

Логистическая регрессия

	precision	recall	f1-score	support
0	0.86	0.33	0.48	1847
1	0.31	0.85	0.45	645
accuracy			0.46	2492
macro avg	0.59	0.59	0.46	2492
weighted avg	0.72	0.46	0.47	2492


```
[[ 608 1239]
 [  95  550]]
```

	precision	recall	f1-score	support
0	0.84	0.62	0.72	1847
1	0.38	0.66	0.48	645
accuracy			0.63	2492
macro avg	0.61	0.64	0.60	2492
weighted avg	0.72	0.63	0.66	2492

```
[[1152  695]
 [ 218  427]]
```

На рисунке слева представлены результаты логистической регрессии на нормализованных данных, а справа результаты логистической регрессии на увеличенном датасете с нормализованными данными. Как можно заметить, первая модель в основном записывает уходящих клиентов, когда они не собираются уходить и при этом у нее мало “промахов” по действительно уходящим людям. Во втором алгоритме также все неплохо, но уже больше упущенных действительно уходящих людей. Что из этого лучше, я считаю нужно выбирать бизнесу, потому что как по мне первая модель лучше, ведь мы сможем удержать больше клиентов.

Случайный лес



	precision	recall	f1-score	support
0	0.79	0.94	0.86	1847
1	0.63	0.30	0.41	645
accuracy			0.77	2492
macro avg	0.71	0.62	0.64	2492
weighted avg	0.75	0.77	0.74	2492

```
[[1734 113]
 [ 449 196]]
```

На данном рисунке видно результаты случайного леса. Как можно заметить, оно показывает в среднем результаты лучше обеих предыдущих моделей. Но опять же, у нее высокий recall для 0 класса и низкий для 1 класса. По матрице мы можем видеть, что мы упускаем еще больше людей, чем в двух предыдущих моделях. Опять таки, результаты можно интерпретировать по разному, в зависимости от того, что для бизнеса лучше. Но лично я считаю, что легче удержать старых пользователей, чем привлекать новых, поэтому лично для меня первая модель лучше остальных.