Overview

o <u>Model</u> = Logistic Regression with l2 regularization and class weighted loss function.

o <u>Features</u> = TF-IDF 5000 features with unigrams, from preprocessed tweets

o <u>Metrics</u> = .k-fold cross validation average for every performance metrics.

o The same model structure for both Age and Race prediction.

o For Race the "Other" class includes only "Latino" and "Hispanic" (Multi-racial class hasn't been considered).

Race Prediction

Confusion matrix	Black	Other	White
Black	**68%**	13%	18%
Other	6%	**58%**	36%
White	13%	8%	**79%**

Overall accuracy = 71%

Performances	Black	Other	White
Precision	68%	66%	76%
Recall	68%	58%	79%
F1	68%	62%	77%

Age Prediction

Confusion matrix	< 21	>= 21
< 21	**63%**	37%
>= 21	22%	**78%**

Overall accuracy = 70%

Performances	< 21	>= 21
Precision	65%	75%
Recall	58%	79%
F1	62%	77%