertaining to the majority and the minority.

D. Evaluation Metric

In most ML predictions, accuracy (*Acc*) is used to assess the performance of the trained model. However, in the case of data imbalanced problems, accuracy is not suitable for model assessment because ML tends to be biased towards the majority class. In comparison, the unbalanced problem focuses on the model's ability to identify the rare cases (minority class). The performance of the model in this paper was, therefore, accessed using F1-score, sensitivity (Sens), precision (Prec), geometric mean (Gmean), Mathew's correlation coefficient (MCC), and AUC. These metrics, as defined below, are computed from the confusion matrix presented in Table I.

TABLE I. CONFUSION MATRIX

		Actual	
		Non-credit-worthy (0)	Credit-worthy (1)
Predicted	Non-credit-worthy (0)	True Positive (TP)	False Positive (FP)
	Credit-worthy (1)	False Negative (FN)	True Negative (TN)

$$Acc = \frac{TP+TN}{N} \quad (5)$$

$$Sens = \frac{TP}{TP+FN} \quad (6)$$

$$Prec = \frac{TP}{(TP+FP)} \quad (7)$$

$$F1-score = 2 \times \frac{Prec*Sen}{(Prec+Sen)} \quad (8)$$

$$G-mean = \sqrt{TPR*TNR} \quad (9)$$

$$MCC = \frac{(TP*TN-FP*FN)}{sqrt((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))} \quad (10)$$

Where *N* represents the total observations.

Another metric that benefits unbalanced data model assessment is the Receiver Operating Characteristics (ROC) graph. The graph additionally supports a measure classed AUC. AUC increases and enhances model performance. As a performance metric, it combines performance across all possible categorization levels. One method of examining AUC is determining the probability that the model evaluates a random positive example more highly than a random negative one. The AUC value lies in the 0.0 to 1.0 range. Additionally, since the area under the ROC diagonal line is 0.5, models with AUC values greater than 0.5 perform better than random models, as shown in Fig. 1.

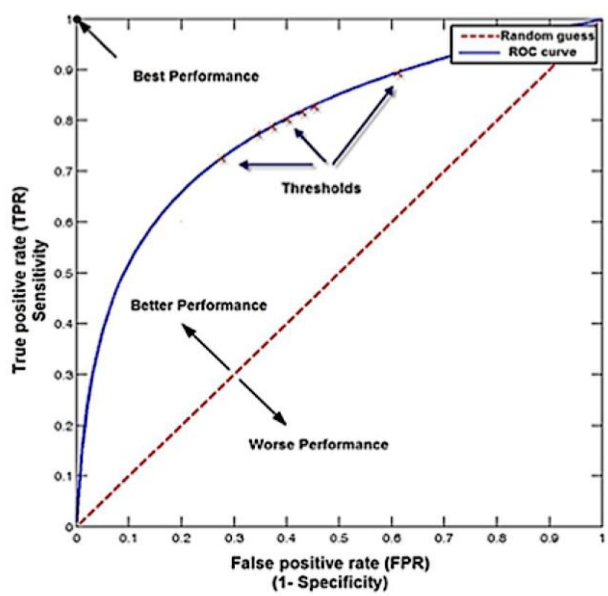

Fig. 1. The AUC Analysis for Model Performance [17]

IV. Experimental Results

This section discusses the experiment setup using R 4.2.1 with all the required packages and Rstudio 2022.07.1. The results obtained using the German credit datasets are discussed below.

A. Data Used

The experiment utilizes a German credit scoring dataset comprising 1000 instances and 21 features, incorporating the target label "credit_risk." Professor Dr Hans Hofmann provided this dataset. Originally, the dataset encompasses 14 categorical variables and 7 numerical variables, out of which contain financial, behavioral, and demographic information. The outcome (class label) describes the status of the client as a good or bad credit history. The detail is presented in Table II.

TABLE II. THE GERMAN DATASET DESCRIPTION

Features	Description	Type
Status	status of the debtor's checking account with the bank.	Categorical
Duration	credit duration in months (quantitative)	Numeric
Credit_history	history of compliance with previous or concurrent credit contracts	Categorical
Purpose	the purpose for which the credit is needed	Categorical
Amount	credit amount in DM	Numeric
Savings	debtor's savings	Categorical
Employment_duration	duration of debtor's employment with current employer	Categorical
Installment_rate	credit installments as a percentage of the debtor's disposable income	Numeric
Personal_status_sex	combined information on sex and marital status; categorical.	Categorical
Other_debtors	Is there another debtor or a guarantor for the credit	Categorical
Present_residence	length of time (in years) the debtor lives in the present residence	Numeric
Property	the debtor's most valuable property, i.e., the highest possible code is used	Categorical
Age:	Age in years	Numeric
Other_installment_plans	installment plans from providers other than the credit-giving bank	Categorical
Housing	type of housing the debtor lives in	Categorical
Number_credits	number of credits, including the current one the debtor has (or had) at this bank	Numeric
Job	definition of debtor's job	Categorical
People_liable	number of persons who financially depend on the debtor (i.e., are entitled to maintenance)	Numeric
Telephone	Is there a telephone landline registered in the debtor's name	Categorical
Foreign_worker	Is the debtor a foreign worker?	Categorical
Credit_risk: Target label	Has the credit contract been complied with (good) or not (bad)	Categorical

B. *Data Preprocessing*

Data preprocessing steps, including missing value, dummy variable encoding, and min-max normalization, were performed. Previously, it was noted that all the features in the dataset were designated as integer data types from the data source. The indication is that a significant portion of the features in the German dataset lacked proper definition. Utilizing the dataset in this state could lead to misleading results. Consequently, a preprocessing step was undertaken to align the dataset with its original information and description, using the 'as.factor()' function.

C. *Exploratory Analysis (EDA)*

The EDA helps to visualize datasets to gain some valuable insights about the data. This section provides insights into the numeric and categorical variables. It was observed that the duration of credit and loan amount have a positive skew distribution. It does appear that most of the credit amounts are relatively small, with a peak frequency for lower credit amounts (around 0 to 2,500). The frequency decreases as the credit amount increases, indicating that higher loan amounts are less common.

Similarly, most credits are of a shorter duration, with a peak in frequency at the lower end of the duration scale. It implies that the number of credits gradually decreases for longer durations, indicating that the duration of the credit is a critical factor in assessing the risk associated with a loan. Shorter durations may be associated with lower risk as the borrower is obligated for a shorter period. In comparison, longer durations may be seen as a higher risk due to the increased uncertainty over time, as shown in Figs 2 and 3. On the other hand, the age distribution suggests that the dataset contains a higher frequency of individuals in the middle age range, with a peak around the 30-40 age range. The frequency decreases for younger and older ages, implying age is an important factor, as shown in Fig. 4. The rest of the visualizations are provided in Figs. 5 - 8, showing the relation between the response label "credit_risk" and the individual predictors.

While the visualizations are essential to gain fore-insight into credit management, it is imperative to note the distribution of the response variable. The distribution of the target label "credit_risk" shows that there are 700 "good" records and 300 "bad" records, as shown in Fig. 9. The implication is that the predictive outcome "1: good" has unequal instances to "0: bad", resulting in an unbalanced problem. ML classifier tends to bias towards good outcomes if the challenge is not addressed. Whereas the objective is to be able to detect "bad borrowers." The results of the methods adopted in solving the problem are discussed in a later section.

D. *Boruta Result*

It is a known fact that not all the variables in a dataset contribute effectively to building an accurate model. Therefore, feature selection is a crucial step taken to improve model accuracy and reduce computation costs. [12].

The Boruta technique uses an RF method to generate an important variable. Boruta technique generates four importance levels, namely shadow, tentative, important, and unimportant variables. The first iteration result, as illustrated in Fig. 10, shows important features in green; the blue represents shadow features, the yellow represents tentative, and the red corresponds to unimportant variables. The tentative is treated to yield the final result, as shown in Fig. 11, with the following final features selected:

"duration", "amount", "installment_rate", "age", "number_credits", "status", "credit_history", "purpose", "savings", "other_debtors", "property", "other_installment_plans", "housing"

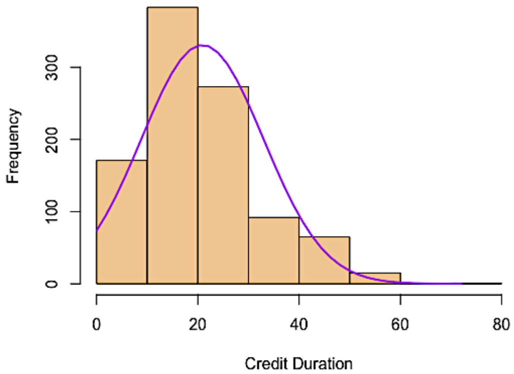

Fig. 2. Credit Duration Distribution

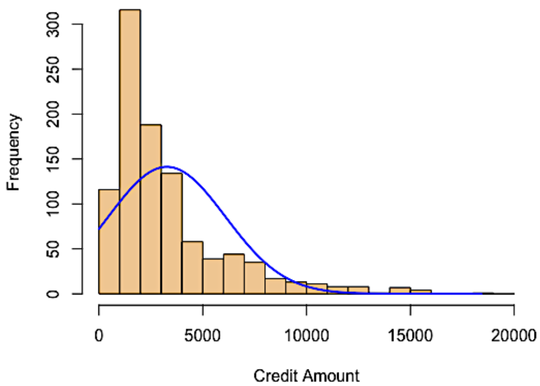

Fig. 3. Credit Amount Distribution

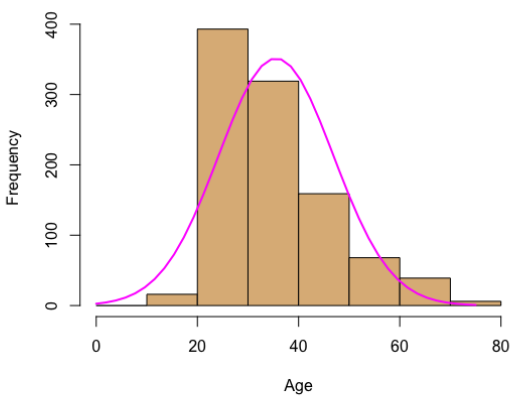

Fig. 4. Age Distribution in CS

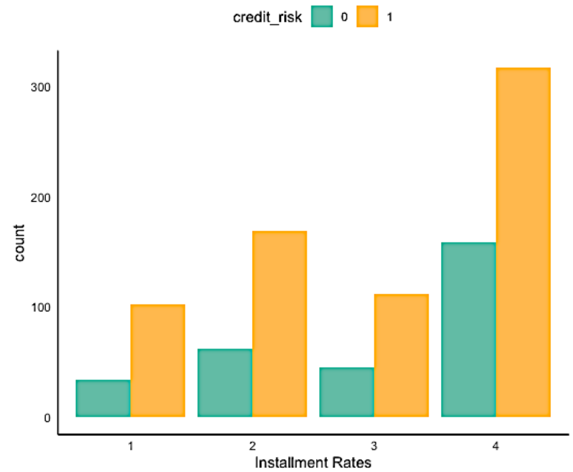

Fig. 5. Effect of Installment Rate on Credit Risk

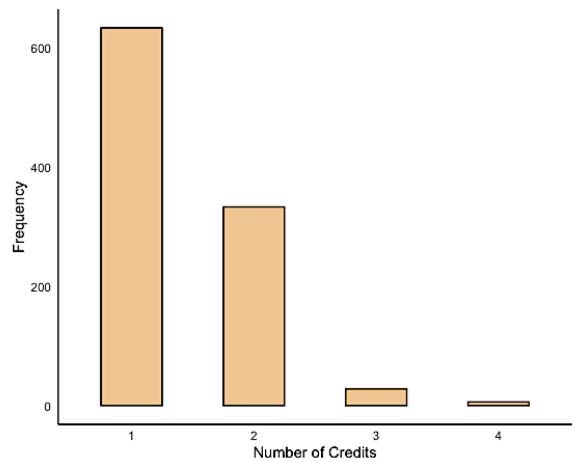

Fig. 6. Number of Credit Distribution

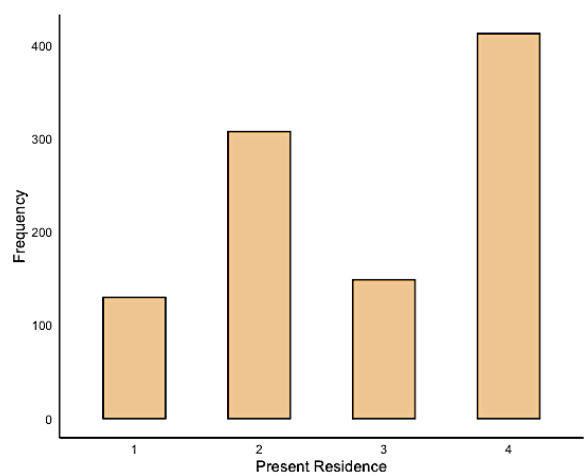

Fig. 7. Present Residence Distribution

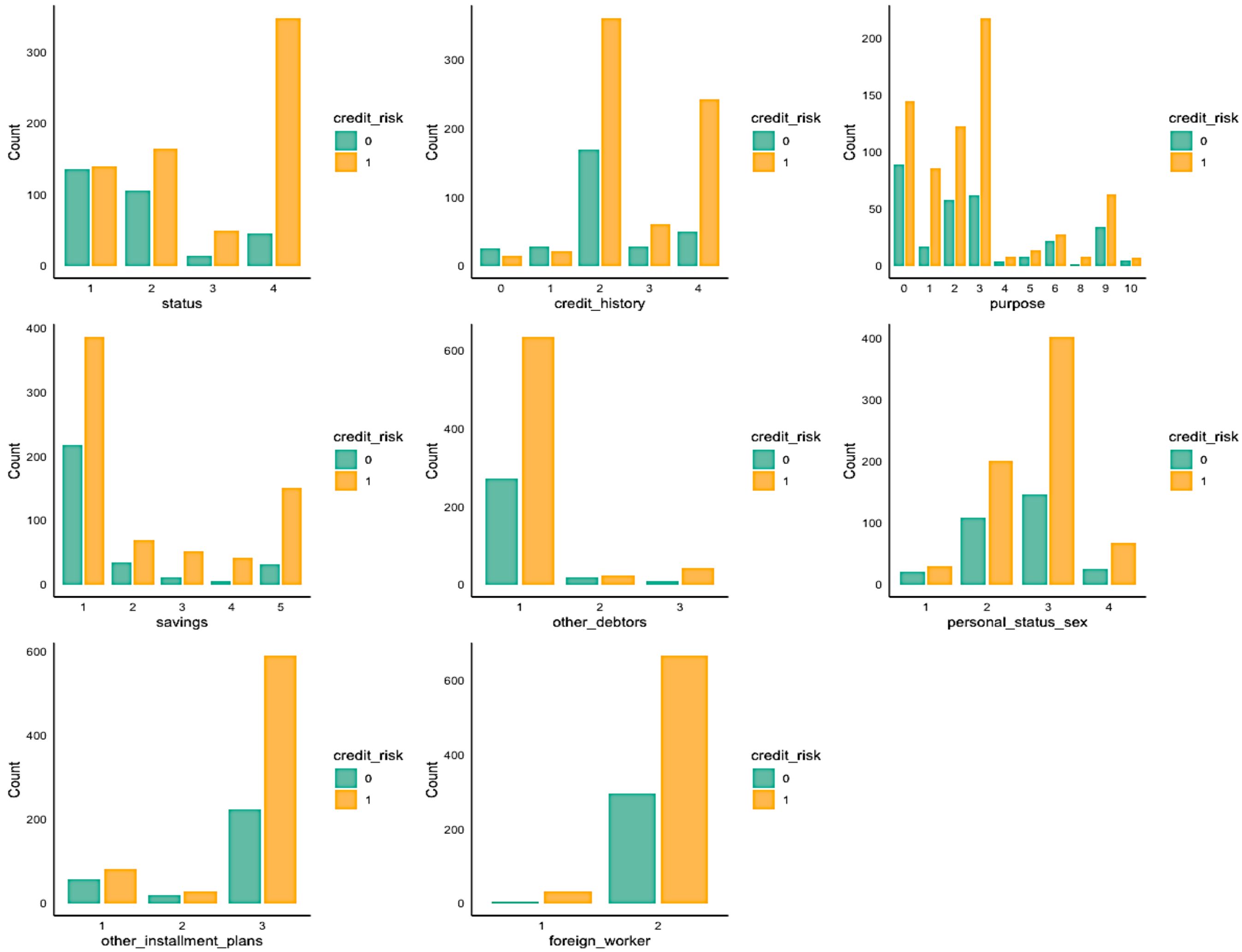

Fig. 8. Visualization Results of Categorical Variables

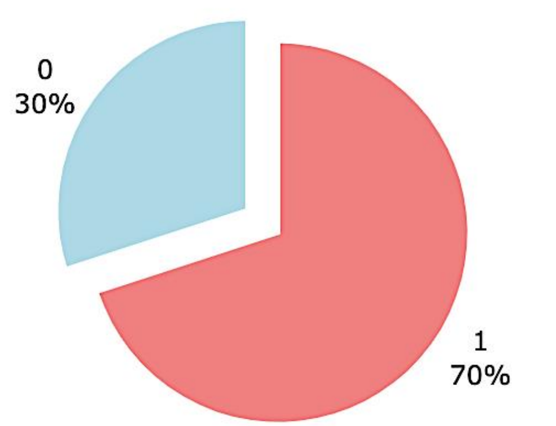

Fig. 9. Class Distribution showing the unbalanced nature of the German Datasets

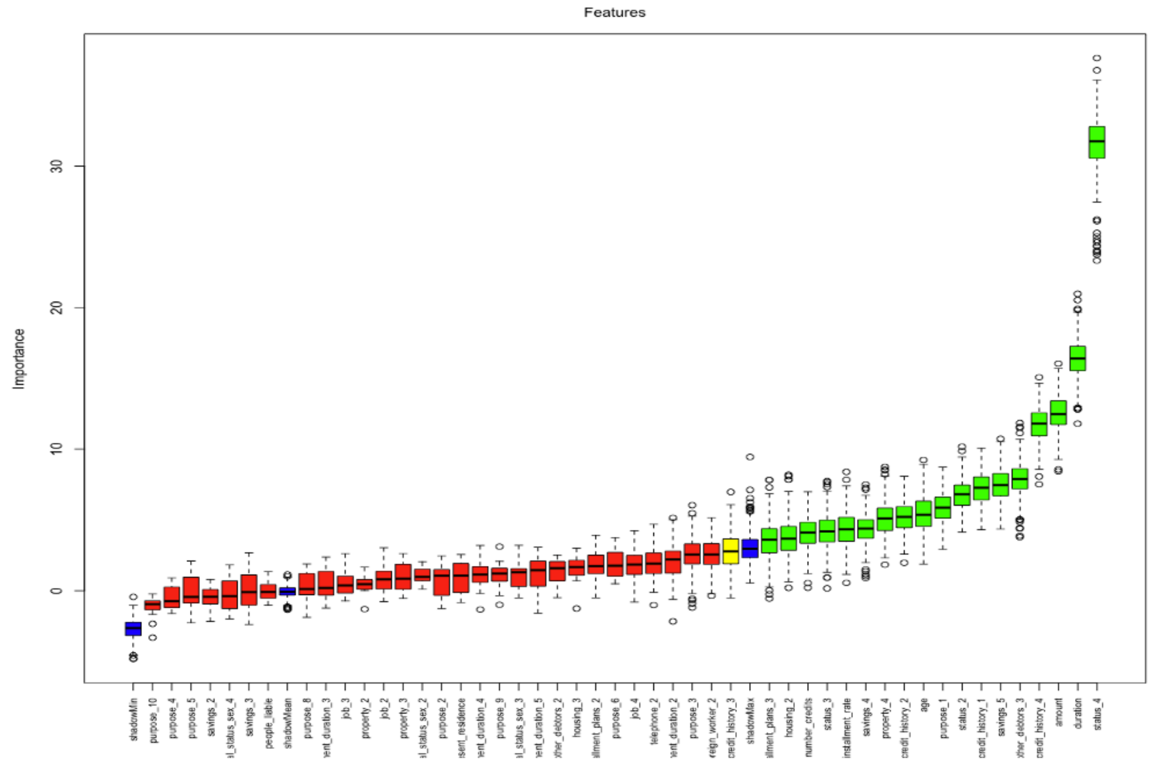

Fig. 10. Boruta Feature Selection Result in the First Runs Iteration

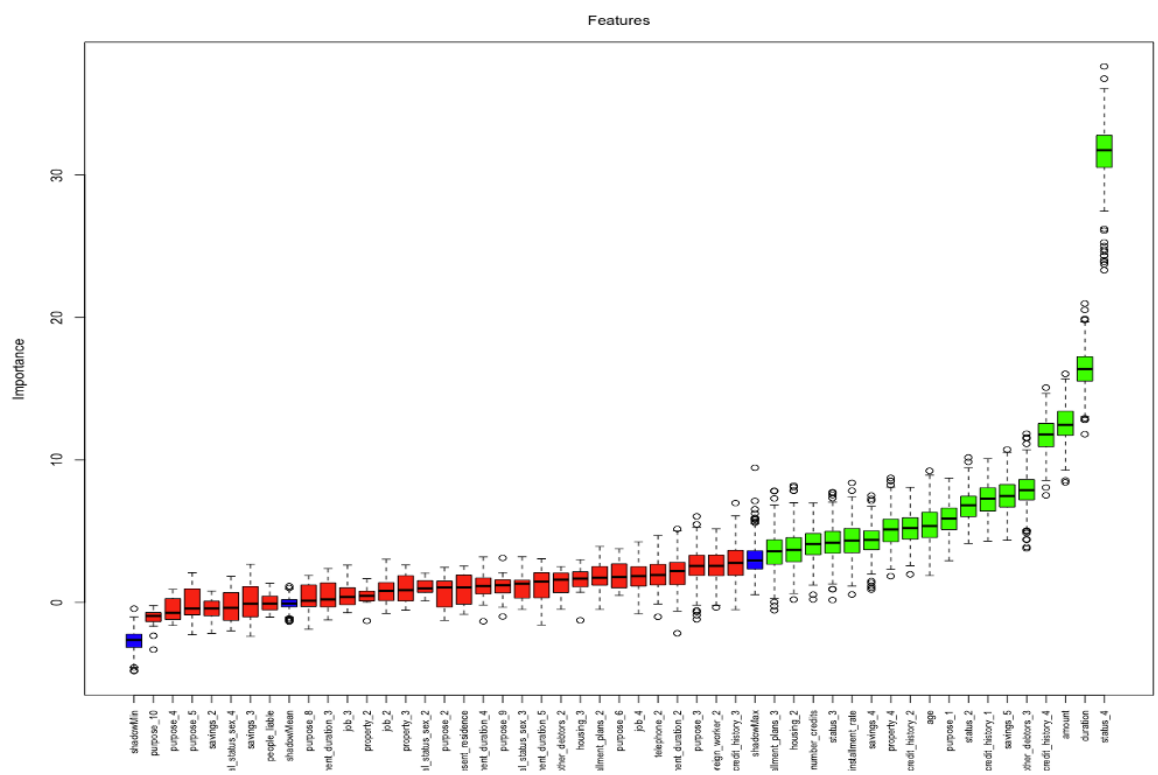

Fig. 11. Boruta Feature Selection Result in the Final Runs Iteration with Treatment

E. Empirical Result and Discussion

The dataset was split into training and test sets using an 80:20 ratio. Initially, the ctree model was trained without applying feature selection methods, utilizing three distinct sampling techniques. Table III illustrates the resampling distribution before and after the application of each method. The ROS technique resulted in an equal distribution of 50% for both classes. In comparison, SMOTE produced a slightly higher distribution for class 1 compared to class 0. In contrast, SMOTEtomek exhibited a notable marginal difference in distribution, favoring the 0 class over class 1. The effect of these distributions on model performance is discussed in the section below.

TABLE III. CLASS INSTANCES RESAMPLING

Sampling Method	No Sampling	ROS	SMOTE	SMOTEtomek
0	0.29125	0.5	0.4511133	0.6557377
1	0.70875	0.5	0.5488867	0.3442623

After training and evaluating the model, the findings reveal that Prec was notably higher for the No_sampling approach. This suggests a potential bias of the model towards the majority class. In contrast, when data sampling methods were employed, the model exhibited significantly higher Sens values across all methods. It implies that data sampling enhances the model's ability to identify rare cases more effectively. The ROS method achieved the highest Sens of 78.91%, outperforming other techniques. In terms of the F1-score, which provides a balanced tradeoff between Prec and Sens, SMOTE yielded the best result at 79.86%, as illustrated in Fig. 12. Examining the overall performance using AUC, SMOTE outperformed other methods, particularly without feature selection applied, as depicted in Fig. 13.

Subsequently, the Boruta technique was applied to obtain relevant features, as stated in Section IV (D). The model was retrained with the selected features as in the previous experiments. Findings show that a potential bias of the model towards the majority class is still observed whenever data is unbalanced. When data sampling methods were employed, the model became highly sensitive to each class distribution. The ROS method achieved the highest performance in all the metrics, maintaining its consistent performance. Notably, obtaining Sens and F1-score of 79.07% and 77.86%, respectively, to outperform other techniques, as shown in Figure 14. In addition, it is observed that MCC was at an all-time low for the No_sampling approach. Whereas ROS maintained a consistent best result. Examining the overall performance using AUC, ROS outperformed other methods, particularly as depicted in Figure 15.

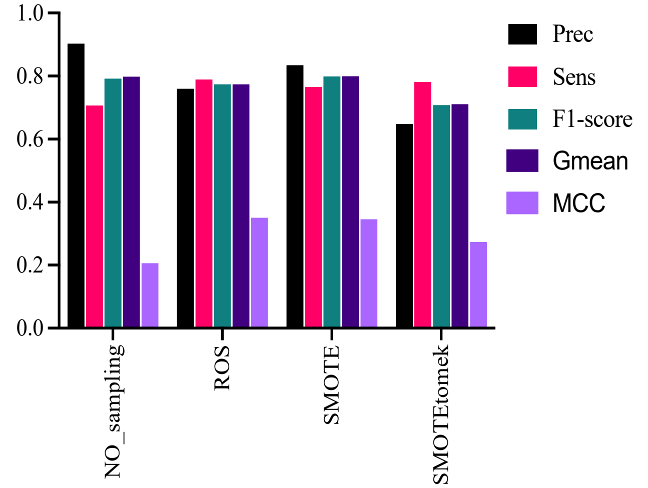

Fig. 12. Comparative Analysis of Data Sampling Method Without Boruta

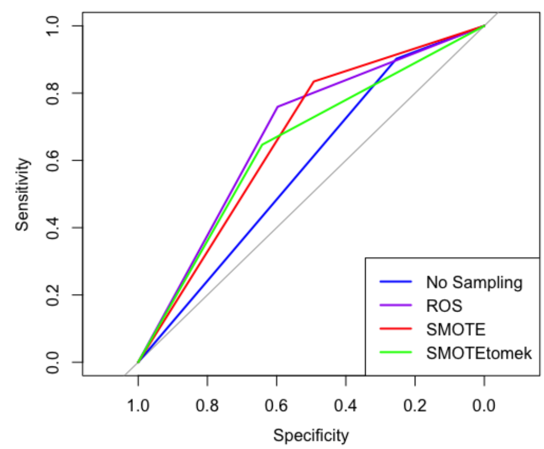

Fig. 13. AUC Comparision of Data Sampling Method Without Boruta

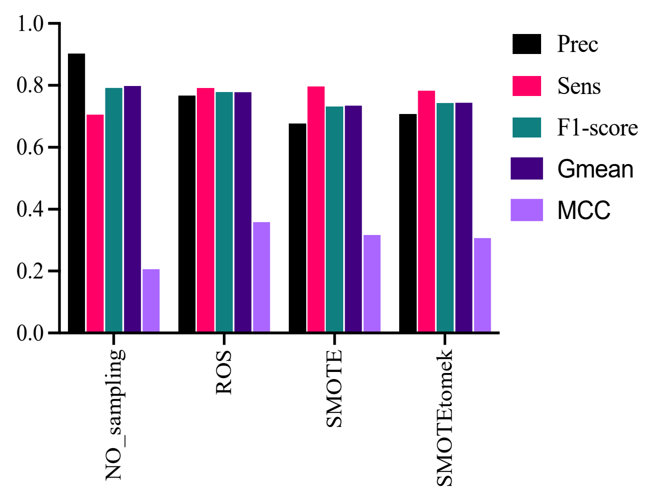

Fig. 14. Comparative Analysis of Data Sampling Method With Boruta

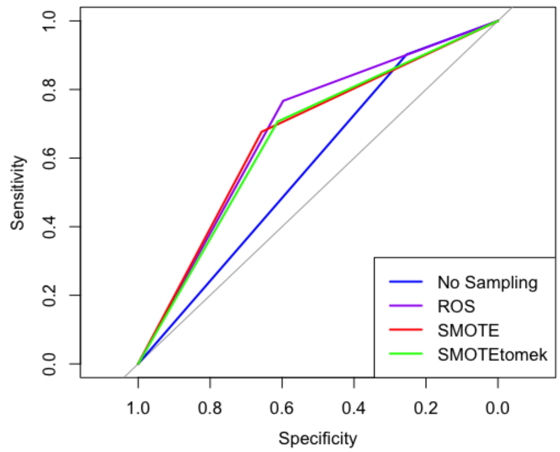

Fig. 15. AUC Comparision of Data Sampling Method With Boruta

V. CONCLUSION

The paper has discussed the CS prediction model, highlighting its crucial role in financial risk management. Financial institutions, especially banks and lending agencies, face intricate challenges that impact both their business and clientele. The study underscored that CS plays a pivotal role in mitigating these risks and reducing potential losses. It was observed that for CS to be effective, comprehensive, and accurate, customer information is necessary. The completeness of the information collected is paramount in assessing the creditworthiness of a customer. Financial risk management, being a delicate and crucial aspect, will continue to rely heavily on CS to enhance decision-making processes.

The study also delved into the limitations of the traditional method used for assessing a substantial volume of credit applications daily through scorecards. It became apparent that this fundamental selection approach lacks the assurance of accurate classifications. Consequently, addressing the challenges posed by extensive and intricate CS datasets necessitates the development of new, precise, and automated methods to minimize prediction errors.

Employing ML techniques to construct a predictive model for CS encounters challenges, particularly the curse of dimensionality and uneven class distribution. This paper investigated the validity of this concern by proposing an agnostic feature selection method, Boruta, coupled with three different sampling techniques. A comparative analysis was conducted, revealing that the combination of Boruta+ROS+*ctree* yielded the best results for handling CS prediction using the German dataset. This combination achieved a Sens of 79.07% and an F1-score of 77.86%.

For future research, it is recommended to explore the efficacy of other high-performing ML techniques such as XGBoost, SVM, or DNN. Additionally, considering alternative feature selection methods along with other CS datasets would contribute to a more comprehensive comparative analysis in addressing the challenges posed by CS prediction.

REFERENCES

[1] A. A. Hussin Adam Khatir and M. Bee, "Machine Learning Models and Data-Balancing Techniques for Credit Scoring: What Is the Best Combination?," Risks, vol. 10, no. 9, 2022, doi: 10.3390/risks10090169.

[2] A. Kumar, S. Sharma, and M. Mahdavi, "Machine learning (Ml) technologies for digital credit scoring in rural finance: a literature review," Risks, vol. 9, no. 11, pp. 1–15, 2021, doi: 10.3390/risks9110192.

[3] E. Dumitrescu, S. Hué, C. Hurlin, and S. Tokpavi, "Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects," Eur J Oper Res, vol. 297, no. 3, pp. 1178–1192, 2022, doi: 10.1016/j.ejor.2021.06.053.

[4] X. Dastile, T. Celik, and M. Potsane, "Statistical and machine learning models in credit scoring: A systematic literature survey," Applied Soft Computing Journal, vol. 91, p. 106263, 2020, doi: 10.1016/j.asoc.2020.106263.

[5] N. C. 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.

[6] X. Feng, Z. Xiao, B. Zhong, J. Qiu, and Y. Dong, "Dynamic ensemble classification for credit scoring using soft probability," Applied Soft Computing Journal, vol. 65, pp. 139–151, 2018, doi: 10.1016/j.asoc.2018.01.021.

[7] Y. Hayashi et al., "Recursive-rule extraction algorithm with J48graft and applications to generating credit scores," Journal of Artificial Intelligence and Soft Computing Research, vol. 6, no. 1, pp. 35–44, 2016, doi: 10.1515/jaiscr-2016-0004.

[8] S. A. Zonneveldt, K. B. Korb, and A. E. Nicholson, "Bayesian network classifiers for the German credit data," pp. 1–11, 2010.

[9] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in 2017 International Conference on Computing Networking and Informatics (ICCNI), IEEE, Oct. 2017, pp. 1–9. doi: 10.1109/ICCNI.2017.8123782.

[10] Z. Zhao, S. Xu, B. H. Kang, M. M. J. Kabir, Y. Liu, and R. Wasinger, "Investigation and improvement of multi-layer perception neural networks for credit scoring," Expert Syst Appl, vol. 42, no. 7, pp. 3508–3516, 2015, doi: 10.1016/j.eswa.2014.12.006.

[11] S. K. Trivedi, "A study on credit scoring modeling with different feature selection and machine learning approaches," Technol Soc, vol. 63, no. October 2017, p. 101413, 2020, doi: 10.1016/j.techsoc.2020.101413.

[12] S. Jadhav, H. He, and K. Jenkins, "Information gain directed genetic algorithm wrapper feature selection for credit rating," Applied Soft Computing Journal, vol. 69, pp. 541–553, 2018, doi: 10.1016/j.asoc.2018.04.033.

[13] I. Ramos-Pérez, Á. Arnaiz-González, J. J. Rodríguez, and C. García-Osorio, "When is resampling beneficial for feature selection with imbalanced wide data?," Expert Syst Appl, vol. 188, p. 116015, 2022, doi: 10.1016/j.eswa.2021.116015.

[14] B. Lal and S. Kumar, "Predictive Model on Churn Customers using SMOTE and XG-Boost Additive Model and Machine Learning Techniques in Telecommunication Industries," Int J Sci Res Sci Technol, pp. 646–661, 2021, doi: 10.32628/ijsrst218498.

[15] A. C. Kae, C. Xinying, J. O. Victor, and K. K. Wah, "Employee Turnover Prediction by Machine Learning Techniques," Journal of Telecommunication, Electronic and Computer Engineering, vol. 13, no. 4, pp. 49–56, 2021, [Online]. Available: https://jtec.utem.edu.my/jtec/article/view/6148/4063

[16] J. O. Victor, X. Chew, K. W. Khaw, and M. H. Lee, "A Cost-Based Dual ConvNet-Attention Transfer Learning Model for ECG Heartbeat Classification," Journal of Informatics and Web Engineering, vol. 2, no. 2, pp. 90–110, 2023, doi: 10.33093/jiwe.2023.2.2.7.

[17] M. Hassouna, A. Tarhini, T. Elyas, and M. S. Abou Trab, "Customer Churn in Mobile Markets: A Comparison of Techniques," International Business Research, vol. 8, no. 6, pp. 224–237, 2015, doi: 10.5539/ibr.v8n6p224.