Maximizing Accuracy under Fairness Constraints (C-LR and C-SVM) & Information Theoretic Measures for Fairness-Aware Feature selection (FFS)

Group 5: Chang Lu, Jiaxin Yu, Marcus Loke, Xiran Lin, Zaigham Khan
Presenter: Xiran Lin

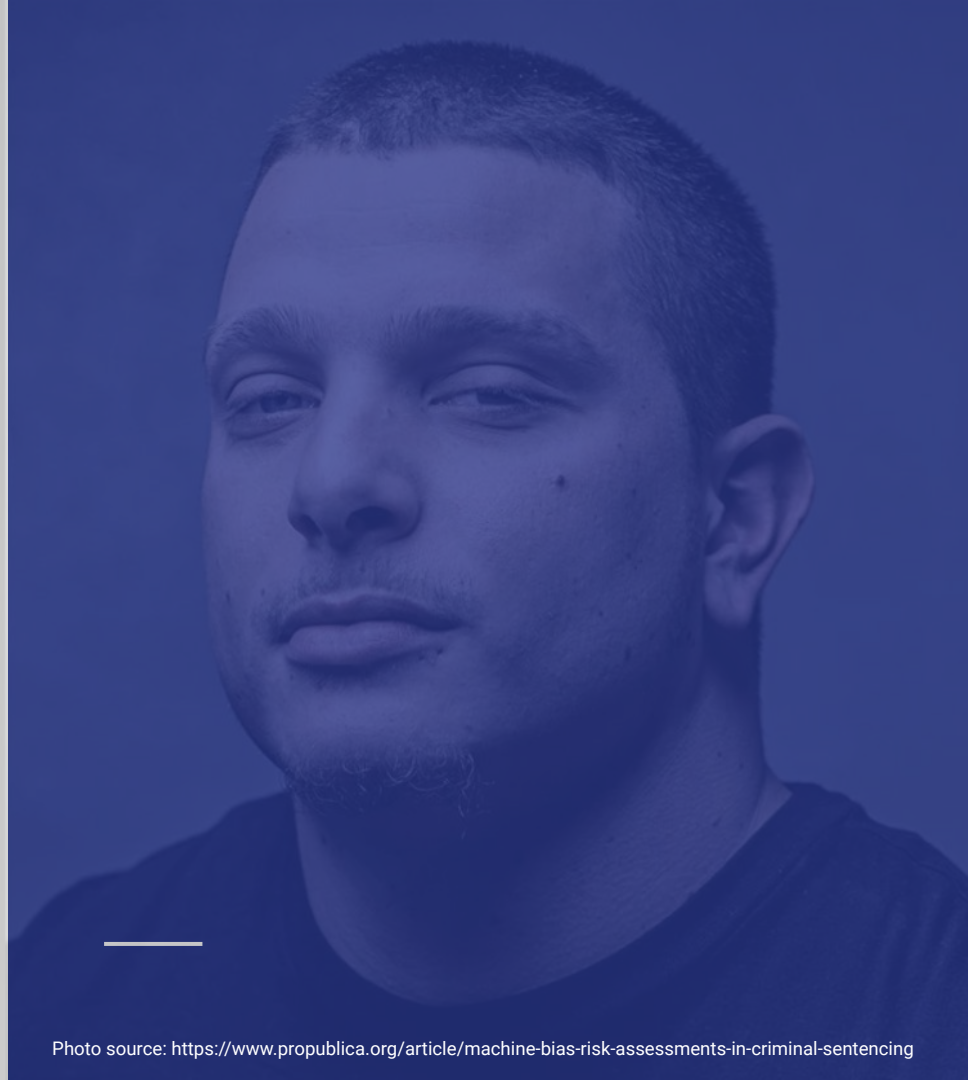
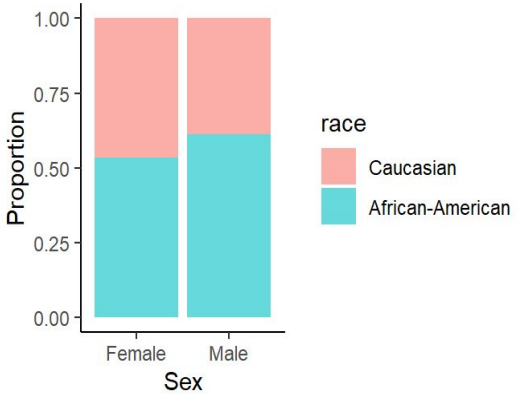
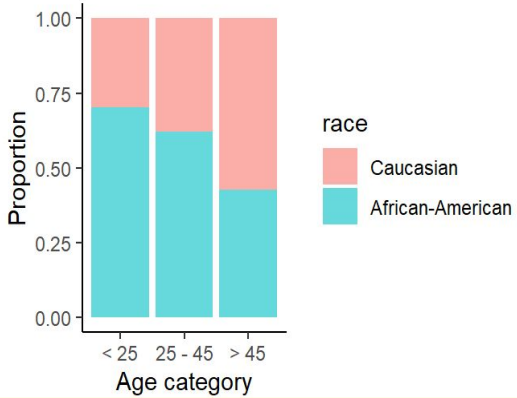
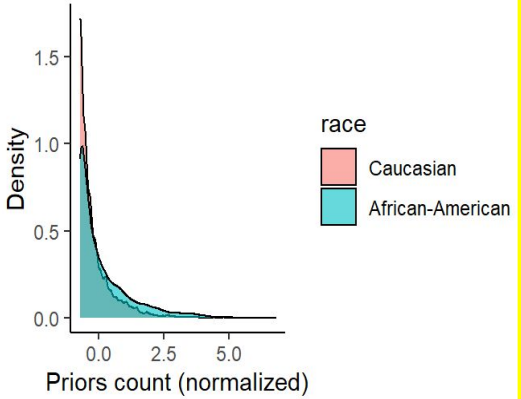


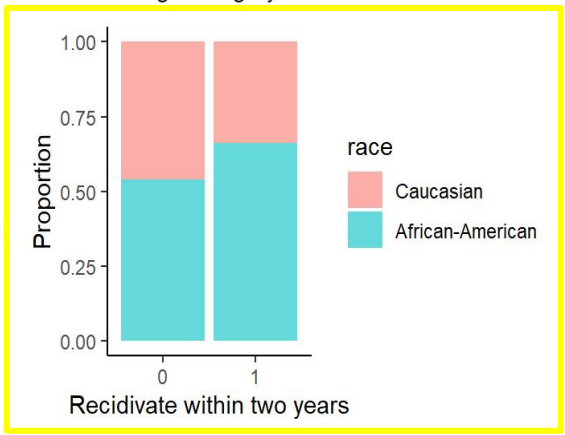
Photo source: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Exploratory Data Analysis










```
##
## Call:
## glm(formula = two_year_recid ~ ., family = "binomial", data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.0379  -1.0104  -0.6439   1.0865   1.9708
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)            0.38552    0.09604   4.014 5.96e-05 ***
## sexMale                0.31628    0.07143   4.428 9.52e-06 ***
## age_cat25 - 45        -0.72934    0.07091 -10.285  < 2e-16 ***
## age_cat> 45           -1.44579    0.09282 -15.576  < 2e-16 ***
## raceAfrican-American   0.09351    0.05905   1.583   0.1133
## priors_count           0.74182    0.03829  19.375  < 2e-16 ***
## c_charge_degreeM      -0.12941    0.06026  -2.148   0.0317 *
## length_of_stay         0.14413    0.03232   4.459 8.22e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8188.2  on 5914  degrees of freedom
## Residual deviance: 7338.3  on 5907  degrees of freedom
## AIC: 7354.3
##
## Number of Fisher Scoring iterations: 4
```

A2 Algorithm

- A2: Maximizing accuracy under fairness constraints

$$\begin{aligned} \text{minimize} \quad & -\sum_{i=1}^{N} \log p(y_i|\mathbf{x}_i, \boldsymbol{\theta}) \\ \text{subject to} \quad & \frac{1}{N}\sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}})\, \boldsymbol{\theta}^T \mathbf{x}_i \leq \mathbf{c}, \\ & \frac{1}{N}\sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}})\, \boldsymbol{\theta}^T \mathbf{x}_i \geq -\mathbf{c}, \end{aligned}$$

C-LR: minimize -ve log likelihood subject to cross-covariance between sensitive variables and the distance to the hyperplane.

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{b}\|^2 + C\sum_{i=1}^{n} \xi_i \\ \text{subject to} \quad & y_i(\mathbf{b}^T[-1\ \mathbf{x}_i]) \geq 1 - \xi_i, \forall i \in \{1, \ldots, n\} \\ & \xi_i \geq 0, \forall i \in \{1, \ldots, n\}, \\ & \frac{1}{N}\sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}})\, \mathbf{b}^T[-1\ \mathbf{x}_i] \leq \mathbf{c}, \\ & \frac{1}{N}\sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}})\, \mathbf{b}^T[-1\ \mathbf{x}_i] \geq -\mathbf{c}. \end{aligned}$$

C-SVM: minimize the margin between support vectors under penalty where **b** is the weight vectors; subject to cross-covariance between sensitive variables and the distance to the hyperplane.

In both case, **c** controls the trade-offs between accuracy and parity (or p-rule).

A7 Algorithm

- A7: Information Theoretic Measures for Fairness-aware Feature selection (FFS)

In this paper, the authors propose two information-theoretic measures that separately quantify the accuracy and discriminatory impact of subsets of features. They also deduce the marginal impacts of each feature using Shapley-value analysis

Accuracy Coefficient is given by:

$$v^{Acc}(X_S) = I(Y; X_S|\{A, X_{S^c}\}) = UI(Y; X_S\backslash\{A, X_{S^c}\}) + CI(Y; X_S, \{A, X_{S^c}\})$$

Discrimination coefficient is given by:

$$v^D(X_S) \triangleq SI(Y; X_S, A) \times I(X_S; A) \times I(X_S; A|Y)$$

Marginal impact of each feature is then calculated using the Shapley values. The Shapley value is a solution concept used in game theory that involves fairly distributing both gains and costs to several actors working in coalition.

A7 Algorithm

- A7: Information Theoretic Measures for Fairness-aware Feature selection (FFS)

	Feature	Shapley Discrimination	Shapley Accuracy
0	priors_count	25508.281363	1.264251
1	length_of_stay	25483.034007	1.048422
2	age_cat	21627.423734	1.096104
3	sex	20962.580750	0.941318
4	c_charge_degree	20764.750822	1.036236

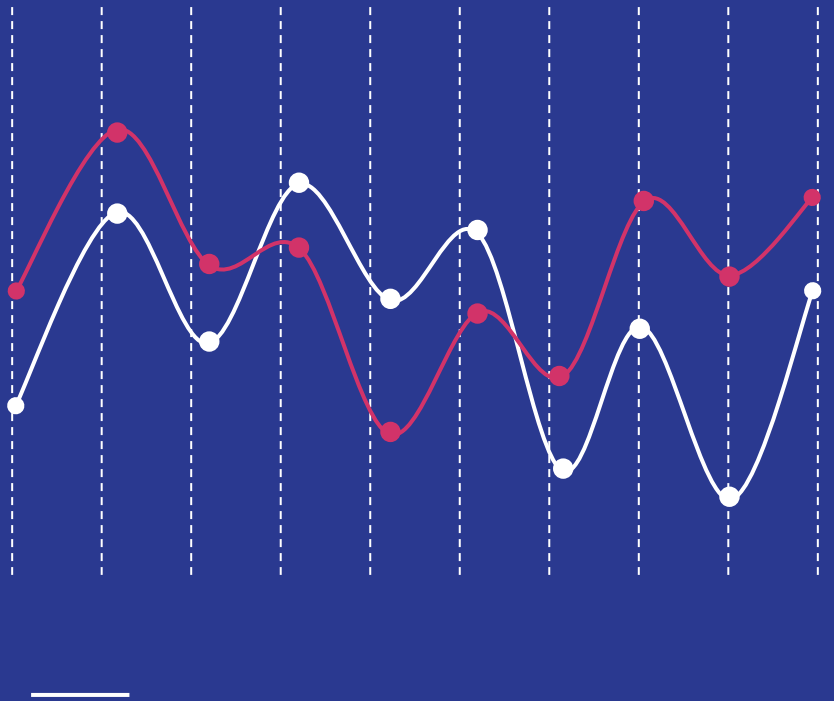
Results Summary and Evaluation

We evaluate the performance of each model using accuracy and calibration.

Classifier	LR	C-LR	SVM	C-SVM	FFS-LR	FFS-SVM
Set	Test	Test	Test	Test	Test	Test
Accuracy (%)	64.957746	46.084507	65.183099	46.028169	65.070423	65.577465
P-rule (%)	61.64272	99.955856	60.848593	99.955856	61.94468	61.602509
Protected (%)	33.888889	99.861111	30.972222	99.861111	34.583333	33.75
Not protected (%)	54.976303	99.905213	50.900474	99.905213	55.829384	54.78673
Calibration (%)	1.005793	9.53594	2.262375	9.441153	0.115192	0.664165

- **FFS-LR** was selected because of good calibration scores and high accuracy
- But if parity is important to user, then we will use the constrained models
- Although FFS takes slightly longer to run (~10sec), it is not detrimental to our project

Q&A



References

- https://towardsdatascience.com/optimization-with-scipy-and-application-ideas-to-machine-learning-81d39c7938b8
- https://github.com/mbilalzafar/fair-classification/tree/master/disparate_impact
- https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis
- https://arxiv.org/abs/2106.00772
- https://github.com/SreeranjaniD/Fairness-in-Classification-using-SVM