EXPLAINABLE MACHINE LEARNING FOR LENDING DEFAULT CLASSIFICATION

Golnoosh Babaei [1], Paolo Pagnottoni[2] and Thanh Thuy Do[3]

[1] Department of Engineering, University of Pavia, Pavia, Italy, (e-mail: golnoosh.babaei01@universitadipavia.it)

[2] Department of Economics and Management, University of Pavia, Pavia, Italy, (e-mail: paolo.pagnottoni@unipv.it)

[3] Department of Economics, University of Insubria, Varese, Italy, (e-mail: ttdo@studenti.uninsubria.it)

ABSTRACT: Machine Learning (ML) models are often used to support classification decision-making, such as in peer-to-peer lending. However, they usually lack interpretable explanations. While Shapley values and the computationally efficient variant Kernel SHAP may be employed for this aim, the latter makes the assumption that the features are independent. We explain classifiers through a Kernel SHAP method able to handle dependent features in the context of credit risk management for peer-to-peer lending. We demonstrate the effectiveness of our method by considering linear and non-linear models with varying degrees of feature dependence, showing that our approach yields credible estimates of true Shapley values across model and dependence specifications.

KEYWORDS: feature dependence; Shapley values; machine learning; explainability.

1 Introduction

Obermeyer & Emanuel, 2016 pointed out that ML model interpretability enhances medical, healthcare, credit scoring, and fraud detection. Explaining complex ML model predictions is a challenging task, and the model's explanation is crucial for both reliability of the estimates and for fairness and compliance with respect to General Data Protection Regulation compliance. Peer-to-peer lending requires creditworthiness, namely transparent and trustworthy explanations to build trust and help lenders and borrowers make well-informed choices. Credit risk analysis determines peer-to-peer lending rates and creditworthiness, and lenders may distrust complicated ML model predictions. Explainable Artificial Intelligence (XAI) improves classification accuracy, model transparency and interpretability via the concept of game-theoretic

Shapley values. Recent model-agnostic explanation methods simplify understanding of how each predictor affects the prediction; in particular, Aas *et al.*, 2021 expand Kernel SHAP to address interdependent characteristics. We exploit such formulation of Kernel SHAP to build predictive classification ML models and relative model explanations for interpretable peer-to-peer credit risk management. We test our proposal on three predictive ML models, i.e. logistic regression, GAMs, XGBoost, and four structures for modelling feature dependence, i.e. the independent case, Gaussian, empirical distribution and copula. This study reveals that linear and non-linear models with variable feature dependencies give consistent and reliable Shapley value estimates. This enhances the understanding of the drivers of peer-to-peer lending credit risk and outlines best practices for its management via machine learning classification techniques.

2 Kernel SHAP for dependent features

Kernel SHAP computes feature importance using weighted linear regression and local linear regression coefficients. In classical machine learning, a predictive model, $f(x)$, is trained using a training set of size n_{train} comprised of sets y $\left\{y^i, x^i\right\}_{i=1,\ldots,n_{train}}$ where $\quad j = 1,\ldots,n_{train}$. This model attempts to closely approximate the response value y. To explain the prediction $f(x^*)$ for a particular feature vector $x = x^*$, the Kernel SHAP technique only uses the independence assumption $p(x_{\bar{S}} \mid x_S) = p(x_{\bar{S}})$ - see Aas *et al.*, 2021.

We examine how the three different ways of accounting for dependence structures in the features increase ML credit risk model accuracy and feature explainability compared to independence.

2.1 Multivariate Gaussian distribution

Given that the feature vector x is obtained from a multivariate Gaussian distribution with mean vector μ and covariance matrix Σ, then the conditional distribution $p\left(x_S \mid x_S = x_S^*\right)$ is also multivariate Gaussian. By expressing $p(x)$ in terms of $p(x) = p(x_S, x_{\bar{S}}) = \mathrm{N}_M(\mu, \Sigma)$ with $\mu = (\mu_S, \mu_{\bar{S}})^\top$ and

$$\Sigma = \begin{bmatrix} \Sigma_{SS} & \Sigma_{S\bar{S}} \\ \Sigma_{\bar{S}S} & \Sigma_{\bar{S}\bar{S}} \end{bmatrix}$$

gives $p\left(x_{\bar{S}}|_S = x_S^*\right) = \mathrm{N}_{|\bar{S}}\left(\mu_{\bar{S}|S}, \Sigma_{\bar{S}_{|S}}\right)$, with

$$\mu_{\bar{S}|S} = \mu_{\bar{S}} + \Sigma_{\bar{S}S}\Sigma_{SS}^{-1}\left(x_S^* - \mu_S\right)$$

and

$$\Sigma_{\bar{S}|S} = \Sigma_{\bar{S}\bar{S}} - \Sigma_{\bar{S}S}\Sigma_{SS}^{-1}\Sigma_{S\bar{S}}$$

2.2 Gaussian Copula

A d-dimensional copula is a multivariate distribution, C, characterized by uniformly distributed marginal probabilities $\mathrm{U}(0,1)$ over the unit interval of $[0,1]$. Sklar's theorem states that for each multivariate distribution F with univariate distributions $F_1, F_2, \ldots, F_d$ can be written as

$$F(x_1, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)),$$

for some appropriate d-dimensional copula C. In fact, the copula from (12) has the expression

$$C(u_1, \ldots, u_d) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_d^{-1}(u_d)\right)$$

where the F_j^{-1} s are the inverse distribution functions of the marginals. Assuming a Gaussian copula, the following methodology can be employed to generate samples from $p\left(x_S \mid x_S = x_S^*\right)$.

2.3 Empirical conditional distribution

We propose a non-parametric method if x's dependence structure and marginal distributions depart from the Gaussian. The kernel estimator, a classical non-parametric density estimation method, has been modified and improved over the decades. The kernel estimator is impeded by the curse of dimensionality, which rapidly restricts its applicability in multivariate problems. Additionally, the non-parametric estimation of conditional densities is limited to a small number of techniques, particularly when either x_S or $x_{\bar{S}}$ is not one-dimensional. Ultimately, most kernel estimation methods generate a non-parametric density estimate, however, samples from the estimated distribution must be produced. Consequently, we have formulated an empirical conditional method to approximately sample from $p\left(x_{\bar{S}} \mid x_S^*\right)$.

3 Empirical Findings

We compare accuracy and prediction explanations from different ML models and feature dependence settings on four predictive models using the suggested technique. Logistic regression and three more complex predictive models—GAMs, RF, and XGBoost—are chosen. Lending Club (LC) has 2260701

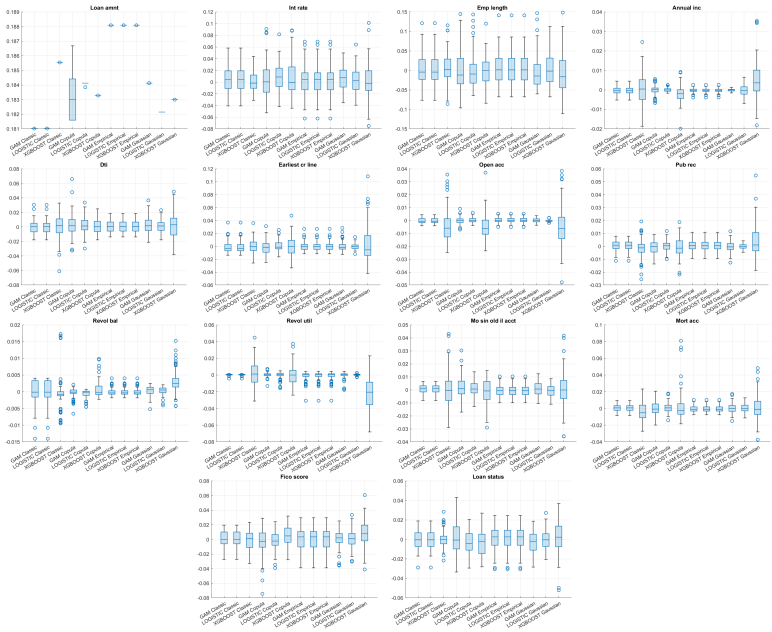

Figure 1. *Distribution of Shapley values from random subsampling for each variable, model and feature dependence structure.*

observations on individual borrowers and their requested loans from 2007 to the fourth quarter of 2018. In this study, we preprocess data and keep 14 variables to analyze the impact of dependencies on the explanations produced by the different ML models. We perform test data random sub-sampling, which provides Shapley values for each of the $n = 100$ iterations. Results are contained in Figure 1. The figure shows that Shapley value estimates are very consistent across model specifications, and that loan amount is the variable fostering the discriminatory power of all the classification models employed.

References

AAS, KJERSTI, JULLUM, MARTIN, & LØLAND, ANDERS. 2021. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence*, **298**, 103502.

OBERMEYER, ZIAD, & EMANUEL, EZEKIEL J. 2016. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, **375**(13), 1216.