|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | Model | AUC | Accuracy | Recall\_Pos. | Recall\_Neg. | Precision\_Pos. | Precision\_Neg. |
| Original |  |  |  |  |  |  |  |
|  | Logistic Reg. | 0.802 | 0.860 | 0.250 | 0.965 | 0.551 | 0.882 |
|  | KNN | 0.742 | 0.853 | 0.062 | 0.989 | 0.500 | 0.859 |
|  | Random Forest | 0.795 | 0.871 | 0.250 | 0.978 | 0.667 | 0.883 |
| BL\_SMOTE |  |  |  |  |  |  |  |
|  | Logistic Reg. | 0.800 | 0.743 | 0.688 | 0.752 | 0.324 | 0.934 |
|  | KNN | 0.750 | 0.753 | 0.578 | 0.795 | 0.327 | 0.916 |
|  | Random Forest | 0.797 | 0.855 | 0.391 | 0.935 | 0.510 | 0.899 |
| SMOTE |  |  |  |  |  |  |  |
|  | Logistic Reg. | 0.798 | 0.752 | 0.688 | 0.763 | 0.331 | 0.934 |
|  | KNN | 0.685 | 0.690 | 0.578 | 0.714 | 0.259 | 0.908 |
|  | Random Forest | 0.782 | 0.850 | 0.359 | 0.935 | 0.489 | 0.894 |

**Recall\_Pos:** Percentage of employees actually intending to quit correctly “detected”

* Logistic Regression fit to the synthetic datasets best
* All models fit to synthetic datasets \*far\* outperformed the models fit to the original datasets on this metric

**Recall\_Neg:** Percentage of employees actually not intending to quit correctly “detected”

* KNN fit to the original dataset best
* All models fit to original datasets \*far\* outperformed the models fit to the synthetic datasets

**Precision\_Pos:** Percentage of positive predictions that were accurate (predicted as intending to quit)

* Random forest fit to the original dataset best
* All models fit to original datasets \*far\* outperformed the models fit to the synthetic datasets

**Precision\_Neg:** Percentage of negative predictions that were accurate (predicted as not intending to quit)

* Logistic Regression fit to the synthetic datasets was best