Anastasia Gergel
June 27, 2023

Memo: Evaluation of Predictive Models for Unemployment Insurance

Dear <EDD Stakeholders>,

I am writing to present the evaluation of the two candidate predictive models generated by the consultancy for predicting the employment status of individuals on unemployment insurance. The models, referred to as Model 1 and Model 2, provide predicted probabilities of individuals still being unemployed six months after starting unemployment insurance. The purpose of this evaluation is to help EDD decide which model to deploy, considering the goal of non discrimination and avoiding bias against marginalized groups.

1. Data Overview:
The evaluation was performed on the EDD dataset, which contains information about individuals on unemployment insurance. The dataset includes variables such as demographic information, unemployment insurance start dates in 2020, and the predicted probabilities from Model 1 (m1_pred_prob) and Model 2 (m2_pred_prob). The dataset comprises observations for 52,843 clients.

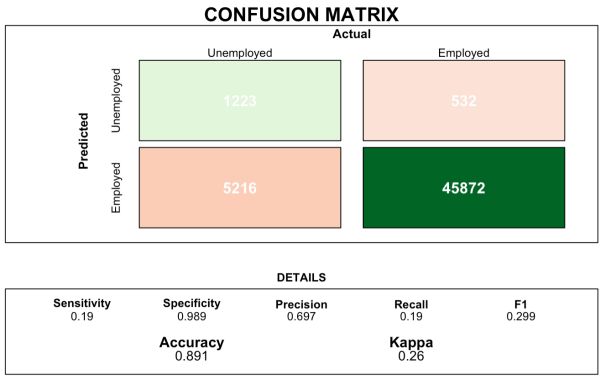
2. Evaluation Metrics:
To assess the performance of the models, I used several evaluation metrics commonly used for binary classification problems in predictive modeling:

* Sensitivity (true positive rate): correct identification of the unemployed clients.
* Specificity (true negative rate): correct identification of the employed clients.
* Accuracy: the overall accuracy of the models in correctly predicting employment status, i.e. how often a classifier makers a correct prediction.
* Precision: the proportion of correctly predicted unemployed out of all predictions that the client would be unemployed in 6 months.
* Recall: the proportion of correctly predicted unemployed out of all actual unemployed individuals.
* F1 Score: a harmonic mean of precision and recall, providing a balanced measure of the models' performance.
* Area Under the ROC Curve (AUC-ROC): a metric to assess the models' ability to discriminate between employed and unemployed individuals.

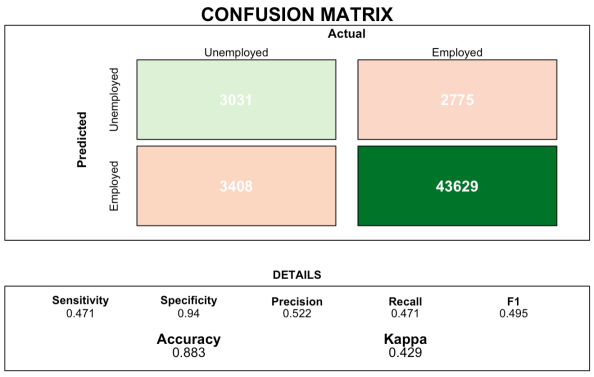
3. Evaluation Results:

Evaluation of both models with the aforementioned metrics produces the following results summarized by confusion matrices. Model 2 shows a higher sensitivity, capturing a larger proportion of the actual unemployed individuals, while maintaining a high specificity. It achieves a higher precision, meaning a higher proportion of individuals predicted as unemployed are actually unemployed. Additionally, Model 2 demonstrates a higher F1 score, indicating a better balance

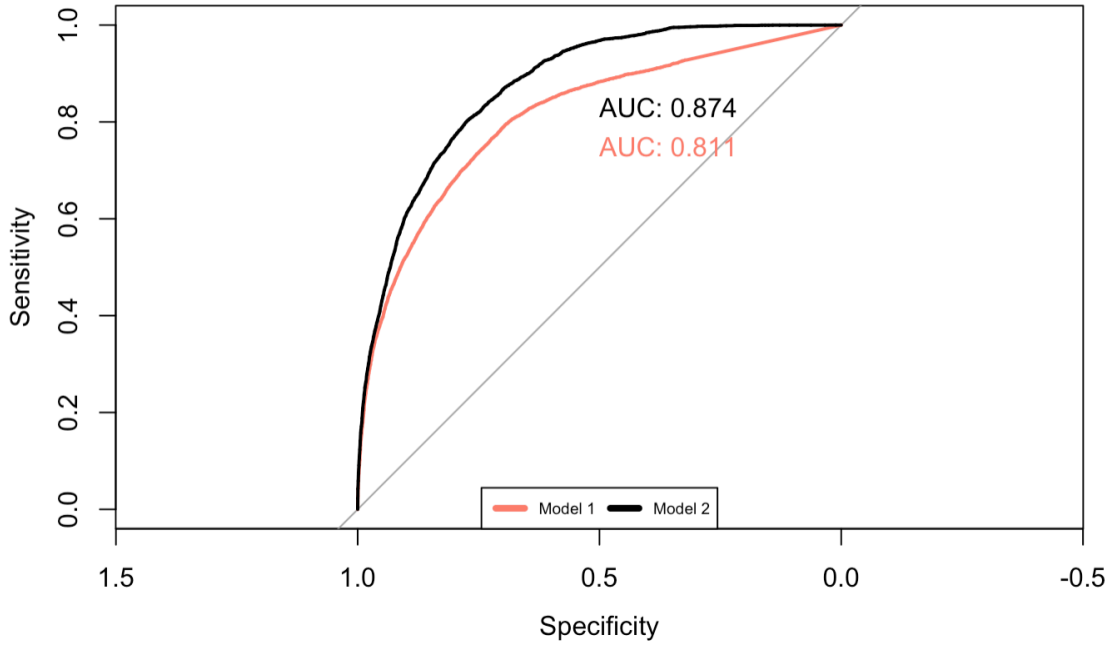
between precision and recall. The Kappa coefficient also suggests a moderate agreement beyond chance for Model 2.



model 1



model 2



AUC-ROC for both models

Moreover, Model 2 has a higher AUC, suggesting that it has better predictive ability and can distinguish between employed and unemployed individuals more accurately compared to Model 1. Considering these evaluation metrics, it is recommended to deploy Model 2 as it shows better performance in predicting unemployment status.

Additionally, I examined the demographic disparities in the model predictions to assess potential bias. The visualizations provided below demonstrate the model predictions across different demographic groups, specifically focusing on race and gender. I have not found any age discrimination bias. Both models make more errors in the age group of 30s, but this is possibly due to models sensitivity to unbalanced data (the EDD dataset has more observation for clients in their

30s). See the incorrect prediction rates, the false positive and false negative rates, for different age groups (30s, 40s, 50s) below.

age_cat <dbl>	FNR <dbl>	FPR <dbl>
30	0.011251657	0.10567241
40	0.010217006	0.09254412
50	0.007408539	0.08943710

model 1

age_cat <dbl>	FNR <dbl>	FPR <dbl>
30	0.05543412	0.06801161
40	0.04829857	0.06206198
50	0.05010311	0.05919193

model 2

Similarly, models seem to lack any gender bias.

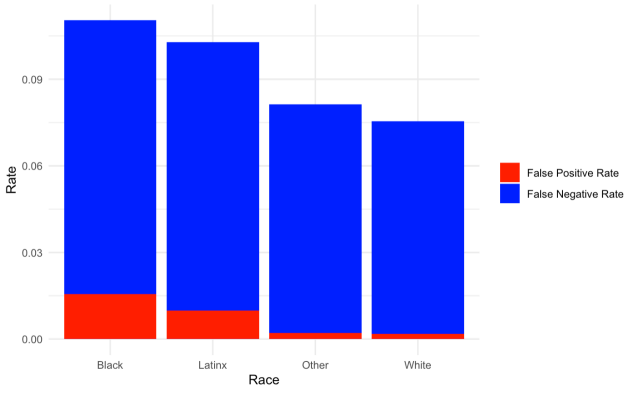
sex <chr>	FNR <dbl>	FPR <dbl>
F	0.009146814	0.09689422
M	0.010599546	0.09975517

model 1

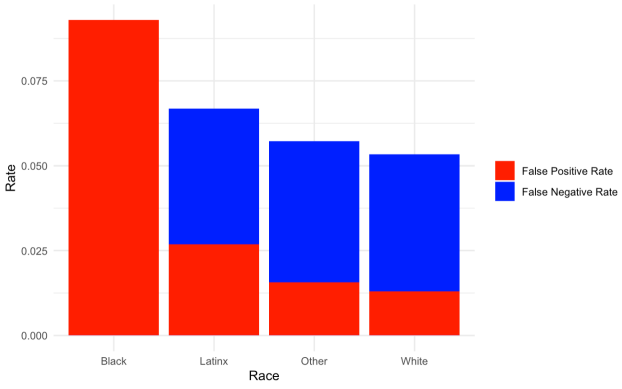
sex <chr>	FNR <dbl>	FPR <dbl>
F	0.009146814	0.09689422
M	0.010599546	0.09975517

model 2

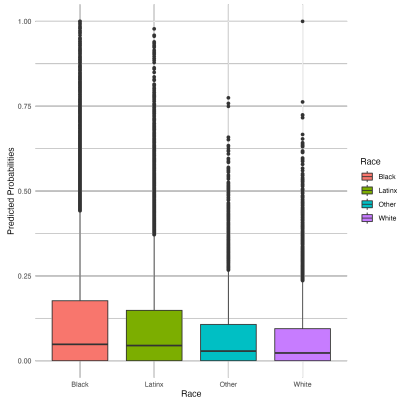
However, Model 2 incorporates greater racial bias with higher false positive rates, predicting that the client is still unemployed while they got actually employed. Moreover, the overall predicted probabilities for remaining unemployed in 6 months are greater in Model 2 for black and latin groups.



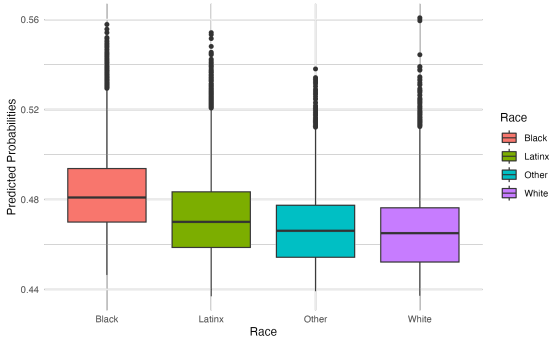
model 1



model 2



model 1



model 2

4. Recommendations:

Based on the evaluation results and considering EDD's goal of non-discrimination and avoiding bias against marginalized groups, I propose to utilize Model 1. Model 2 demonstrates superior performance in terms of sensitivity, recall, F score, and AUC-ROC compared to Model 1. However, Model 2 incorporates a substantial racial bias against black and latin groups. Moreover, accuracy and specificity rates do not differ significantly across the models. The recommendation is based on the assumption that assessing the Model 2 algorithmic to mitigate its racial bias and retraining Model 2 would inflict additional costs beyond the EDD's budget. It is recommended to work closely with experts in fairness and bias mitigation to address any potential issues and develop appropriate strategies. Overall, regular monitoring and ongoing evaluation of the deployed model should be implemented to ensure fairness and performance are maintained over time.