## Part 1

My model’s performance: ~0.8613 accuracy

|  |  |
| --- | --- |
| Better Performing | Worse Performing |
| |  |  | | --- | --- | | Gradient Boosting Classifier | 0.782-0.880 | | MLP Classifier | 0.840-0.871 | | Random Forest Classifier | 0.825-0.873 | | SVC | 0.664-0.897 | | |  |  | | --- | --- | | Decision Tree Classifer | 0.777-0.798 | | ExtraTree Classifier | 0.737-0.775 | | GaussianNB | 0.511 | | KNeighbours Classifer | 0.838-0.854 | | LinearSVC | 0.484-0.836 | | Logistic Regression | 0.836-0.842 | | Passive Aggressive Classifier | 0.773-0.776 | | SGD Classifier | 0.813-0.819 | |

## Part 2

1. Before implementing the bias mitigation strategy, the model was fairly accurate (80-84%)
2. The African-American subgroup had relatively high likelihood (~25%) of being predicted to reoffend. Other metrics: Asian and Native subgroups are extremely unlikely to reoffend (close to 0%), Caucasian and “Other” subgroups are both unlikely (~5%), Hispanic subgroup is very unlikely (~4%).
3. The African-American subgroup was unreasonably likely to be predicted to reoffend. The African-American subgroup had a higher likelihood of being classified as a false positive than all other subgroups: ~6% of all positives were falsely assigned, while for all other classes ~0% of all positives were falsely assigned.
4. With the class-balanced training, the accuracy is relatively lower (unbalanced: ~92%, balanced 82%) but the bias is significantly lower. The model is overall less likely to classify samples as positive (risk of reoffense). However, in terms of unreasonably high likelihoods of being classified positively, the bias has been sufficiently mitigated: nearly 0% of samples were classified incorrectly (1/729).