| | |
|---|---|
| **Project Name** | *Boosting for Fairness-Aware Classification* |
| **Project Type** | *Replication/Research* |
| **Team Name** | *Group 20* |
| **Project Team** | *Daniil Babin - experience with classification problems (Logistic Regression, KNeighbours, Decision Trees) and NLP*<br>*Sudarut Kasemsuk - experienced with Logistic Regression for computation loan default rates.*<br>*Waralak Pariwatphan - experienced with Decision Trees to create the take-up models on loan.* |
| **Who suggested** | *Nina Mazyavkina* |
| **Difficulty** | *Easy/Medium* |
| **Proposal** | *Fairness-aware classification, which aims to reduce the algorithms' prejudice towards the minorities (e.g. in terms of race, gender or nationality) in various domains such as loan credit and/or job hiring, is a hot topic in the international ML society right now. This problem can in some way be reduced to the imbalanced classification and minority class prediction. In the provided papers authors try different algorithms for imbalanced classification for popular real-world datasets (UCI Adult Census, UCI KDD Census etc).*<br><br>*In this project, we will focus on reproducing, improving and creating the new algorithms and models for fairness-aware classification for reproduction. In addition, we propose an AdaBoost model to create weight and adjust the process into balanced error for preventing discrimination. The datasets for training and testing are four real-world datasets: Adult census income, Bank, Compass, and KDD census income.* |
| **Specific Tasks & Expected results** | ➢ *1) Implement several algorithms from the listed articles and/or other algorithms, provided by the students*<br>➢ *2) Evaluate and compare those models in terms of predictive and fairness performance on the listed datasets (more on fairness evaluation can be found in https://arxiv.org/pdf/1909.08982.pdf)*<br>➢ *3) Provide the interpretation of the showed results (find the features that increase/decrease unfairness in the provided datasets, how fairly balancing the datasets affects the overall model accuracy etc)* |
| **Relevant papers** | ➢ *Vasileios Iosifidis, Eirini Ntoutsi. 2019. AdaFair: Cumulative Fairness Adaptive Boosting. https://arxiv.org/pdf/1909.08982.pdf*<br>➢ *Nitesh V. Chawla, Aleksandar Lazarevic, Lawrence O. Hall, Kevin W. Bowyer. 2003. SMOTEBoost: Improving Prediction of the Minority Class in Boosting. Knowledge Discovery in Databases: PKDD 2003, 107-119. https://link.springer.com/chapter/10.1007/978-3-540-39804-2_12* |

| | |
|---|---|
| | ➢ *Emmanouil Krasanakis, Eleftherios Spyromitros-Xioufis, Symeon Papadopoulos, Yiannis Kompatsiaris. 2018. Adaptive Sensitive Reweighting to Mitigate Bias in Fairness-aware Classification. https://dl.acm.org/doi/pdf/10.1145/3178876.3186133?download=true*<br>➢ *Muhammad Bilal Zafar, Lsabel Valera, Manuel Gomez Rodriguez, Krishna P. Gummadi. 2017. Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification without Disparate Mistreatment. https://arxiv.org/pdf/1610.08452.pdf* |
| **Grading Scheme** | *70% successfully implement the models from the articles (at least 2), compare the models and report the results in terms of performance and fairness evaluation, provide some interesting interpretation of the results*<br><br>*90% successfully implement the models from the listed or any other articles (more than 2) and or create your own approach to deal with the datasets' imbalance, compare the models and report the results in terms of performance and fairness evaluation, provide some interesting interpretation of the results*<br><br>*All the experiments should be done using the mentioned datasets (Adult Census, KDD Census, Bank, Compass)* |