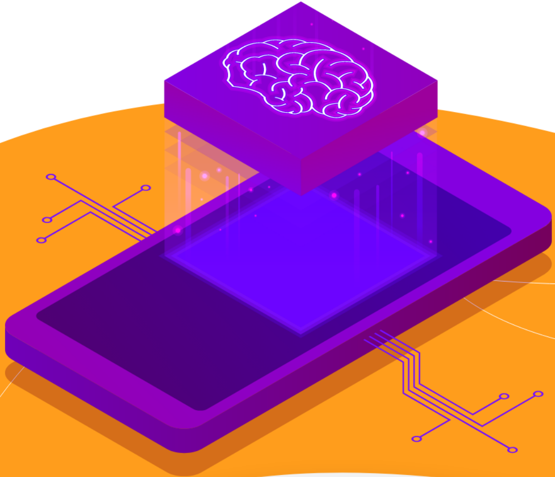


HUB DE INTELIGÊNCIA ARTIFICIAL E ARQUITETURAS COGNITIVAS

Measuring fairness of synthetic minority oversampling on credit datasets

Prof. Marcos M. Raimundo - University of Campinas

AFT2023 - A NeurIPS Workshop

December 15, 2023

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

GOVERNO FEDERAL BRASIL UNIÃO E RECONSTRUÇÃO

Softex

ELDORADO INSPIRAÇÃO PELO NOVO

UNICAMP

Instituto de Computação UNIVERSIDADE ESTADUAL DE CAMPINAS

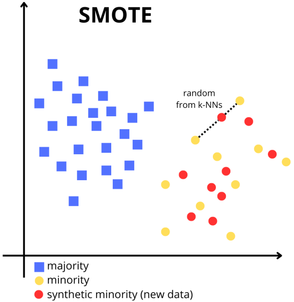
FEEC

Proposal

- **Oversampling** the minority class through synthetic generators has become a **popular solution for balancing data**, giving rise to many rebalancing techniques, like ADASYN and SMOTE.
- Practitioners usually **lean on performance metrics to either refute or advocate** for adopting some resampling method.
- Increasing ethical and legal demands for fair machine learning models demand **testing the neutrality of these methods concerning fairness**.
- We investigated **the effects of oversampling on gender bias** by analyzing statistical parity difference (SPD) and equal opportunity difference (EOD) obtained from four credit datasets.

Oversampling Methods

- SMOTE [1]
- Borderline-SMOTE [2]
- SVM-SMOTE [3]
- ADASYN [4]



Fairness

The inequality in credit scoring is evident since AB classes have 18% of denied credit. In contrast, CDE has 41% even with significantly lower amounts (above 105 thousand against thousands of reais) [5], and women have lower credit limits (23% against 28% of limits above 7600 reais) [6].

Statistical Parity Difference [7] quantifies the independence between the decision $\hat{y}(X)$ and the protected attribute Z, and is given by:

$$SPD \quad = \quad P(\hat{y}(X) = gain \mid Z = unpr) - P(\hat{y}(X) = gain \mid Z = priv).$$

Equal Opportunity Difference (EOD) [8] assesses the disparity in access to a favorable outcome $\hat{y}(X) = gain$ between the unprivileged group $Z = unpr$ and the privileged group $Z = priv$, when an individual rightfully deserves that outcome, that is, $Y = gain$ (true positive rate):

$$EOD \quad = \quad P(\hat{y}(X) = gain \mid Z = unpr,\ Y = gain) - P(\hat{y}(X) = gain \mid Z = priv,\ Y = gain).$$

Performance Metrics

The **Balanced Accuracy** is given by:

$$BalancedAccuracy = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right).$$

The **Area Under the Curve** is given by:

$$AUC = P[p(y=1|X_i) > p(y=1|X_j)|y_i = 1, y_j = 0].$$

Experimental Setup

Classifiers

- Logistic Regression
- Random Forest
- XGBoost

Oversamplers

- ADASYN
- SMOTE
- SVM-SMOTE
- Borderline-SMOTE

- Algorithms were optimized considering a utility measure to emulate a performance-driven use of imbalanced methods.
- Baselines were trained to their standard procedure (loss for logistic regression and entropy for the tree-based classifiers)
- Hyperparameter tuning used balanced accuracy to find the best parameters.
- Experiments were performed on the **same test set** (20% of total data).

Datasets

- Four **publicly available financial datasets** commonly used in credit scoring research.
- We excluded the categorical features from training data for all datasets, except for gender, the protected characteristic defining groups in our analysis.

Dataset	# samples	# features	class split (good vs. bad payers)
Home Credit	150,000	121	93.3% vs. 6.7%
Taiwan	30,000	24	78% vs. 22%
German	1,000	20	70% vs. 30%
PAKDD	39,988	28	80.2% vs. 19.8%

Results

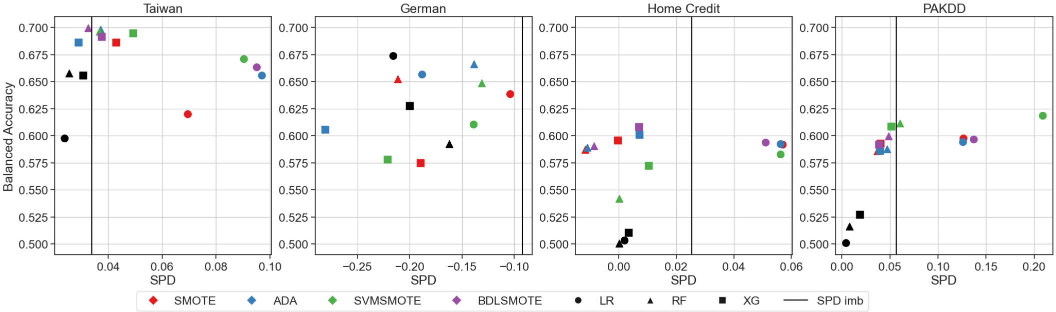


Figure 1: Fairness metrics vs. Balanced Accuracy for models with hyperparameters tuned to maximize Balanced Accuracy with a **fixed decision threshold of 0.5**.

Results

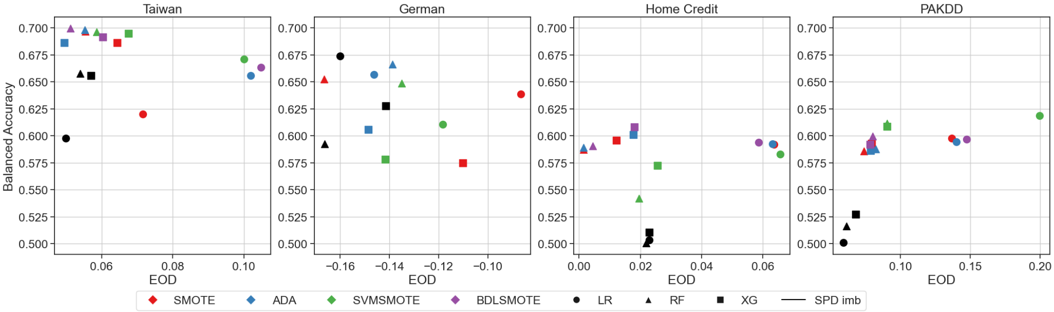


Figure 2: Fairness metrics vs. Balanced Accuracy for models with hyperparameters tuned to maximize Balanced Accuracy with a **fixed decision threshold of 0.5**.

Results

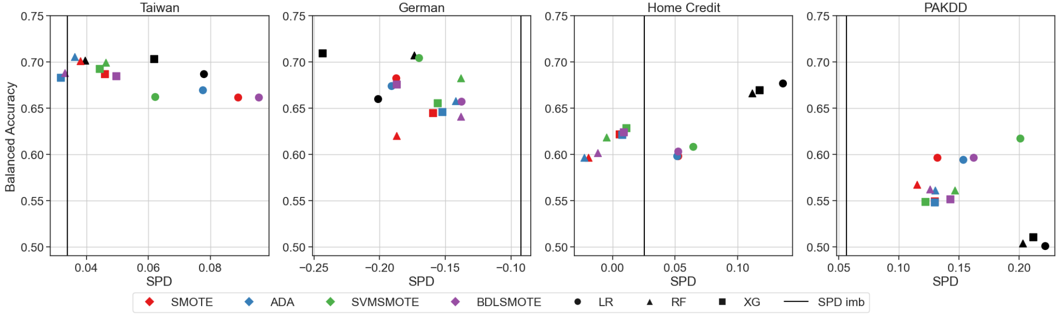


Figure 3: Fairness metrics vs. Balanced Accuracy for models with hyperparameters tuned to maximize AUC with a **threshold defined by the maximal difference between TNR and FPR**.

Results

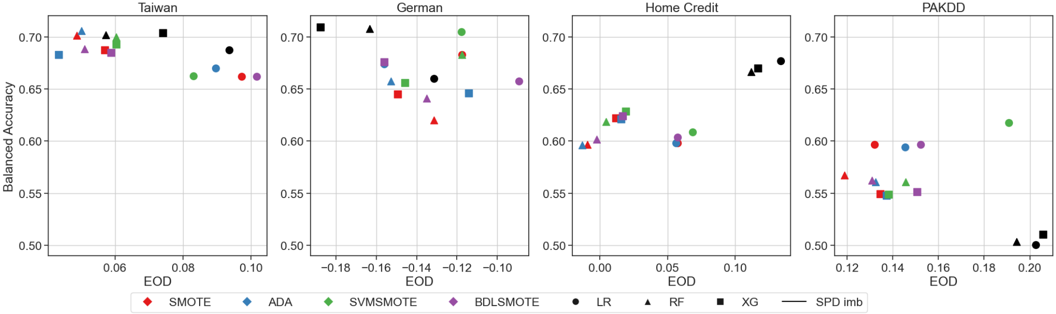


Figure 4: Fairness metrics vs. Balanced Accuracy for models with hyperparameters tuned to maximize AUC with a **threshold defined by the maximal difference between TNR and FPR**.

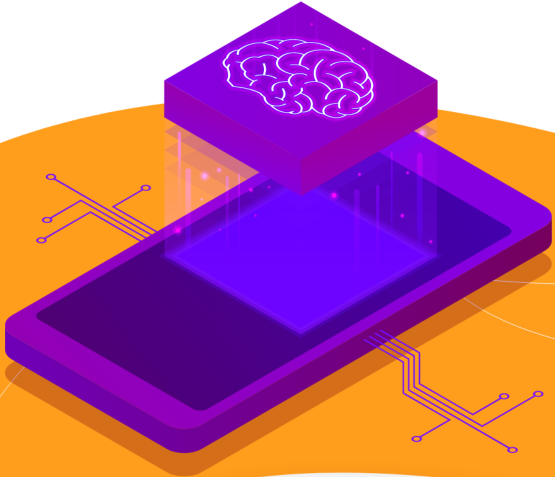
Conclusion

- Measuring the success of oversamplers requires reflections on ethical implications.
- Augmented synthetic samples may noisily replicate already discriminating samples.
- The unfettered use of oversampling can disseminate social bias.

Acknowledgements

This project was supported by the Ministry of Science, Technology, and Innovation of Brazil, with resources granted by the Federal Law 8.248 of October 23, 1991, under the PPI-Softex. The project was coordinated by Softex and published as Intelligent agents for mobile platforms based on Cognitive Architecture technology [01245.013778/2020-21].

H.IAAC | HUB DE INTELIGÊNCIA ARTIFICIAL E ARQUITETURAS COGNITIVAS



Measuring fairness of synthetic minority oversampling on credit datasets

Prof. Marcos M. Raimundo - University of Campinas

AFT2023 - A NeurIPS Workshop

December 15, 2023

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES · GOVERNO FEDERAL BRASIL UNIÃO E RECONSTRUÇÃO · Softex · ELDORADO INSPIRAÇÃO PELO NOVO · UNICAMP · Instituto de Computação UNIVERSIDADE ESTADUAL DE CAMPINAS · FEEC

References i

Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer.
SMOTE: synthetic minority over-sampling technique.
Journal of artificial intelligence research, 16:321–357, 2002.

Hui Han, Wen-Yuan Wang, and Bing-Huan Mao.
Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning.
In *International conference on intelligent computing*, pages 878–887. Springer, 2005.

Yuchun Tang, Yan-Qing Zhang, Nitesh V Chawla, and Sven Krasser.
Svms modeling for highly imbalanced classification.
IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39(1):281–288, 2008.

References ii

Haibo He, Yang Bai, Edwardo A Garcia, and Shutao Li.
ADASYN: Adaptive synthetic sampling approach for imbalanced learning.
In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, pages 1322–1328. Ieee, 2008.

Levantamento Serasa/Opinion box.
O papel do crédito em um momento de retomada.
Serasa, 2021.

IBGE.
Acesso à internet e à televisão e posse de telefone móvel celular para uso pessoa.
IBGE, 2019.

Kush R. Varshney.
Trustworthy Machine Learning.
Independently Published, Chappaqua, NY, USA, 2022.

References iii

Moritz Hardt, Eric Price, and Nati Srebro.
Equality of opportunity in supervised learning.
Advances in neural information processing systems, 29, 2016.