Assessing Credit Risk Using Machine Learning Techniques

†FERDINANTOS KOTTAS, ‡ ANDREAS KIZIRIDIS, ††PANAGIOTIS MOUTSIOUNAS, ‡‡CHRYSANTHOS CHRISTOU
†Postgraduate Student, AUTH; Risk Data Scientist, Creditinfo, Iceland
‡ Erasmus Student, AUTH; Postgraduate Student, UNIVERSITÄT LEIPZIG
††Postgraduate Student, AUTH; Data Scientist
‡‡ Postgraduate Student, AUTH; MD

Overview

- Credit risk assessment is crucial for defining bank policies and credit limit strategies.
- The article presents the traditional approach of the FIs (Logistic regression models) for assessing the credit risk
- Performance of popular credit risk models
- Explainability of white and black boxes model. For black box using eXplainable Artificial Intelligence (XAI) tools - SHAP.

Does it add value the effort to adopt the Black box models in the financial market for better financial decisions?

Motivation

- To show that the performance of binned and non-binned datasets differ.
- Comparison of black box and white boxes performance-wise.(Feature Importance & SHAP).
- Explainability power between black box and white box models.

Accurate credit risk models are essential to address risks in mobile lending.

Methodology

On the 3 datasets, the following procedures technics were performed.

- Dataset understanding.
  Data were originally split into subsidiary datasets.
- Data preprocessing.
  - Missing Values Handling
  - Feature Engineering
  - Important Variable Creation
  - Subsidiary dataset merging.
- Descriptive statistical analysis.
- Outlier detection & clearance.
- Train, validation & test dataset split.
- Binning & Weight of Evidence
  "Binning or bucketing is a technique purposed to reduce impact of statistical noise"[1]
  - *The used algorithm the* Binning solver : CP
  - Transforms continuous variables into discrete.
  - Metrics for variable selection:
    - Gini coefficient
    - Information Value(IV) = $\Sigma((\%good/\%bad)*\text{Weight of Evidence}) = \Sigma((\%good/\%bad)*\ln(\%good/\%bad))$
- Classification models:
  - 3 white box models.
    - Decision Trees
    - Logistic Regression
    - Naïve Bayes
  - 2 black box models.
    - Random Forest
    - XGBoost
  - Stacking, uses both white and black box models.
- Model calibration.
  - Feature selection: statistical significant features selected based on Gini score.
- Hyperparameter tuning, protects from overfitting.
  - Grid search algorithm on our selected algorithms.
- Performance metrics.
  - Calibrarion curve
  - Confusion matrix.
  - ROC curve.
  - Cost matrix according to Beiling (2015)

Cost Matrix

How many € are lost for bad decisions?

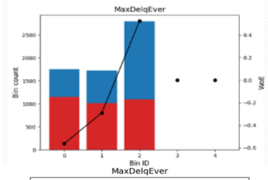
Credit risk		Ground truth	
		bad	good
Prediction	bad	0	5.386€
	good	585.947€	0

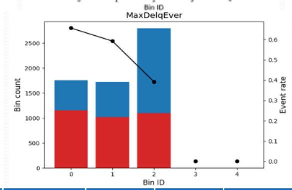
a. Cost for False Negatives
(Rate a customer as good even though defaults)
… was calculated by building the average of the customers' credit amount (drop extreme values) → the **expected loss**

b. Cost for False Positives
(Rate a customer as bad even though he pays back his credit)
… was calculated by multiplying the average credit amount with an interest rate out of the year the dataset dates from (2018) → **the opportunity of profit that is missed**

Binning of Variables

Quality score: 653.66, Gini Coef: 0.25

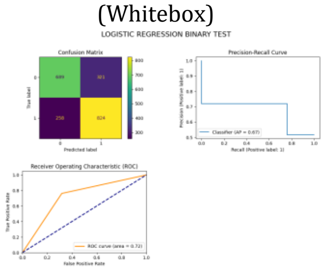


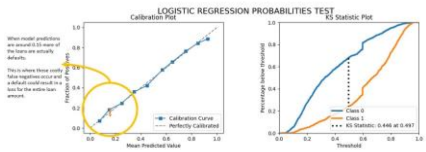


Bin	Count/ Count(%)	Event/ Event rate	WoE	IV	JS
(-inf, 5.50)	1757/0.28	1154/66%	-0.56	0.09	0.01
[5.50, 6.50)	1723/0.27	1021/59%	-0.29	0.02	0.003
[6.50, inf)	2795/0.45	1096/39%	0.52	0.12	0.01
Special	0/0	0/0%	0.0	0.0	0.0
Missing	0/0	0/0%	0.0	0.0	0.0
	6275/1	3271/52%		0.23	0.03

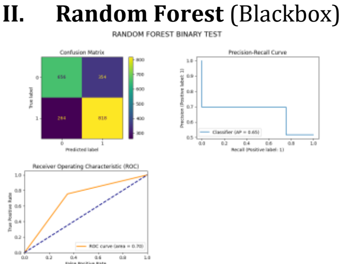
As a reference, we state the performance of specific models on dataset 3 including one whitebox and one blackbox model:

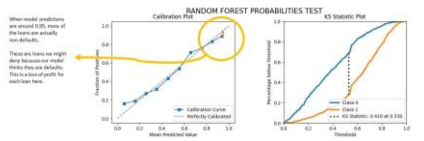
I. Logistic Regression (Whitebox)





II. Random Forest (Blackbox)





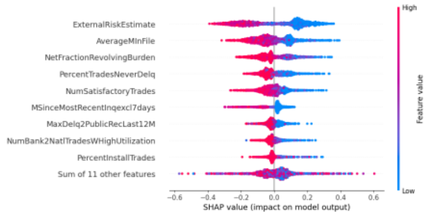
Explainability

Our black box and white box model show a similar performance. But can we also understand how both come to their decisions?

Feature Importances (Top 5) – Logistic Regression

Feature	Importance
NetFractionRevolvingBurden	0.112
NumTrades60Ever2DerogPubRec	0.085
MaxDelq2PublicRecLast12M	0.085
PercentTradesWBalance	0.068
NumInstallTradesWBalance	0.068

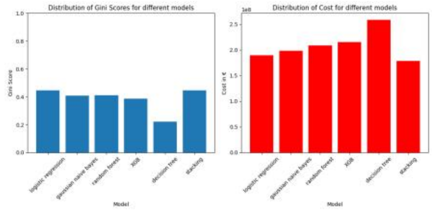
SHAP Values – Random Forest



Empirical Results

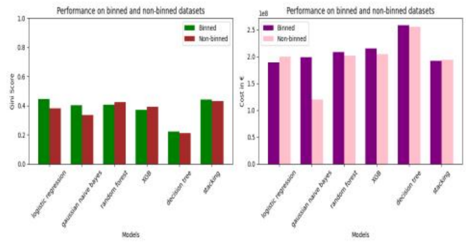
1. Performance

In the following, we will compare the results from the different models for the binned version of dataset 3.



2. Binned vs. non-binned datasets

Even though we assumed that the binning might have a considerable impact on our predictions, avoiding overfitting and making the model more generalizable, we observe overall differences between binned and not binned datasets:



Datasets Employed

	Dataset	Number of Features/Instances	Country(ies)	Period Covered
1.	Credit Default Risk	122/247032	Various	Not specified
2.	Default of Credit Card Clients	24/30000	Taiwan	April 2005 to September 2005
3.	Home Equity Line of Credit(HELOC)	23/10459	Various	Not specified

1. Home Credit Default Risk | Kaggle [Internet]. Available from: https://www.kaggle.com/competitions/home-credit-default-risk/overview
2. UCI Machine Learning Repository: default of credit card clients Data Set [Internet]. Available from: http://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients
3. Home Equity Line of Credit(HELOC) | Kaggle [Internet]. Available from: https://www.kaggle.com/datasets/averkiyoliabev/home-equity-line-of-creditheloc

White vs Black Boxes

- White and black boxes refer to different levels of transparency and understanding in machine learning models.
- White box models are characterized by their transparency, where the inner workings and decision-making processes are fully interpretable and explainable.
- Black box models are highly complex, making it difficult to comprehend their internal mechanisms and decision-making processes.
- Black boxes often exhibit superior performance due to their ability to capture intricate patterns and relationships in data, their lack of interpretability can be a challenge.
- Interpretability of black boxes refers to the ability to understand and explain the reasoning behind the model's predictions. To achieve interpretability, techniques such as feature importance analysis, surrogate models, and visualization can be employed, enabling researchers to gain insights into the decision-making process of black box models.

References

[1] "Monotone optimal binning algorithm for credit risk modeling", Pavel Mironchyk, Viktor Tchistiakov, September 6, 2017
[2] "Optimal scoring cutoff policies and efficient frontiers", Journal of the Operational Research Society, 56(9), 1016–1029, Beling, P., Covaliu, Z., & Oliver, R. M, 2005
[3] "Tuning White Box Model With Black Box Models: Transparency in Credit Risk Modeling", Gürtler, Marc and Zöllner, Marvin, May 1, 2023

ARISTOTLE UNIVERSITY OF THESSALONIKI
SCHOOL OF INFORMATICS Data & Web Science MSc Program
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΑΠΘ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ Τεχνητή Νοημοσύνη