Analyzing the Impact of the Threshold on Machine Learning Models for Credit Risk Prediction Using Business Intelligence

Assoc. Prof. Dr. Yanka Aleksandrova
University of Economics - Varna, Varna, Bulgaria
yalexandrova@ue-varna.bg

Chief Assist. Prof, Dr. Mariya Armyanova
University of Economics - Varna, Varna, Bulgaria
armianova@ue-varna.bg

Abstract

Making decisions based on predictions generated from machine learning models requires users to have a clear understanding of the mechanisms and logic behind every prediction. From one side, business users must be convinced in the ability of the models to generate correct predictions. Predictive power, expressed by the different performance measures, is not sufficient for building trust and acceptance of machine learning models. Business users need additional techniques and tools for model interpretation and evaluation of the effects from decisions based on machine learning predictions. In this research paper we propose an approach for analyzing the impact of different thresholds for converting probabilities into predictions for binomial classification machine learning models applicable for credit risk prediction in Peer-to-Peer Lending platforms. We define a set of measures to explore global and local impact on decision-making process and present different scenarios in a Business Intelligence application built in an analytical and business intelligent platform. Based on the presented results we can draw conclusions that when choosing the best model and threshold users should consider a broad set of measures not only for model accuracy, but also should consider misclassification costs, financial results, asset portfolio structure, etc.

Keywords: machine learning, business intelligence, credit risk prediction, explainable AI, artificial intelligence

JEL Code: O33

DOI: 10.56065/IJUSV-ESS/2023.12.2.79

Introduction

Machine learning models are used ubiquitously for decision making. One area of application is the risk assessment (Petrov et al., 2021) and particularly credit risk prediction, where a prediction must be made whether a loan applicant will repay the credit or not (Lohani et al., 2022), (Cetin et al., 2023), (Meshref, 2020). Credit risk prediction is especially important in the field of Peer-to-Peer lending business models. Crowd lending market is one of the steady growing areas of alternative finance along with digital payments in recent year and especially after the COVID-19 pandemic (Cambrridge Center of Alternative Finance, 2023). Unquestionably, Peer-to-Peer lending benefits investors as well as platforms, organizational, and individual borrowers. However, the unique characteristics of this industry and the ever-changing environment also present risks to this business model. An increase in the percentage of loans in default is one of the key dangers(Kehayova-Stoycheva et al., 2023). Bad loans pose a major risk to borrowers, online P2P lending platforms, and a large number of new investors joining the sector. This establishes the critical significance of the procedure for evaluating borrowers and estimating the likelihood that the loan will be or not be repaid.

When making decisions based on machine learning models, experts must have trust in models' accuracy and predictive power. The trustworthiness of a machine learning model depends heavily on the ability to interpret the model behavior, algorithm and discovered knowledge. The ability of a model to be explained and understandable to users is associated with its explainability and interpretability (Gall, 2018) (Miller, 2019) (Molnar, 2020). The interpretability of the model determines how much a person can understand the reasons behind the generation of a specific prediction in a supervised learning (Miller, 2019). The higher the degree of interpretation of a model, the easier it is to understand the reasons and mechanisms for making a decision or generating a particular prediction. The explainability of machine learning models, on the other hand, determines the possibility for model results to be presented in an understandable form to users.

The interpretation of the model requires its presentation in an appropriate textual or graphical form to the users, allowing easy adoption of the mechanism of generating the predictions. Some of the machine learning models are easier to explain because of their characteristics. Examples of such models with intrinsic interpretability are decision trees (Lundberg & Lee, 2017) (Ariza-Garzon et al., 2020). However, due to the limited ability of users to adopt more complex models, it should be noted that decision trees with more splits, end segments and long decision paths, as well as random forests, gradient boosted decision trees, decision jungles, etc. are equally difficult and impossible to perceive as neural networks classified as "black box" models (Molnar, 2020).

Due to the impossibility of achieving a satisfactory level of transparency of the model, additional agnostic models are built on top of the trained models to explain the generated predictions. Regarding this Explainable Artificial Intelligence (xAI) aims at developing models and algorithms that can provide understandable and clear explanation and interpretation of the predictions or algorithms characteristics to humans. The primary goal of xAI is to make the decision-making based on complex machine learning models more transparent and interpretable thus increasing the trustworthiness and applicability of trained models.

Every decision made based on machine learning models’ predictions has certain consequences that usually can be evaluated and measured locally and globally. For example, if the company approves a credit request based on the negative (fully paid) prediction, generated by the model, the immediate outcome is that the money is granted to the borrower and the loan is issued through the platform. If the prediction had been correct, the P2P lending company and investors will generate profit from interest payments and fees during the credit term. In case of a wrong prediction, however, the borrower would stop repaying the loan hence generating loss to both the P2P lending platform and investors. When the threshold for converting probabilities into predictions is set it has its effect on all the decisions based on the machine learning model. Globally this effect can be measured with the cost matrix which includes costs, associated with the wrong predictions (Gutzwiller et al., 2020). The total misclassification costs are calculated as sum of the costs for false positive and false negative predictions each multiplied by the respective counts of false positive and false negative cases. Cost matrix, as well as confusion matrix, is tightly related to the chosen threshold as every change in the threshold changes the counts of true and false positive and negative predictions. Cost matrix is a very effective way of measuring the financial result due to the application of machine learning models for decision making. However, there are other consequences of the implementation of the models besides the financial results. One such consequence is the default ratio, i.e. the proportion of default loans to all loans issued through the platform. P2P lending companies try to keep the default ratio at a reasonable acceptable level, because higher values of default ratio ruin the image of the company and causes investors to withdraw their funds. The default ratio usually varies for different loans categories, terms, etc. so it can be calculated globally or across the credit portfolio. Credit portfolio is another area which is influenced from the application of machine learning models. The structure of the portfolio is aligned with the corporative strategy and policy and because of this company must measure and evaluate the effects of the machine learning models on the structure of the credit portfolio. Besides default ratio and credit portfolio structure, there can be other consequences to be considered which require their measuring and evaluation.

The interpretation of a machine learning model in different dimensions requires data visualization and exploration. It can be successfully implemented as an analytical app, built in a Analytics and Business Intelligence (ABI) platform. These platforms provide excellent opportunities for creating various forms of presenting resultant information, defining metrics and key performance indicators, filtering, aggregating, and slicing data. Current ABI platforms are focused on decision-making processes through delivering contextualized insights and advanced analytics (Schlegel et al., 2023).

1. Research methodology

Binomial classification models are used for predicting probabilities for belonging to either of the two target classes. For convenience we assume that one of the classes is 0 (or the negative class) and the other – 1 (or the positive class). Hence the probabilities generated by the binomial classification models are p0 – probability for belonging to class 0 and p1 – probability for belonging to class 1. The sum of p0 and p1 is 1. These probabilities are then converted into prediction by applying a threshold. The importance of the chosen threshold is crucial for the evaluation of model's performance with measures derived from the confusion matrix. When the p1 threshold increases, the number of positive predictions decreases leading to fewer true and false positives. At the same time the number of negative predictions increases thus increasing true and false negatives. The opposite is observed when p1 threshold decreases. In these cases, the number of true and false positives increases, and the number of true and false negatives decreases. Every change in positive and negative predictions causes respective changes in accuracy, sensitivity, and specificity. Since these measures rely heavily on the chosen threshold some machine learning models automatically set the threshold such as to achieve the best performance based on complex measures like F1 score or AUC which don't depend on the threshold. This approach, however, does not consider the misclassification costs expressed in the cost matrix. When a classification model is used for decision-making users must be fully aware not only of the accuracy of the predictions but also of the implications of the decision taken. Based on this we propose to define a set of measures for evaluating the effects of implementation of the machine learning models, when the implications are related to financial results after the model implementation. The proposed set includes the following measures:

- $Total\ Financial\ Result\ After\ ML = \sum_{i=1}^{j} Financial\ Result_i$, where $Financial\ Result_i$ is the result from i-th approved credit request and j is the total number of approved requests. Total Financial Result after machine learning implementation for decision making evaluates financial effect of the predictions generated of the model with a certain threshold. For correctly classified negative (True Negatives - TN) cases the result can be calculated as the average profit from a paid loan and for incorrectly classified negative (False Negatives - FN) cases the result is equal to the average loss generated from default loans.
- $Financial\ Result\ Lift\ (\%) = \frac{Total\ Financial\ Result\ After\ ML}{Total\ Financial\ Result\ Before\ ML} - 1$. This measure calculates the increase or decrease in % of the financial result after the machine learning model is implemented for decision making compared to the financial result without machine learning, i.e. the original data.
- $Default\ Ratio\ After\ ML = \frac{Fully\ paid\ loans}{All\ loans} = \frac{TN}{TN+FN}$
- Portfolio structure – default ratio by different loans categories, purposes, terms, etc.

Financial results are calculated from the provided Lending Club dataset. The dataset contains data about actual profit or loss generated from every loan. The average profit from one fully paid loan is 2600 USD and the average loss from a default loan is 10500 USD. In this research we consider the actual profit or loss and not the averaged results.

The importance of the chosen threshold for converting probabilities into predictions can be presented in the decision-making process depicted in Figure 1. When the probability for credit default is greater than the chosen threshold, the prediction is positive, and the credit request is denied. If the probability of default (p1) is less than the threshold, a negative prediction is generated, and the credit request is approved. All financial results are derived from the approved credit requests. The profit is acquired from correctly predicted negative classes and loss is generated from incorrectly predicted negative classes. The process from fig.1 simulates a real-time scenario for implementation of machine learning model for decision-making thus alternative profits and losses from positive predictions are not considered here. Since the probability is converted into a prediction using the threshold cutoff, the threshold value must be chosen in such a way as to achieve two main goals – model accuracy and optimal financial result. Sometimes these goals require setting different threshold values meaning that the most accurate model is not necessarily the model insuring the maximum profit. In these cases, a compromise must be made by prioritizing one goal or the other. Business users involved in the decision-making process should be aware of the implications and been able to compare different scenarios.

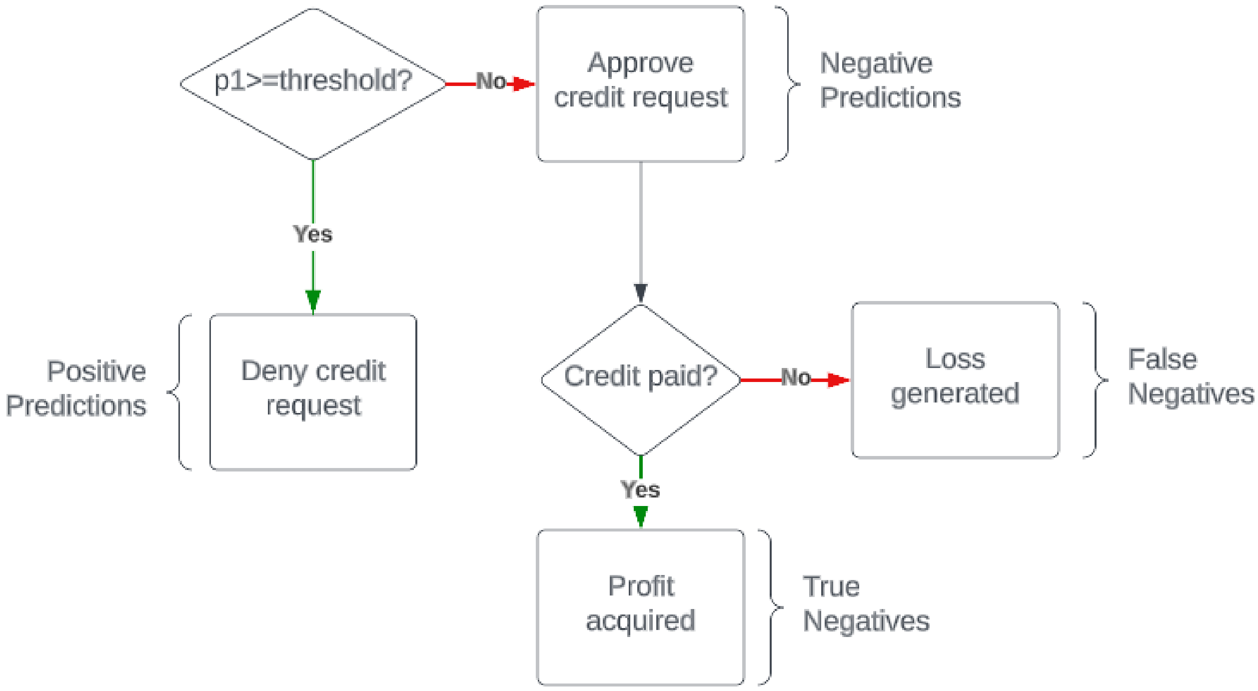


Figure 1. Decision-making using machine learning model

Business intelligence platforms support the needed functionality to explore and analyze key performance indicators (KPIs) measuring the effectiveness of machine learning implementation in real-time using suitable visuals. This requires building an analytical application for interpreting generated predictions and simulation of machine learning implementation for decision-making with different probability thresholds. The interpretation of model's prediction is done in different levels. On a global level measures like financial result, profit lift and default ratio after machine learning implementation is used for evaluation of the implications of the model's implementation. On a local level the prediction with the set threshold is compared with the actual target value to estimate its accuracy.

2. Training machine learning models

The dataset is downloaded from the Lending Club platform's website. This online P2P lending platform provides data in .csv for the loans granted since 2007. By 2016, the data are provided on an annual or several-year basis, and due to the increased number of loans from 2016 data is published each quarter. The last set of data provided by Lending Club at the time of the survey is for the second quarter of 2020, updated on 28.07.2020.

Data provided by Lending Club is a subject of numerous studies. Serrano-Cinca (Serrano-Cinca et al., 2015) selects 18 factor variables classified into five groups: borrower assessment by the lending organization, credit characteristics, credit history and indebtedness. Similar are the variables selected by Carmichael (Carmichael, 2014). Ariza-Garzon (Ariza-Garzon et al., 2020) also includes the employment length at the current work, previous experience with the P2P lending platform, state code of the borrower, FICO score.

When selecting variables involved in the structure of the data set subject to subsequent analysis, the following criteria are considered:

- low percentage missing values. All selected variables have a percentage of missing values below 3%.
- lack of constant values for all observations.
- Information value based on theoretical research in the field of creditworthiness assessment.

The final data set contained 1 467 296 cases describing credits from 2012 to 2017. The chosen independent variables can be classified into six categories:

1. General characteristics of the loan applicant – employment length, home ownership
2. Financial characteristics of the loan applicant – annual income, FICO score, bank accounts from different types (satisfactory, bankcard, installment, mortgage), current balance, credit limit
3. Loan characteristics – amount of the loan, purpose, term in months (36 or 60)
4. Characteristics of applicant's indebtedness – ratio between current balance and credit limit, debt payments divided by income monthly, total revolving balance, revolving utilization rate.
5. Credit history of the loan applicant – months since first credit line, months since first bank installment account, months since first revolving account, total number of credit lines, etc.
6. Macroeconomic indicators - effective funding rate and unemployment rate (Federal Reserve Bank of St.Louis, 2021).

Data has been checked for missing values and variables with more than 30% missing values has been removed. Category variables have been transformed into numerical using one-hot encoding method. Outliers detection is performed and different outliers removal techniques has been implemented according to the outliers distribution and ratio. To eliminate the influence of different numerical scales a Robust Scaler has been implemented on the training dataset. This scaler is used for data standardization by removing the median and scaling the data according to the Interquartile range. The scaling is based on percentiles and is robust toward the presence of outliers.

The model to predict credit status is trained using XGBoost algorithm. This algorithm is chosen based on its excellence predictive power, revealed on numerous papers in the area (Setiawan et al., 2019), (Aleksandrova, 2021), (Zhang et al., 2023). The optimization of models hyperparameters is crucial for the model performance and we have implemented Bayesian Optimization to optimize hyperparameters values. Bayesian Optimization is suitable for optimizing cost intensive functions (Frazier, 2018). Parameter grid space is set for the following hyperparameters:

- Max depth of the trees (max_depth)– from 5 to 20
- Learning rate (learning_rate) – from 0.01 to 1
- Number of trees (n_estimators) – from 50 to 500
- Subsampling ratio (subsample) – from 0.5 to 1
- Sampling of the columns (colsample_bytree) – from 0.3 to 1
- Gamma (gamma) – from 0 to 3

The objective function is set to maximize Area Under ROC curve (AUC). The initial dataset has a strongly imbalanced structure with 80% of cases belonging to the negative class (Fully Paid) and 20% belonging to the positive class (Default). To accommodate this class imbalance, we set the parameter scale_pos_weight for scaling the minority class to 4. After the optimization the following values of hyperparameters have been calculated: colsample_bytree=0.9135[1], gamma= 2.6838, learning_rate=0.0942, max_depth=6, n_estimators=126, subsample=0.9391.

The best model is then applied to the test dataset and generated probabilities are added to the dataset. The model generates two probabilities – p0 for belonging to the negative class and p1 for belonging to the positive class. The test dataset along with the probabilities p0 and p1 are saved and exported in a .csv file.

3. Building Business Intelligence application

The chosen ABI platform for implementation of the BI application is Microsoft Power BI. This platform has been chosen because of its excellent functional capabilities, advanced analytics, price/value ratio, support of Python scripts and seamless integration with external environments which position Power BI as a leader in the global ABI market (Schlegel et al., 2023). The application is built in the free Power BI Desktop environment.

To explore the influence of the chosen threshold on the set evaluation measures we created a numeric parameter and used it in the DAX formulas for calculating financial and accuracy results from model implementation. The measures that we created in Power BI are shown in Table 1.

Table 1. Measures in Power BI

Measure	DAX formula	Explanation
cumProf	CALCULATE(SUM(Test[prof_loss_act]), Test[p1]<=SELECTEDVALUE(Threshold[Threshold]))	Cumulative profit/loss at the selected threshold
FN	CALCULATE(COUNT(Test[index]),Test[p1]<SELECTEDVALUE(Threshold[Threshold]),LC200Ksamples[loan_status_final]=1)	False negatives count at the selected threshold
FP	CALCULATE(COUNT(Test[index]),Test[p1]>=SELECTEDVALUE(Threshold[Threshold]),LC200Ksamples[loan_status_final]=0)	False positives count at the selected threshold
Missed prof/loss	CALCULATE(SUM(Test[prof_loss_act]),Test[p1]>=SELECTEDVALUE(Threshold[Threshold]))	Missed profit or loss at the selected threshold
TN	CALCULATE(COUNT(Test[index]),Test[p1]<SELECTEDVALUE(Threshold[Threshold]),LC200Ksamples[loan_status_final]=0)	True negatives count at the selected threshold
TP	CALCULATE(COUNT(Test[index]),Test[p1]>=SELECTEDVALUE(Threshold[Threshold]),LC200Ksamples[loan_status_final]=1)	True positives count at the selected threshold
Prediction	IF(SELECTEDVALUE(Test[p1])>=SELECTEDVALUE(Threshold[Threshold]),1,0)	Prediction at the selected threshold
Accuracy	([TP]+[TN])/([TP]+[TN]+[FP]+[FN])	Accuracy at the selected threshold
Sensitivity	[TP]/([TP]+[FN])	Sensitivity at the selected threshold

[1] Values are rounded up to 4 digit after decimal point

Measure	DAX formula	Explanation
Specificity	[TN]/([TN]+[FP])	Specificity at the selected threshold
DR after	[FN]/([FN]+[TN])	Default ratio after model implementation
ProfitLift	[cumProf_act]/SUM(Test[prof_loss_act])-1	Profit lift after the model implementation

A Power BI dashboard for exploring the influence on the set threshold on default ratio, reject rate, profit lift, cumulative profit, accuracy, sensitivity, specificity and missed profit or loss is shown in Figure 2. The different threshold values are set using the slicer at the top left part of the dashboard. In the example below with the default value of the threshold of 0.5 the company can achieve a drop by 6 points in the default ratio – from 18.78% before model implementation to 12.03% after the implementation. Evaluation measures show that at this threshold the accuracy of the model is 0.678, sensitivity – 0.587 and specificity – 0.699. Applying the machine learning model for decision-making on credit requests with the default threshold value of 0.5 will lead to profit lift of 270.99%, which is approximately 21.45 million USD greater than the base profit without using machine learning model. Visible from the figure the greatest decrease in default ratio is observed for the riskier loans – categories D, E and F and with 60 months terms.

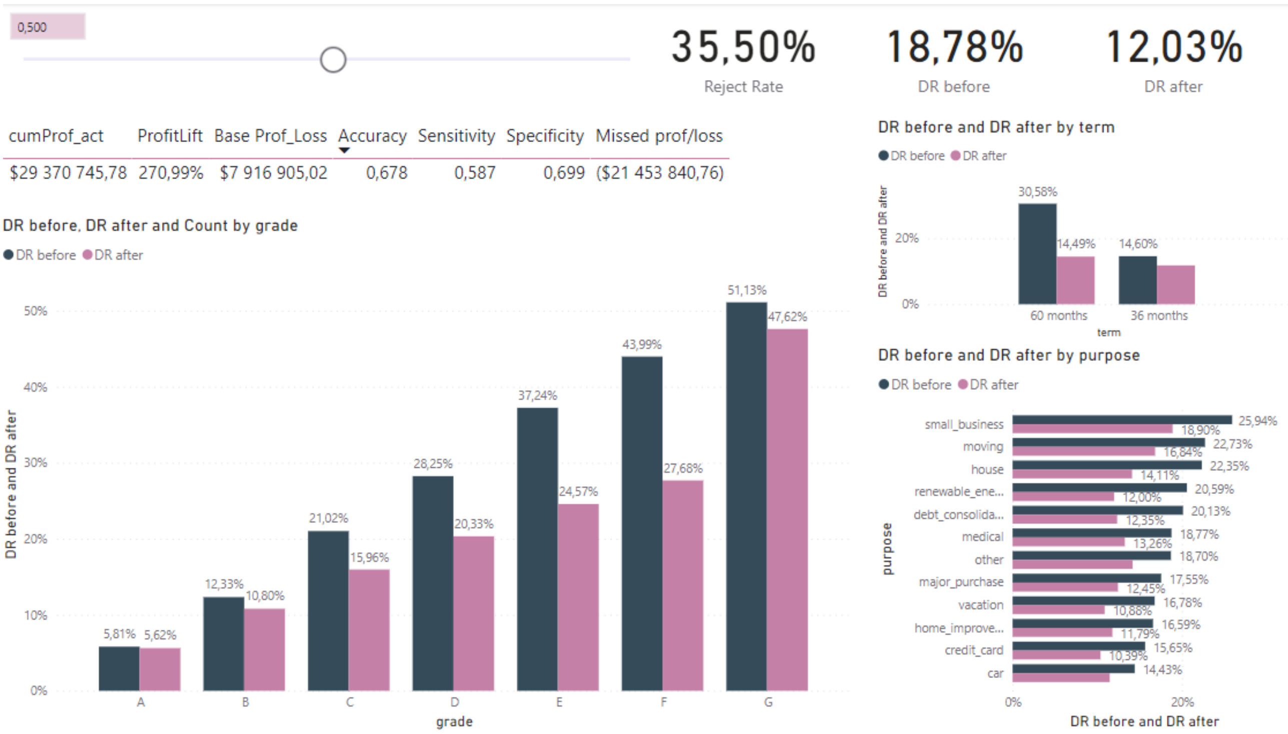
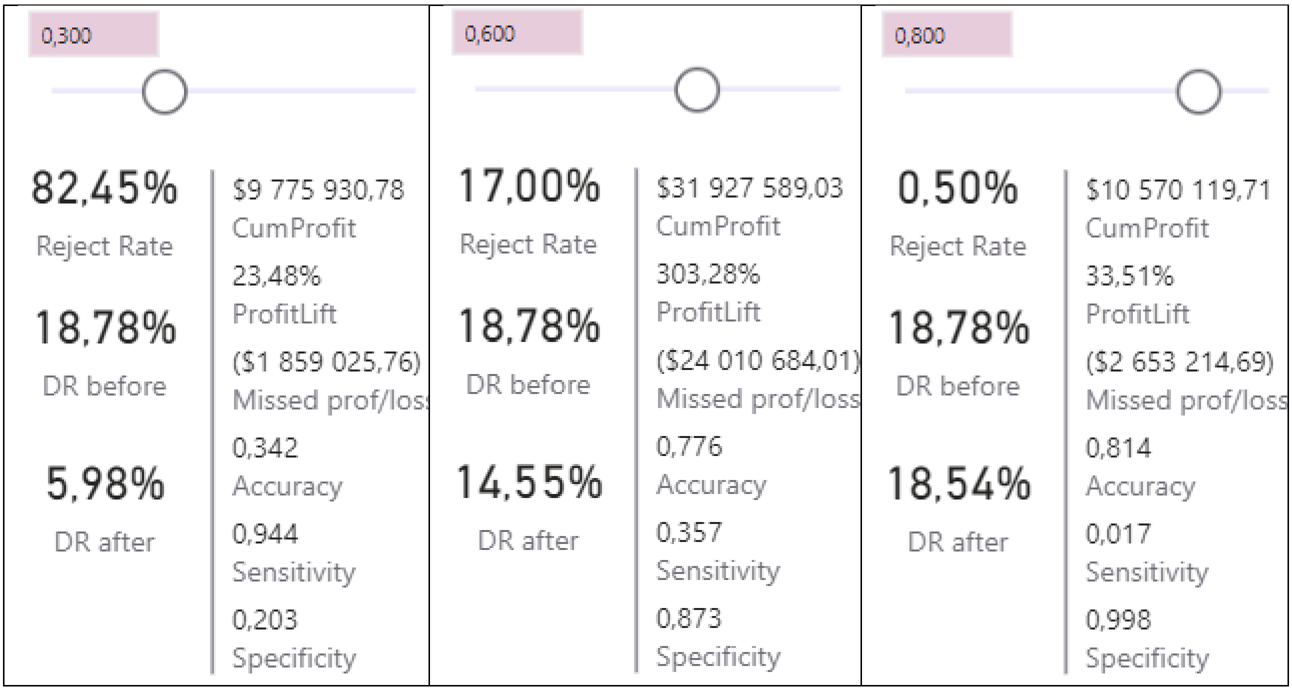


Figure 2. Power BI dashboard for exploring the impact of different thresholds

Power BI dashboards can be used to compare different scenarios and evaluate the outcomes when implementing a certain threshold level. Three different scenarios are depicted in Figure 3 with measures calculated according to three different levels of thresholds – namely 0.3, 0.6 and 0.8. The greatest decrease in default ratio can be achieved with the lowest threshold of 0.3. If this threshold is applied the default ratio will be decreased from 18.78% to 5.98%. The company can still generate profit lift of 23.48%, equivalent to 1.86 million USD, but this will be achieved at the cost of 82.45% reject rate. This means that the company must enforce a very restrictive policy regarding credit requests, approving only 17.55% of them. Such a policy will have a negative effect on the loans supply and will eventually result in an outflow of investors.

The total accuracy of the predictive model with a threshold at 0.3 is 0.342. The first presented scenario is however the one with the greatest sensitivity – 0.944, meaning that more than 94% of the positive cases have been correctly classified as default. At the same time the model has a very low specificity (0.203) as only 20% of the fully paid loans would be correctly predicted as such. All these measures are influenced by the low level of the threshold resulting in 82.45% reject rate.



Scenario 1 (threshold=0.3) Scenario 2 (threshold=0.6) Scenario 3 (threshold=0.8)

Figure 3. Scenario evaluation in Power BI

The greatest profit lift (303.28%) from these scenarios is observed when the threshold is set to 0.6 (scenario 2). With a 17% reject rate the company will be able to ensure enough loan supply through Peer-to-Peer lending platform and keep the potential investors. The default ratio would decrease by 5 points and would reach 14.55%. The accuracy of the model at a threshold at 0.6 is 0.776 which is more than two times higher than the total accuracy at a threshold at 0.3 (scenario 1). The second scenario however has a very low sensitivity (0.357) compared to the first scenario (0.944). The second model can correctly classify only 35.7% of the default loans and still achieve a profit lift of more than 24 million USD. The ability of the model to correctly classify negative cases, expressed in its specificity, is significantly increased at the threshold 0.6 and is equal to 0.873, resulting in a right identification of 87.3% of all fully paid loans.

The lowest reject rate of 0.5% is observed in scenario 3 with a threshold level of 0.8. This results in a very tolerant policy of approving all requests where the probability of default is lower than 80%. With this threshold the model can correctly classify more than 99% of the default loans but at the cost of identifying lower than 2% of the default ones. The total accuracy of the model from scenario 3 is 0.814 which is the greatest of the observed scenarios. The sensitivity of the model is 0.017 which is sufficient for a modest profit lift of 33.51%.

It's verry important to consider all performance measures when analyzing the predictive power of classification models. Every one of the scenarios presented in Figure 3 is better than the rest when we consider different measures. For example, if we compare the scenarios according to total accuracy, the best threshold level is 0.8, at which scenario 3 achieves accuracy of 0.814. If the goal is to get maximum decrease in the default ratio, then scenario 1 with a threshold of 0.3 would be the best one. With this threshold the default ratio after machine learning implementation is reduced from 18.78% to 5.98%. If the primary goal is to optimize the profit, while keeping the default ratio and reject rate at acceptable levels, then scenario 2 is the best one. With its threshold of 0.6 the P2P lending company can accumulate 24 million USD which is equivalent to 303.28% profit lift. The reject rate is 17% and the default ratio as a result of machine learning implementation would decrease by 4 points. The model from scenario 2 has a total accuracy of 0.776, sensitivity of 0.357 and specificity of 0.873. These measures confirm the initial hypothesis that for a model to achieve significant profit lift it should be better at correctly identifying the negative cases (fully paid loans). Although the misclassification costs associated with incorrect classification of a negative case are lower than similar costs for incorrect classification of a positive case (default loans), the distribution of the target classes is 4:1 in favor of negative classes.

Conclusion

Model interpretation and evaluation of effects from decision-making is as important as the performance measures for the acceptance and implementation of machine learning models. In the case of binomial classification machine learning model's threshold for converting probabilities into predictions determines the consequences from decisions based on predictions. Business users can explore different scenarios and choose the optimal value of the threshold. It's very important to define a broad range of measures to monitor and evaluate the achieved levels at different threshold levels. With the demonstration of a business intelligence and analytical app in this research paper we proposed a practical approach suitable for implementation in the financial area for credit risk prediction. The application combines the advantages of machine learning models with the functionality of analytics and Business Intelligence platform to deliver a comprehensive framework to explore and choose the best threshold according to the desired results.

References

1. Aleksandrova, Y. (2021) Comparing Performance of Machine Learning Algorithms for Default Risk Prediction in Peer to Peer Lending, *TEM Journal*, 10(1), pp. 133–143. Available at: https://doi.org/10.18421/TEM101-16 [Accessed 01/05/2024].
2. Ariza-Garzon, M., Arroyo, J., Caparrini, A., & Segovia-Vargas, M.-J. (2020). Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending. *IEEE Access*, 8, pp. 64873-64890. Available at: doi:10.1109/ACCESS.2020.2984412 [Accessed 01/05/2024]
3. Been, K., Khanna, R., & Koyejo, O. (2016). Examples are not enough, learn to criticize! Criticism for interpretability. *29th Conference on Neural Information Processing Systems (NIPS 2016)*, 29, pp. 2280-2288. Barcelona, Spain.
4. Cambridge Center of Alternative Finance. (2023). *The Global COVID-19 Fintech Market Impact and Industry Resilience Study*. [Online] Available from: https://www.jbs.cam.ac.uk/wp-content/uploads/2022/06/2022-global-covid19-fintech-market-study.pdf [Accessed 01/05/2024]
5. Cetin, A.I. and Ahmed, S.E. (2023) Determinants of Credit Ratings and Comparison of the Rating Prediction Performances of Machine Learning Algorithms, *E3S Web of Conferences*, 409. Available at: https://doi.org/10.1051/e3sconf/202340905013 [Accessed 01/05/2024].
6. Frazier, P.I. (2018) A Tutorial on Bayesian Optimization. *arXiv:1807.02811 [stat.ML]*. Available at: https://doi.org/10.48550/arXiv.1807.02811 [Accessed 01/05/2024].

7. Gall, R. (2018). Machine Learning Explainability vs Interpretability: Two concepts that could help restore trust in AI. [Online] Available from: https://www.kdnuggets.com/2018/12/machine-learning-explainability-interpretability-ai.html [Accessed 01/05/2024].
8. Gutzwiller, K.J. and Chaudhary, A. (2020) Machine-learning models, cost matrices, and conservation-based reduction of selected landscape classification errors, *Landscape Ecology*, 35(2), pp. 249–255. Available at: https://doi.org/10.1007/s10980-020-00969-y [Accessed 01/05/2024].
9. Kehayova-Stoycheva, M. and Vasilev, J. (2023) Developing a Household Sustainable Consumption Index Calculator – Marketing and IT Dimensions, *TEM Journal*, 12(4), pp. 2101–2111. Available at: https://doi.org/10.18421/TEM124-21 [Accessed 01/05/2024].
10. Lohani, B.P. et al. (2022) Machine Learning Based Model for Prediction of Loan Approval, *Proceedings of 3rd International Conference on Intelligent Engineering and Management, ICIEM 2022*, pp. 465–470. Available at: https://doi.org/10.1109/ICIEM54221.2022.9853160 [Accessed 01/05/2024]..
11. Lundberg, S., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in neural information processing systems*, pp. 4765-4774 Available from: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf [Accessed 01/05/2024].
12. Meshref, H. (2020) Predicting loan approval of bank direct marketing data using ensemble machine learning algorithms, *International Journal of Circuits, Systems and Signal Processing*, 14, pp. 914–922. Available at: https://doi.org/10.46300/9106.2020.14.117 [Accessed 01/05/2024].
13. Miller, T. (2019). Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artificial Intelligence*, 267, 1-38. Available from: doi:https://doi.org/10.1016/j.artint.2018.07.007 [Accessed 01/05/2024]
14. Molnar, C. (2020). Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. [Online] Available from: https://christophm.github.io/interpretable-ml-book/index.html [Accessed 01/05/2024]
15. Petrov, P., Stoev, S., Radev, M., & Sergeev, A. (2021). RISK MANAGEMENT PROCESSES IN INFORMATION SYSTEMS DEVELOPMENT. *International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM*, 21 (7), pp. 467-473. Available from: doi:10.5593/sgem2021/2.1/s07.18 [Accessed 01/05/2024]
16. Schlegel, K., Sun, J., Pidsley, D., Ganeshan, A., Fei, F., & Popa, A. (2023). Magic Quadrant for Analytics and Business Intelligence Platforms. [Online] Available from https://www.gartner.com/doc/reprints?id=1-2955ETOT&ct=220215&st= sb?ocid=lp_pg398450-_gdc_comm_az [Accessed 01/05/2024]
17. Setiawan, N., Suharjito and Diana (2019) A comparison of prediction methods for credit default on peer to peer lending using machine learning, *Procedia Computer Science*, 157, pp. 38–45. Available at: https://doi.org/10.1016/j.procs.2019.08.139 [Accessed 01/05/2024].
18. Zhang, C. and Zhou, X. (2023) Forecasting Credit Risk of SMEs in Supply Chain Finance Using Bayesian Optimization and XGBoost, *Mathematical Problems in Engineering*, 2023. Available at: https://doi.org/10.1155/2023/5609996 [Accessed 01/05/2024].