Fair-SMOTE vs the AIF360 toolkit

ANDRE LUSTOSA, North Carolina State University, USA

ACM Reference Format:
Andre Lustosa. 2021. Fair-SMOTE vs the AIF360 toolkit. 1, 1 (October 2021), 1 page. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 PROPOSAL

For my CSC-791 Project I will explore the universe of different algorithms made available in the AIF360 toolkit[1]. These different algorithms come from a myriad of different previous works in the literature of fairness. These algorithms will then be compared to the results obtained by Chakraborty et al. [2].

As stated by Chakraborty et al. fairness algorithms will usually fall within one of three categories:

- **Pre-processing** - Optimized Preprocessing [1], Reweighing [4].
- **In-processing** - Adversarial Debiasing [7], Prejudice Remover Regularizer [6].
- **Post-processing** - Equalized Odds [3], Reject Option Classification [5].

On the opposite end of these algorithms, Fair-SMOTE [2] addresses the problem through all three steps, through pre, in and post processing. However the Fair-SMOTE paper fails to address comparisons to these existing algorithms. As such this work has the objective of running all 7 algorithms described above in different datasets and comparing the obtained results to those of Fair-SMOTE.

ACKNOWLEDGMENTS

To Joymallya and Kewen, for explaining me and guiding me through the AIF360 toolkit.

REFERENCES

[1] Flavio P Calmon, Dennis Wei, Bhanukiran Vinzamuri, Karthikeyan Natesan Ramamurthy, and Kush R Varshney. Optimized pre-processing for discrimination prevention. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 3995–4004, 2017.
[2] Joymallya Chakraborty, Suvodeep Majumder, and Tim Menzies. Bias in machine learning software: Why? how? what to do? *arXiv preprint arXiv:2105.12195*, 2021.
[3] Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. *Advances in neural information processing systems*, 29:3315–3323, 2016.
[4] Faisal Kamiran and Toon Calders. Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems*, 33(1):1–33, 2012.
[5] Faisal Kamiran, Asim Karim, and Xiangliang Zhang. Decision theory for discrimination-aware classification. In *2012 IEEE 12th International Conference on Data Mining*, pages 924–929, 2012.
[6] Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma. Fairness-aware classifier with prejudice remover regularizer. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 35–50. Springer, 2012.
[7] Brian Hu Zhang, Blake Lemoine, and Margaret Mitchell. Mitigating unwanted biases with adversarial learning. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pages 335–340, 2018.

[1] Available at https://github.com/Trusted-AI/AIF360

Author's address: Andre Lustosa, North Carolina State University, Raleigh, USA, alustos@ncsu.edu.

XXXX-XXXX/2021/10-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn