Promoting Fairness through Hyperparameter Optimization

André Cruz, Pedro Saleiro, Catarina Belém, Carlos Soares, and Pedro Bizarro

feedzai

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Motivation & Contributions

1. Current hyperparameter optimization is fairness blind, and unwittingly optimizes for unfair models.
2. Guiding the search towards fairer regions is a natural solution that can be achieved through multi-objective optimization.
3. Fair HO promotes fairness without changing the pipeline or the methods themselves, providing no obstacles to its widespread adoption.
4. Results on real-world data show **steep fairness increases** (68% to 147%) at small predictive accuracy costs (4.7% to 7.5%).

Approach

We propose fairness-aware variants of popular algos:

Fairband, **Fair TPE**, **Fair Random Search**.

- We use a scalarizing function to optimize for the best fairness-accuracy trade-off.

$$g(\lambda) = \alpha \cdot a(\lambda) + (1 - \alpha) \cdot f(\lambda)$$

- We propose a heuristic to automatically set the trade-off parameter, α, based on the expected value of the model's accuracy and fairness.

$$\alpha = 0.5 \cdot (\mathbb{E}_{D \subseteq \Lambda}[f(\lambda)] - \mathbb{E}_{D \subseteq \Lambda}[a(\lambda)]) + 0.5$$

- FB-auto requires no domain knowledge and can be **effortlessly integrated** into current pipelines.

Algo.	Adult				COMPAS				Donors Choose			
	Validation		Test		Validation		Test		Validation		Test	
	Acc.	Fair.	Acc.	Fair.	Acc.	Fair.	Acc.	Fair.	Acc.	Fair.	Acc.	Fair.
FB-auto	92.0$^{\blacklozenge}$	94.7$^{\blacklozenge}$	91.6$^{\blacklozenge}$	90.9$^{\blacklozenge}$	74.0$^{\blacklozenge}$	95.8$^{\blacklozenge}$	70.1$^{\lozenge}$	90.0$^{\blacklozenge}$	54.2$^{\blacklozenge}$	98.2$^{\blacklozenge}$	50.7$^{\blacklozenge}$	86.5$^{\blacklozenge}$
FB	92.7$^{\blacklozenge}$	94.0$^{\blacklozenge}$	92.3$^{\blacklozenge}$	89.5$^{\blacklozenge}$	71.2$^{\blacklozenge}$	95.5$^{\blacklozenge}$	67.6$^{\blacklozenge}$	80.7$^{\blacklozenge}$	54.2$^{\blacklozenge}$	97.7$^{\blacklozenge}$	50.4$^{\blacklozenge}$	85.5$^{\blacklozenge}$
FairRS	93.6$^{\blacklozenge}$	79.4$^{\blacklozenge}$	93.8$^{\blacklozenge}$	78.6$^{\blacklozenge}$	67.4$^{\blacklozenge}$	77.4$^{\blacklozenge}$	64.2$^{\blacklozenge}$	67.8$^{\blacklozenge}$	51.7$^{\blacklozenge}$	97.0$^{\blacklozenge}$	50.4$^{\blacklozenge}$	79.5$^{\blacklozenge}$
FairTPE	93.3$^{\blacklozenge}$	82.2$^{\blacklozenge}$	93.5$^{\blacklozenge}$	80.7$^{\blacklozenge}$	67.1$^{\blacklozenge}$	81.8$^{\blacklozenge}$	63.9$^{\blacklozenge}$	69.5$^{\blacklozenge}$	52.3$^{\blacklozenge}$	96.3$^{\blacklozenge}$	50.6$^{\blacklozenge}$	79.1$^{\blacklozenge}$
HB	99.4	53.5	99.0	54.1	78.1	45.4	73.6	51.2	60.9	28.7	53.6	35.0
RS	99.4	55.7	99.1	56.6	77.7	43.8	73.2	43.4	59.9	24.9	53.4	32.4
TPE	99.4	54.9	99.1	55.6	78.0	42.8	73.5	46.6	61.0	27.1	53.3	33.4

Figure 1: Results for all hyperparameter tuners on the Adult Income, COMPAS, and Donors Choose datasets. Over 10K models were trained on each dataset. FB-auto consistently achieves Pareto efficient results.

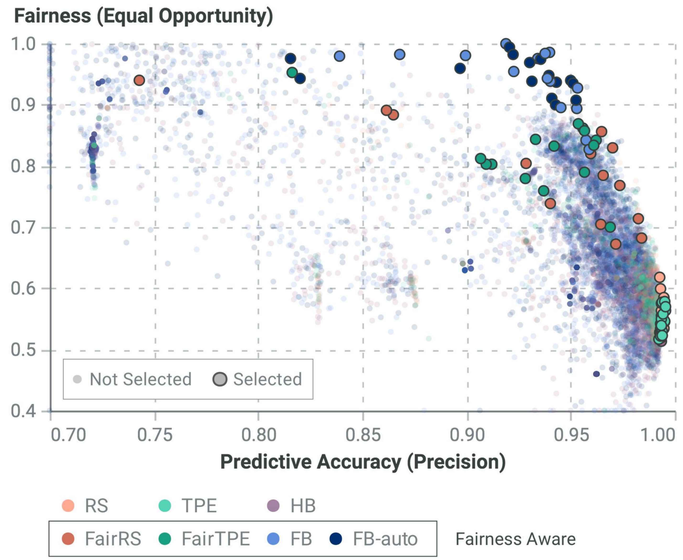

Figure 2: [Adult] Fairness vs predictive accuracy per tuner.

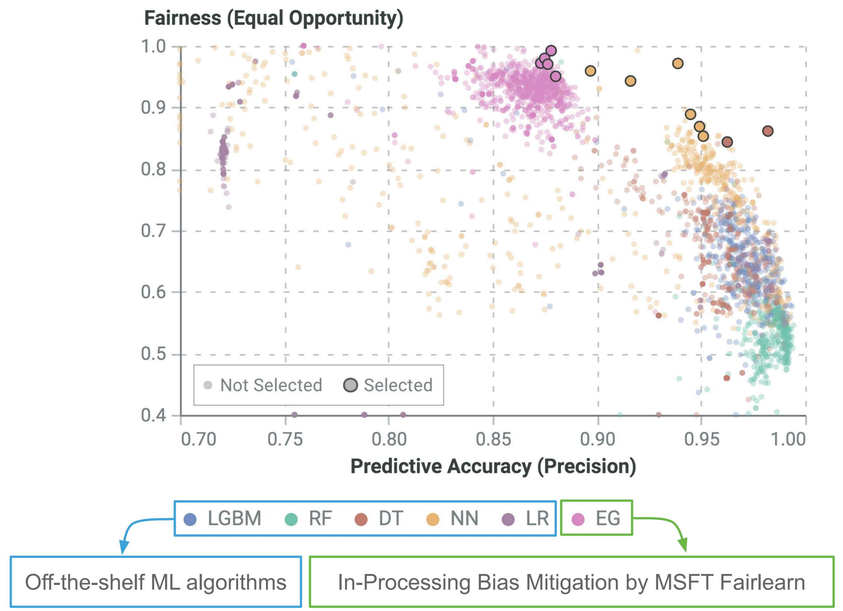

Figure 3: [Adult] Fairness vs predictive accuracy of models selected by FB-auto. Off-the-shelf bias-blind ML algorithms can dominate bias reduction methods via hyperparameter tuning.

The project CAMELOT (reference POCI-01-0247-FEDER-045915) leading to this work is co-financed by the ERDF - European Regional Development Fund through the Operational Program for Competitiveness and Internationalisation - COMPETE 2020, the North Portugal Regional Operational Program - NORTE 2020 and by the Portuguese Foundation for Science and Technology - FCT under the CMU Portugal international partnership.