Skoltech
Skolkovo Institute of Science and Technology

BOOSTING FOR FAIRNESS-AWARE CLASSIFICATION

<COURSE : MACHINE LEARNING 2022>

GROUP : 20

TEAM MEMBER

- DANIIL BABIN
- SUDARUT KASEMSUK
- WARALAK PARIWATPHAN

OUTLINE :

01 PROBLEM STATEMENT AND RELATED WORK

02 ALGORITHMS AND MODELS

03 EXPERIMENTS AND RESULTS

04 CONCLUSION

01
PROBLEM STATEMENT & RELATED WORK

Motivation :

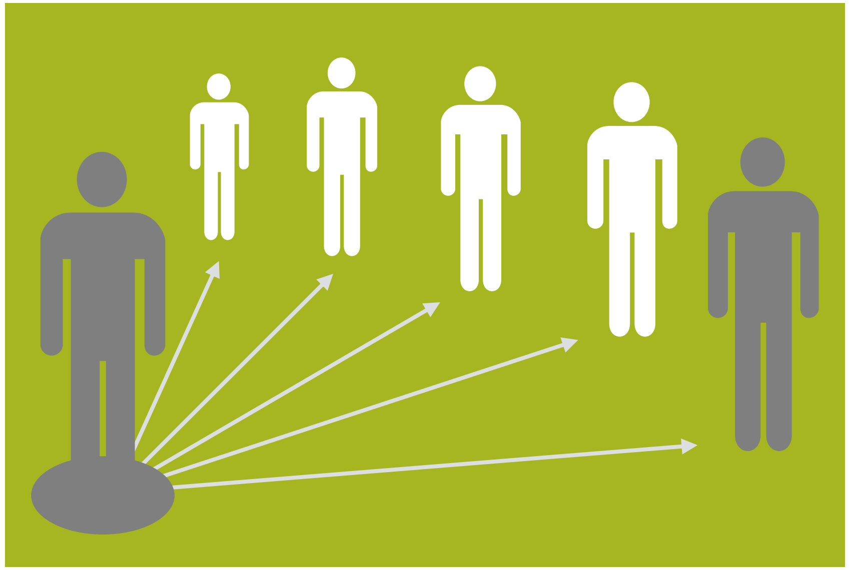
- Machine Learning is useful tool and currently popular in many industries.
  ML algorithms are used in the financial industry and others to decision-making.

- For example, Machine Learning are asked to predict which items will be placed in your shopping cart based on your previous purchase history.

- But for some predictions, algorithms can lead to discrimination.
  Such as job hiring, loan credit or recidivism.

BOOSTING FOR FAIRNESS-AWARE CLASSIFICATION

Skoltech
Skolkovo Institute of Science and Technology



Job Hiring

ML's algorithms can give different conclusions. It may lead to unfair discrimination depends on sensitive variables such as gender or race. In case of Google's AdFisher tool showed significantly more high-paying job ads to men than women. Meanwhile, there is research paper titled "Man is to Computer Programmer as Women is to Homemaker? Debiasing Word Embeddings", which showed man → computer-programmer, and women → homemaker. For this reason, AdaFair Algorithms helped to solve the problem of fairness in AI algorithms.

BOOSTING FOR FAIRNESS-AWARE CLASSIFICATION

Skoltech
Skolkovo Institute of Science and Technology

Related Work

AdaBoost:
Boosting is an ensemble technique that combines weak learners to create a strong learner. AdaBoost calls a weak learner iteratively by adjusting the instance weights in each iteration based on misclassified instances. Boosting is a promising technique for fairness-aware classification as it divides the learning problem into multiple sub-problems and then combines their solutions into an overall model. In order to apply AdaBoost for fairness one has to carefully change the underlying data distribution between consecutive rounds so that both predictive performance aspects and fairness-related aspects are considered.

SMOTE:
Synthetic Minority Oversampling Technique was proposed to counter the effect of having few instances of the minority class in a data set. SMOTE creates synthetic instances of the minority class by operating in the “feature space” rather than the “data space”. By synthetically generating more instances of the minority class, the inductive learners, such as decision trees or rule-learners, are able to broaden their decision regions for the minority class. In the nearest neighbor computations for the minority classes use Euclidean distance for the continuous features and the Value Distance Metric for the nominal features.

02
ALGORITHMS AND MODELS

BOOSTING FOR FAIRNESS-AWARE CLASSIFICATION

Skoltech
Skolkovo Institute of Science and Technology

Problem Statement and Related Work | **Algorithms and Models** | Experiment and Results | Conclusion

Basic Concepts:

Dataset D consist of n samples drawn from a joint distribution P(F, S, y)

S - denotes sensitive attribute such as gender and race
F - denote other non-sensitive attributes
y - is the class label

Consider y in binary class: $y \in \{+, -\}$
A single attribute S also binary: $S \in \{s, s'\}$
with s is protected and s' is non-protected group

The goal of classification is to find a mapping function $f: (F, S) \rightarrow y$
to predict the class labels of future unseen instance.

BOOSTING FOR FAIRNESS-AWARE CLASSIFICATION

Skoltech
Skolkovo Institute of Science and Technology

AdaFair Algorithm

Algorithm 1 AdaFair

Input: $D = (x_i, y_i)_1^N, T, \epsilon$
Initialize $w_i = 1/N$ and $u_i = 0$, for $i = 1, 2, ..., N$
for $j = 1$ **to** T **do**
(a) Train a classifier h_j to the training data using weights w_i.
(b) Compute the error rate $err_j = \frac{\sum_{i=1}^N w_i I(y_i \neq h_j(x_i))}{\sum_{i=1}^N w_i}$
(c) Compute the weight $\alpha_j = \frac{1}{2} \cdot log(\frac{1-err_j}{err_j})$
(d) Compute fairness-related $\delta\text{FNR}^{1:j}$
(e) Compute fairness-related $\delta\text{FPR}^{1:j}$
(f) Compute fairness-related costs u_i
(g) Update the distribution as
$w_i \leftarrow \frac{1}{Z_j} w_i \cdot e^{\alpha_j \cdot \hat{h}_j(x) I(y_i \neq h_j(x_i))} \cdot (1 + u_i)$
//Z_j is normalization factor; $\hat{h}_j$ is the confidence score
end for
Output: $H(x) = \sum_{j=1}^T \alpha h_j(x)$

SMOTEBoost Algorithm

Algorithm 2 SMOTEBoost

Given: Set $S(x_1, y_1), .., (x_m, y_m) x_i \in X$, with label in $y_i \in Y = 1, ..., C$, where $C_p, (C_p < C)$ corresponds to a minority (positive) class
Let $B = (i, j) : i = 1, ..., m, y \neq y_i$
Initialize the distribution D_1 over the examples, such that $D_1(i) = 1/m$
for $i = 1$ **to** T **do**
(a) Modify distribution D_t by creating N synthetic examples from minority class C_p using the SMOTE algorithm
(b) Train a weak learner using distribution D_t
(c) Compute weak hypothesis $h_t = X \times Y \to [0, 1]$
(d) Compute the pseudo-loss of hypothesis h_t : $\epsilon_t = \sum_{(i,j)\in B} D_t(i, y)(1 - h_t(x_i, y_i) + h_t(x_i, y))$
(e) Set $\beta_t = \epsilon_t/(1 - \epsilon_t)$ and $w_t = (1/2) \cdot (1 - h_t(x_i, y) + (h_t(x_i, y_i))$
(f) Update $D_t : D_{t+1}(i, y) = (D_t(i, y)/Z_t) \cdot \beta_t^{w_t}$ where Z_t is a normalization constant chosen such that D_{t+1} is a distribution
end for
Output: $h_{fn} = argmax_{y \in Y} \sum_{t=1}^T (log \frac{1}{\beta_t}) \cdot h_t(x, y)$

Skolkovo Institute of Science and Technology

03
EXPERIMENTS AND RESULTS

BOOSTING FOR FAIRNESS-AWARE CLASSIFICATION

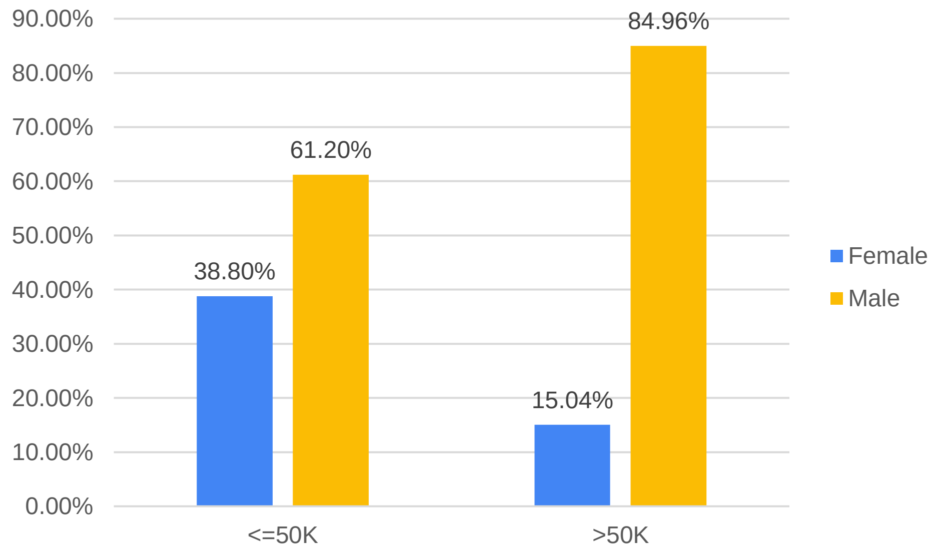
Skoltech
Skolkovo Institute of Science and Technology

Problem Statement and Related Work | Algorithms and Models | Experiment and Results | Conclusion

Datasets Table

Characteristics	Adult	Bank	Compas	KDD
Instances	32,561	49,732	7,214	199,523
Attributes	15	17	53	42
Sensitive Attr.	Gender	Marital	Gender	Gender
Class ratio (+:-)	1:2.02	1:1.52	1:4.17	1.09:1
Positive class	50K+	yes	1	50000+

Bar chart shown discrimination of income between Male and Female



BOOSTING FOR FAIRNESS-AWARE CLASSIFICATION

Skoltech
Skolkovo Institute of Science and Technology

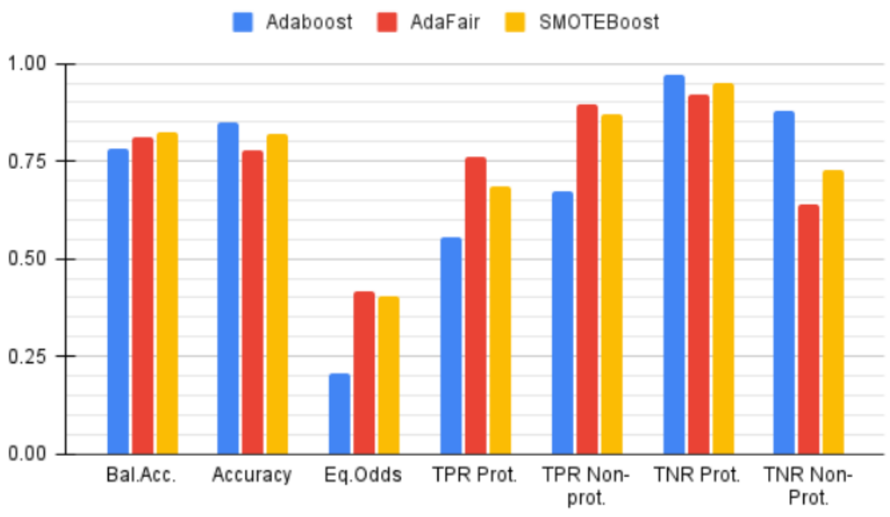


Figure 1. Performance of algorithms on Adult Census dataset.

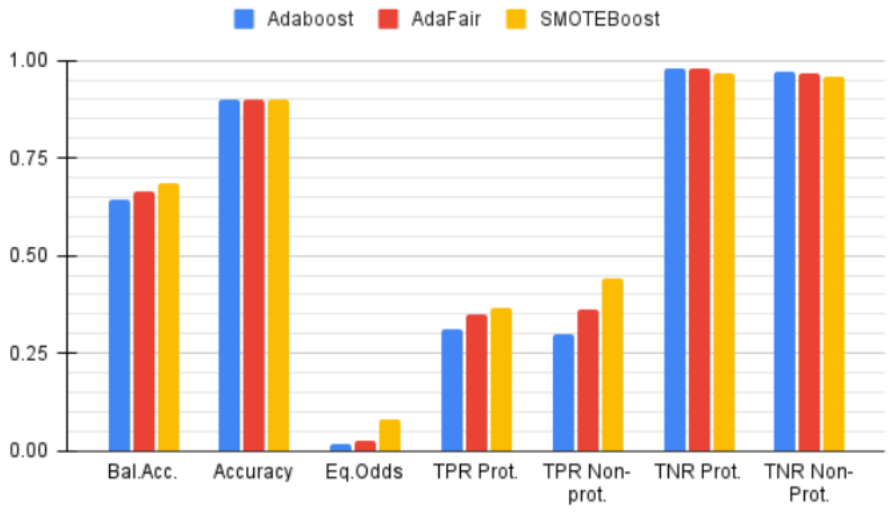


Figure 2. Performance of algorithms on Bank Census dataset.

BOOSTING FOR FAIRNESS-AWARE CLASSIFICATION

Skoltech
Skolkovo Institute of Science and Technology

Problem Statement and Related Work | Algorithms and Models | Experiment and Results | Conclusion

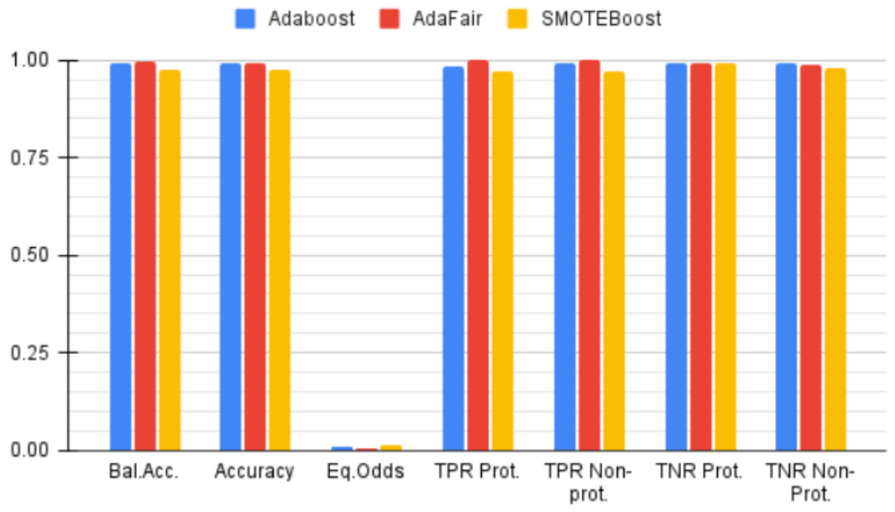


Figure 3. Performance of algorithms on Compas dataset.

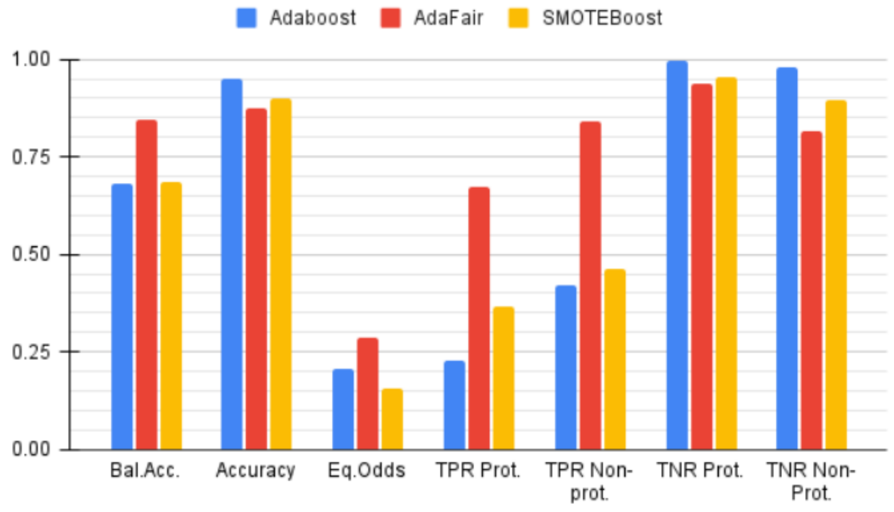


Figure 4. Performance of algorithms on KDD Census dataset.

04
CONCLUSION

- Replicate the AdaFair and SMOTEBoost algorithms to reduce their bias. An evaluate predictive and fairness performance by balance accuracy, accuracy, TPR, and TNR of protected and non-protected classes.

- Moreover, AdaFair is able to achieve the accuracy and fairness (balance accuracy and TPR) in both extreme class-imbalance and nearly class-balance.

- The outcome of analyzing reveals a significant difference in TPR score for both protected and non-protected classes. TPR and balance accuracy can be outperformed by AdaFair and SMOTEBoost. On the other hand, improving unfair classification algorithms can have an effect on accuracy, TNR, and equalized odds scores in some cases.

THANK YOU