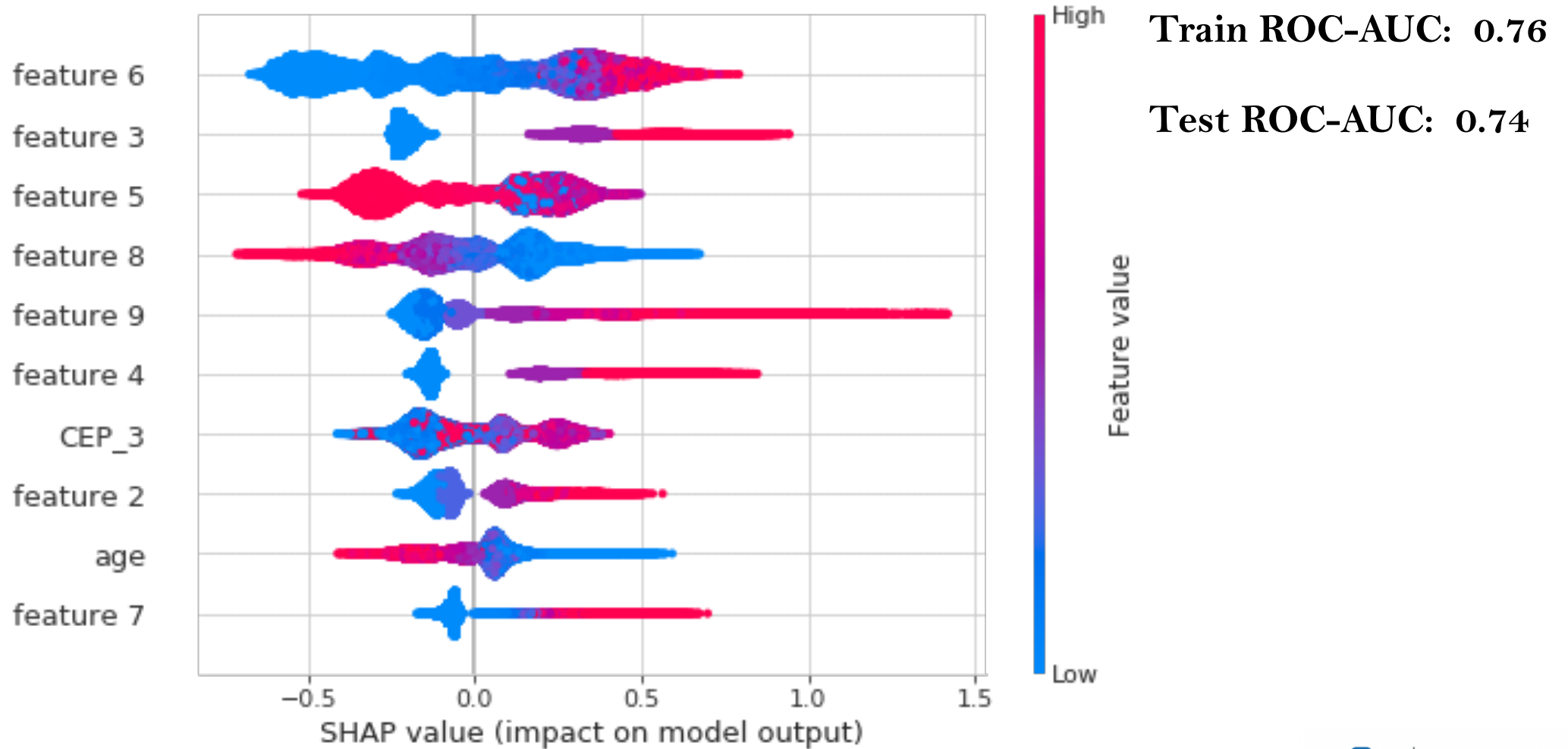# LOCATION INFORMATION AND RACIAL BIASES IN CREDIT SCORING
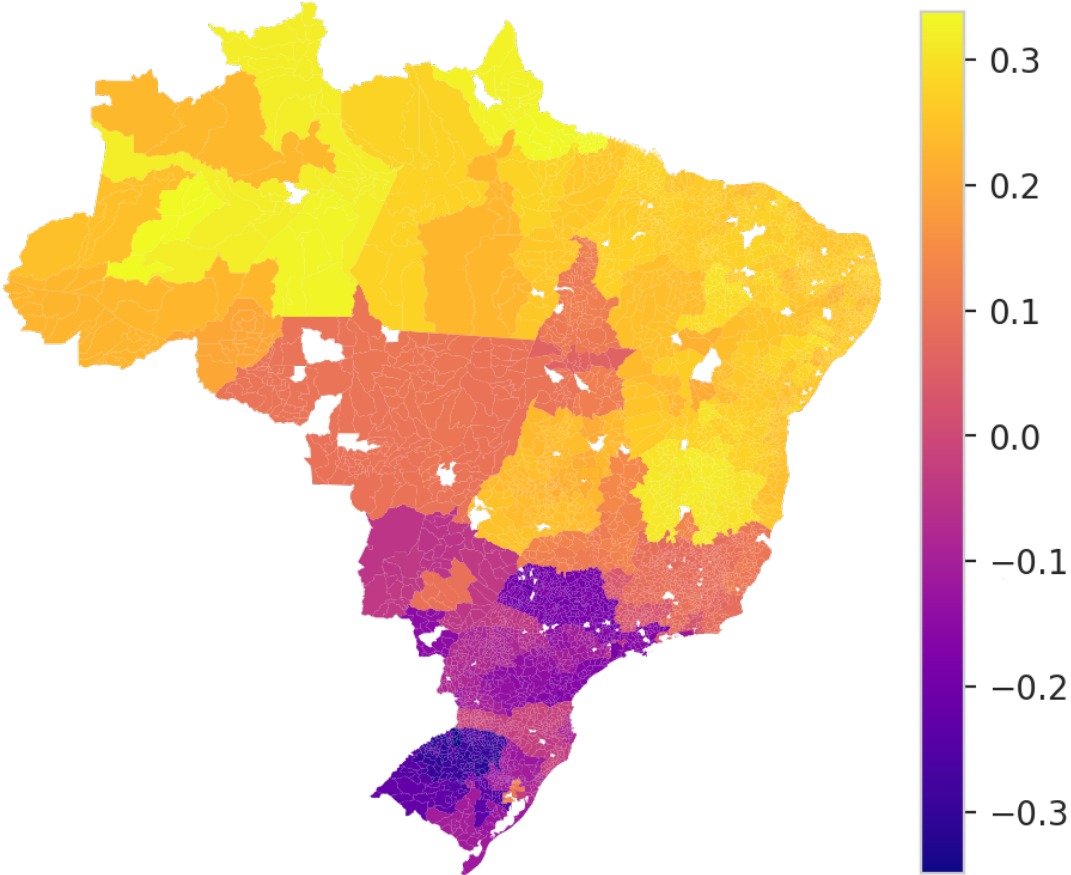
Ramon Vilarino

Experian DataLab LatAm

DataLab

SHAP Values Summary for Tree Gradient Boost Model with 250 tree estimators of maximal depth of 3
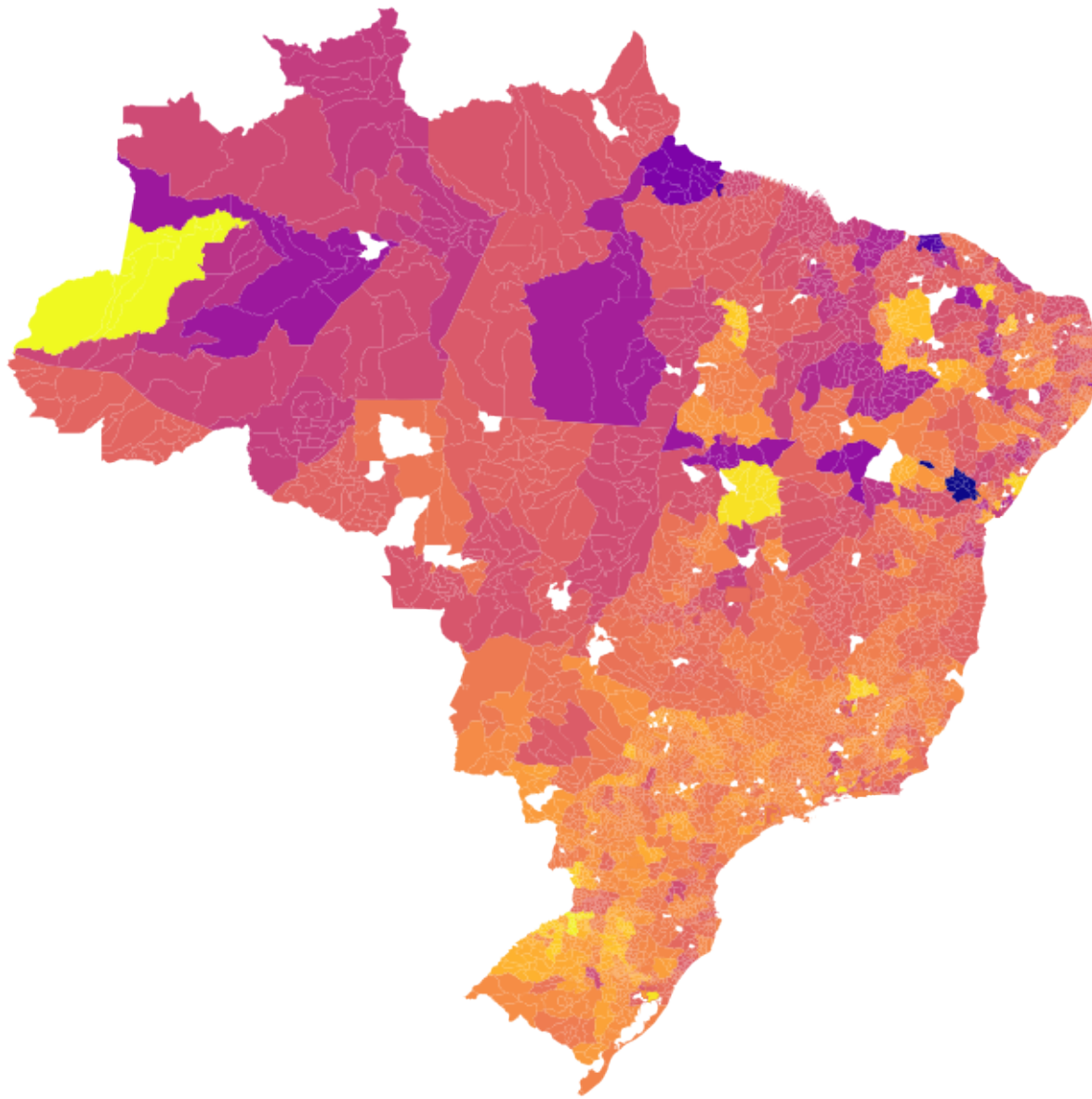
Correlation Coefficient: 82%
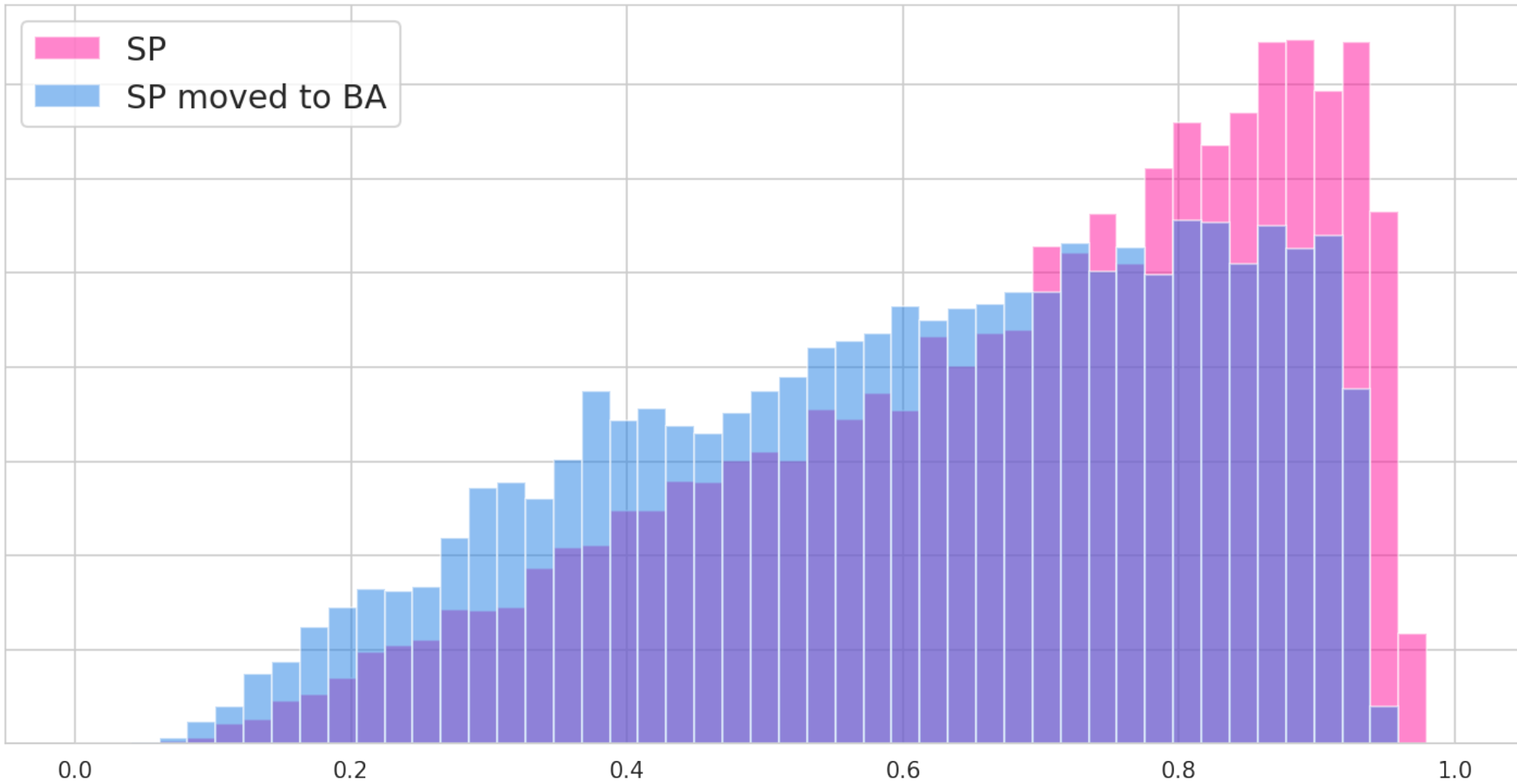

**Model's 3-CEP
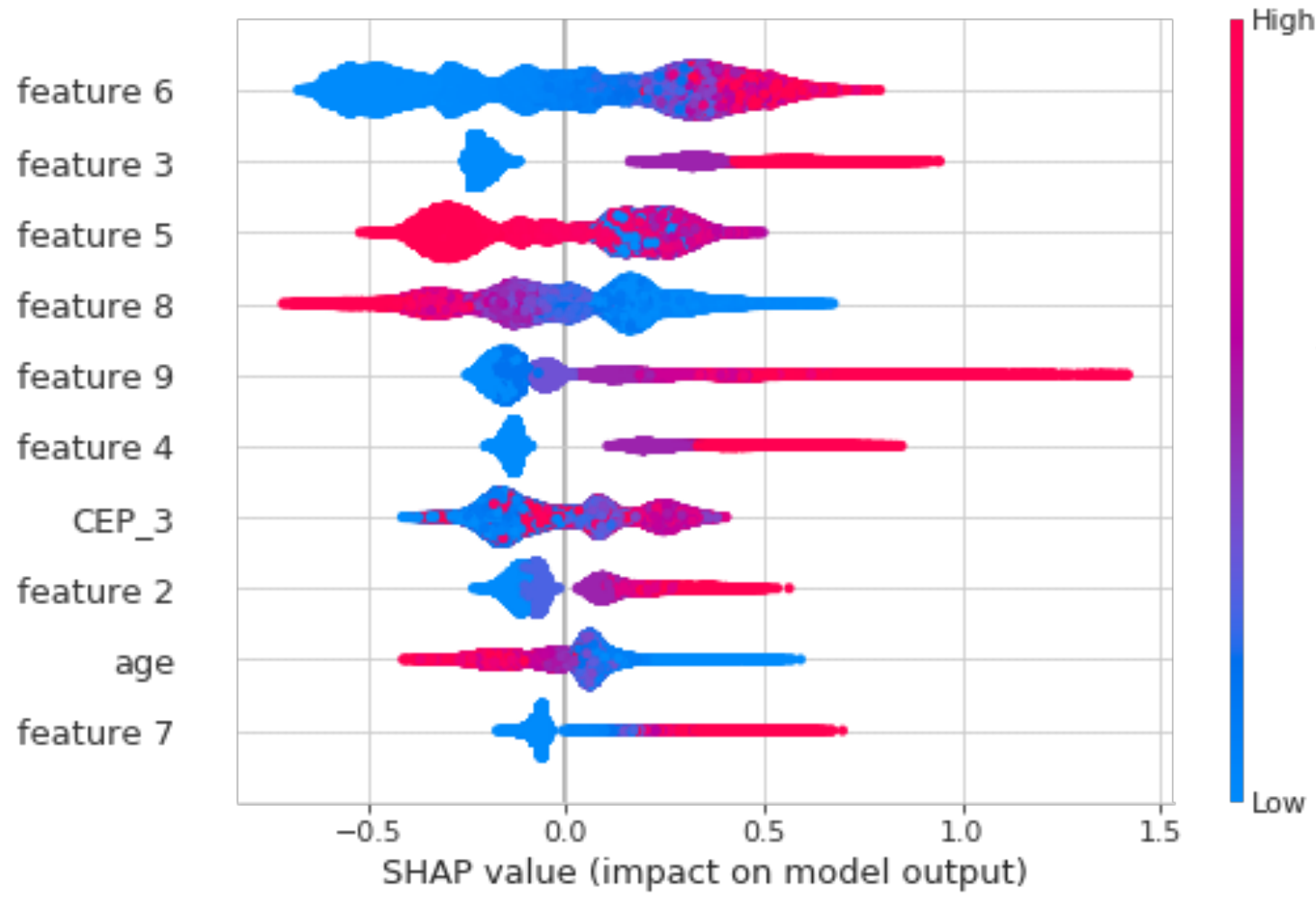SHAP Values**

**Not-White People
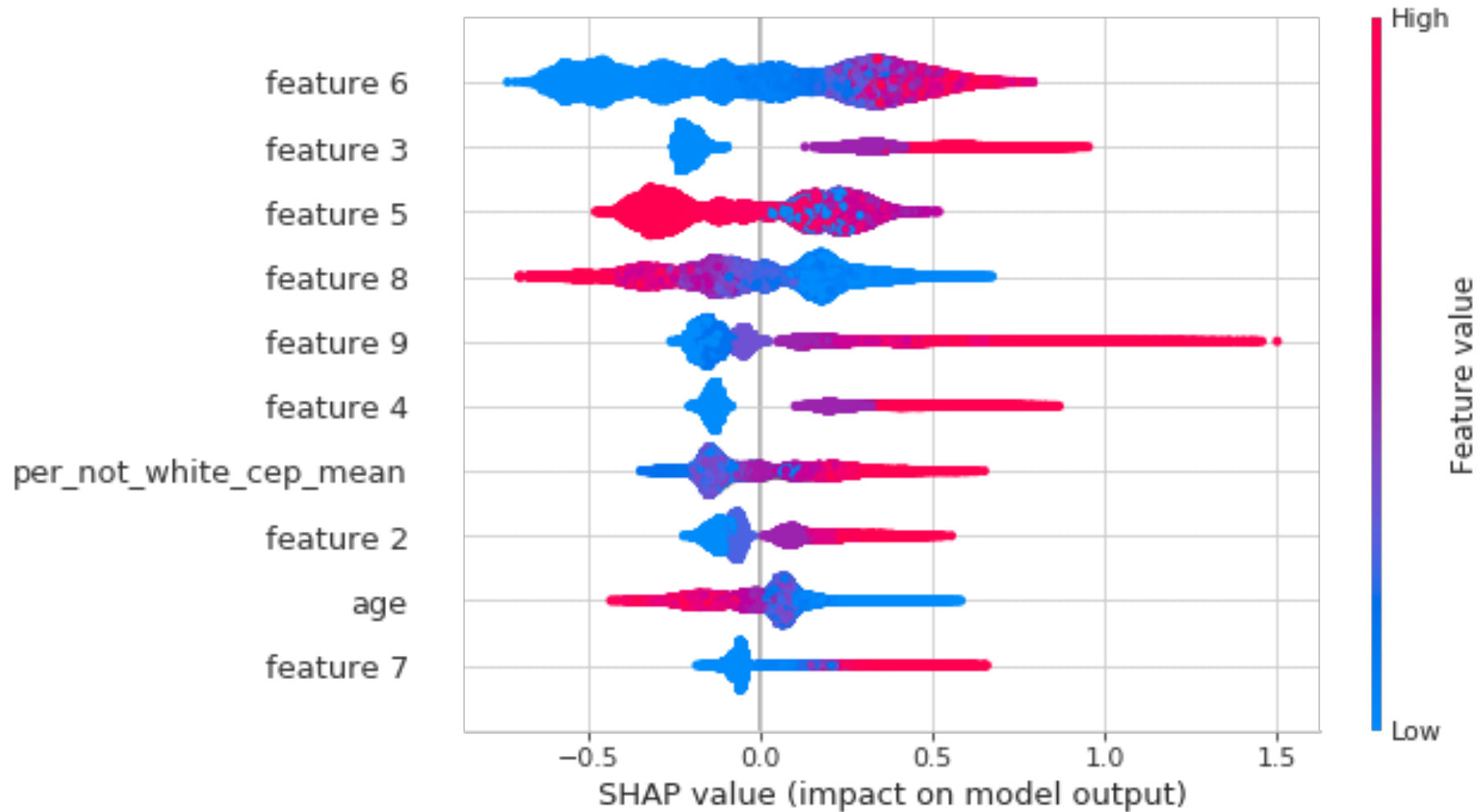Proportion**

Score Distribution

In 99.9% of the cases the score strictly decreased

Train ROC-AUC: 0.76

Test ROC-AUC: 0.74
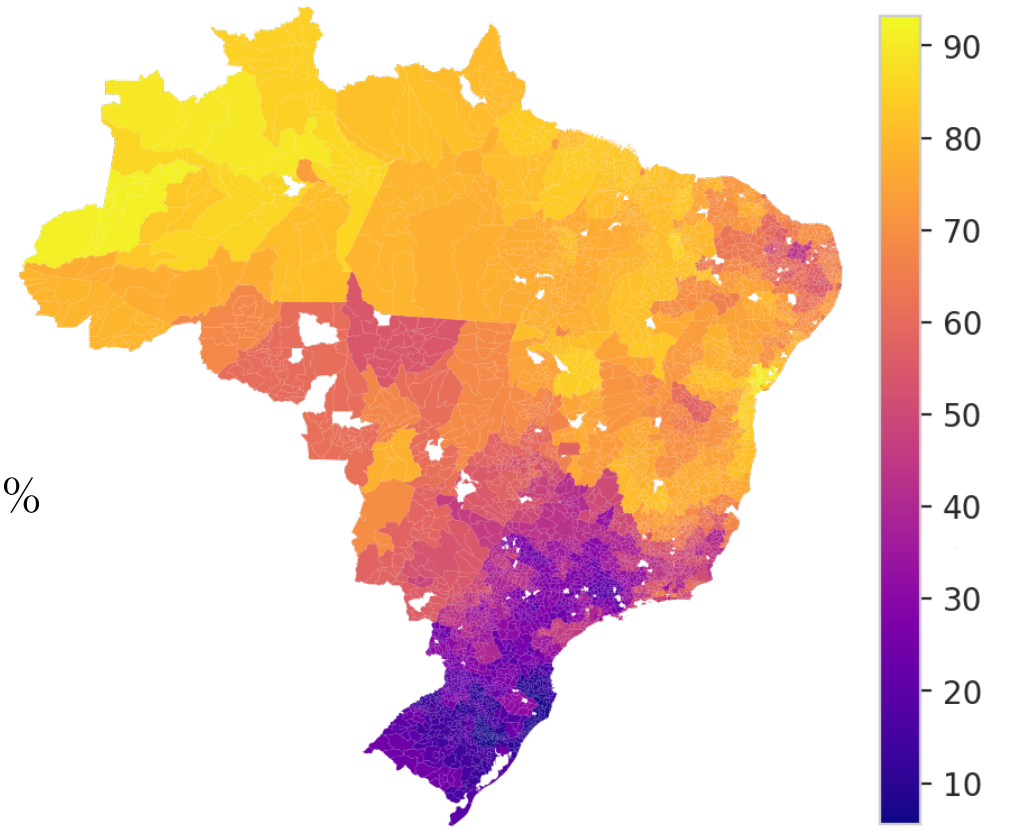
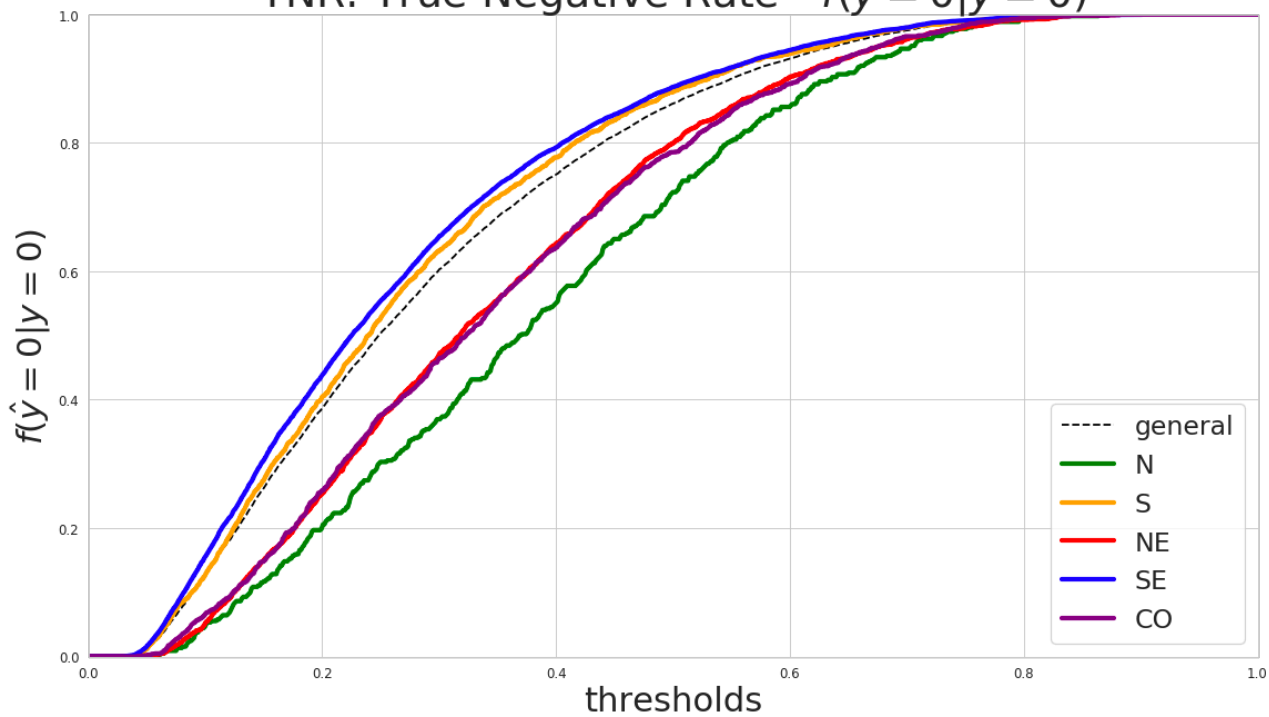Train ROC-AUC: 0.76

Test ROC-AUC: 0.74

Proportion of self-declared not-whites

North: 76.4 %
Northeast: 71.2 %
Central-West: 58.5 %
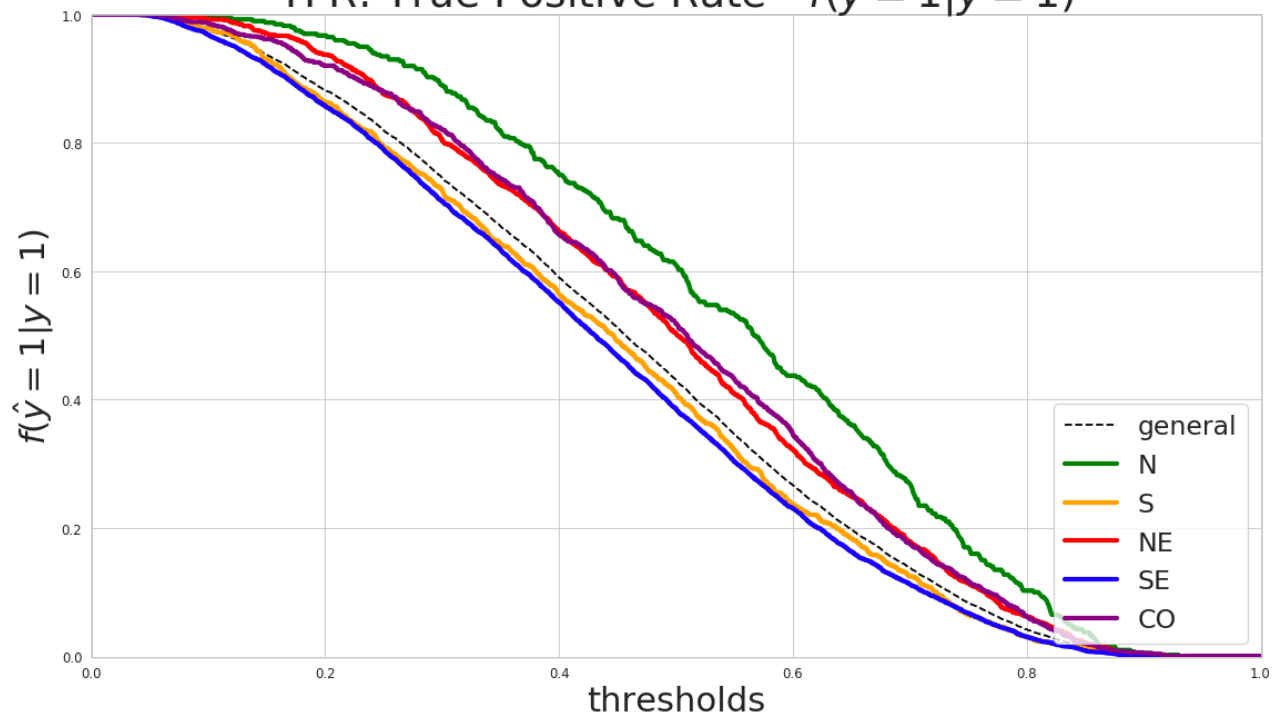**Southeast:** 43.3 %
South: 21.5 %
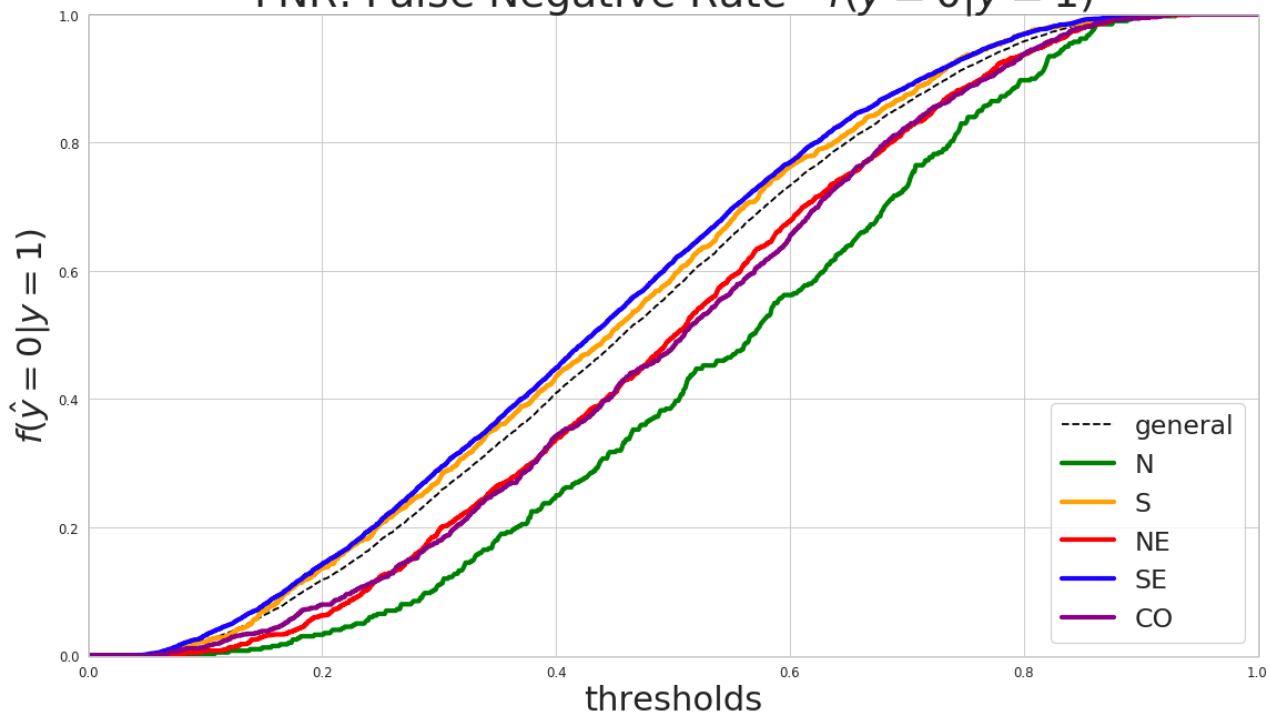
Not-White People Proportion
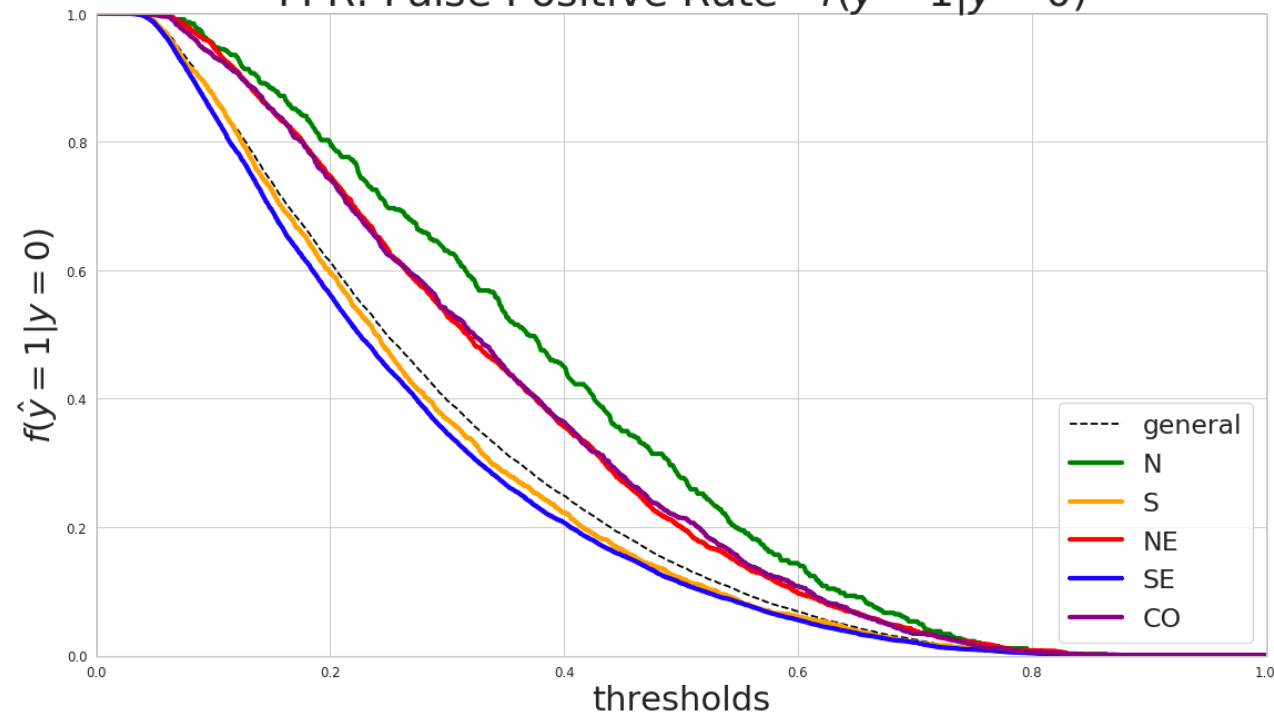
TNR: True Negative Rate - $f(\hat{y} = 0 | y = 0)$

TPR: True Positive Rate - $f(\hat{y} = 1 | y = 1)$

general
N
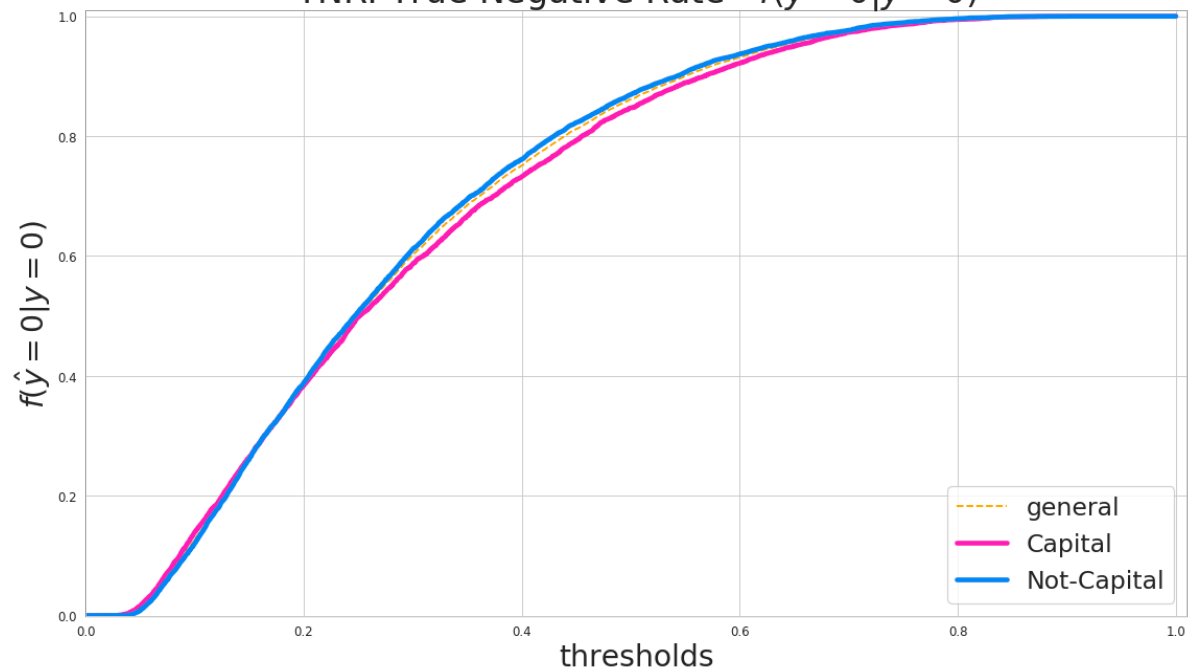S
NE
SE
CO

FNR: False Negative Rate - $f(\hat{y} = 0 | y = 1)$

$f(\hat{y} = 0 | y = 1)$

thresholds

- - - general
— N
— S
— NE
— SE
— CO

FPR: False Positive Rate - $f(\hat{y} = 1 | y = 0)$

$f(\hat{y} = 1 | y = 0)$

thresholds

- - - general
— N
— S
— NE
— SE
— CO

DataLab

TNR: True Negative Rate - $f(\hat{y}=0|y=0)$

FNR: False Negative Rate - $f(\hat{y}=0|y=1)$
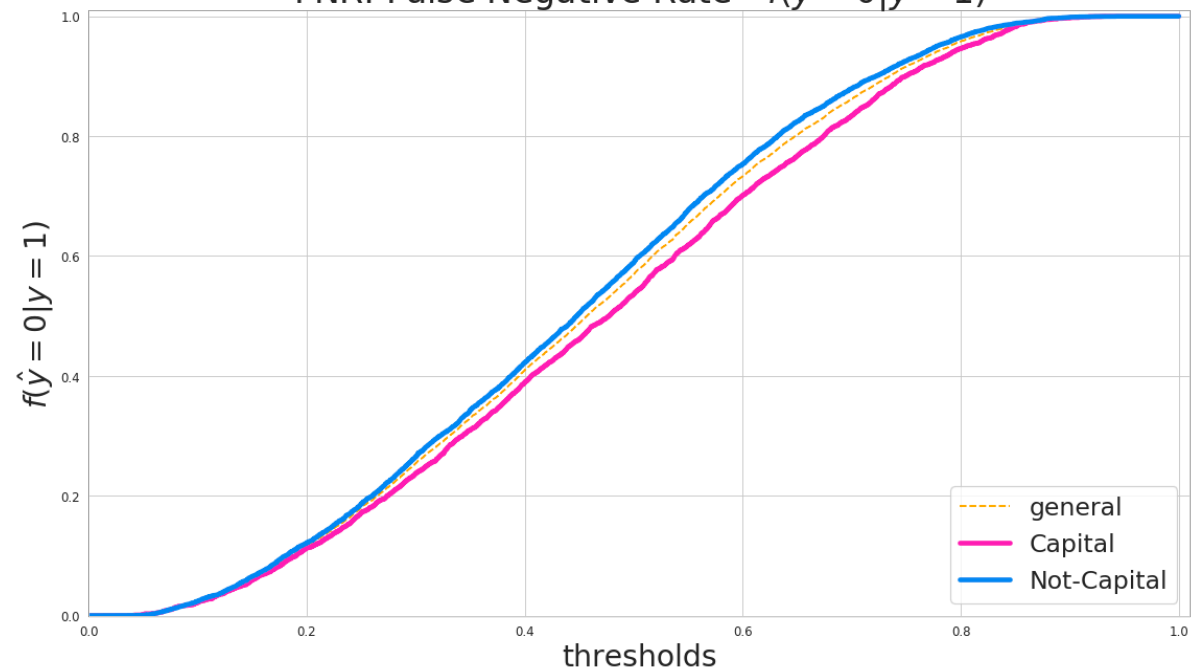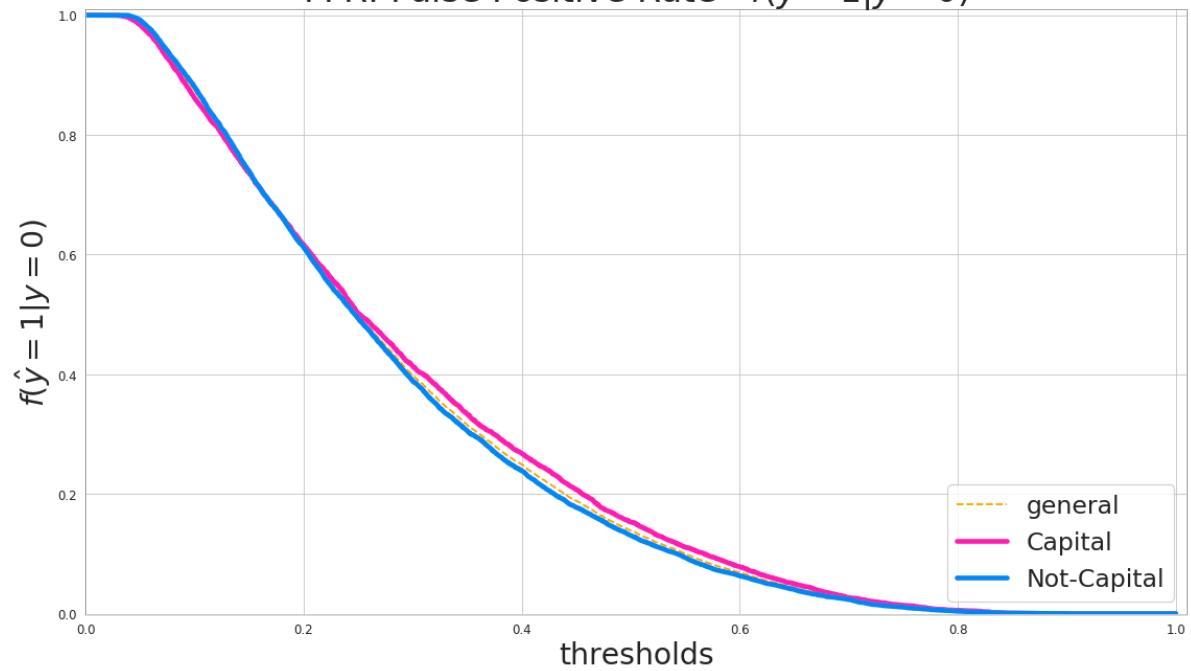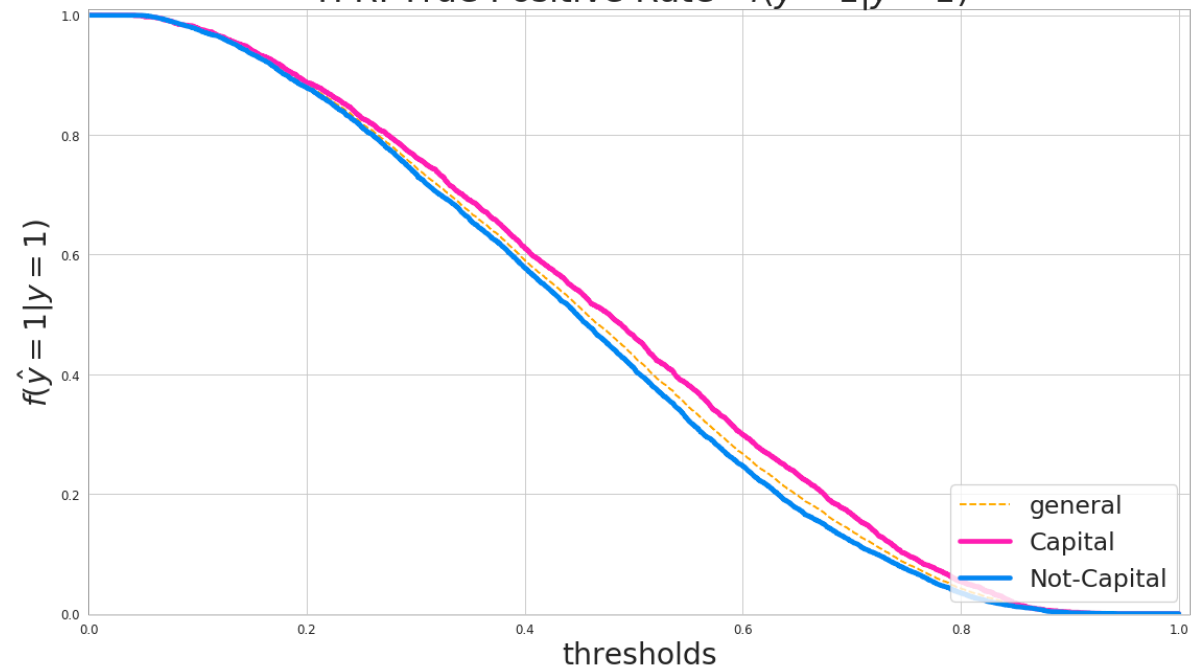
FPR: False Positive Rate - $f(\hat{y}=1|y=0)$

TPR: True Positive Rate - $f(\hat{y}=1|y=1)$

Thank You!

@datalab.experian

linkedin.com/showcase/serasa-experian-datalab/

medium.com/datalab-log

@vilarinoramon

linkedin.com/in/vilarinoramon

ramon.vilarino.santos@usp.br