**Описание проекта:**

Из «Бета-Банка» стали уходить клиенты. Каждый месяц. Немного, но заметно. Банковский маркетологи посчитали: сохранять текущих клиентов дешевле, чем привлекать новых.

Нужно спрогнозировать уйдет клиент из банка в ближайшее время или нет. В проекте предоставлены исторические данные о поведении клиентов и расторжении договоров с банком.

**Постановка задачи:**

Необходимо:

- построить модель с предельно большим значением F1-меры(нужно довести метрику до 0.59).

- проверить F1-меру на тестовой выборке самостоятельно.

- дополнительно измерять AUC-ROC и сравнивать её значение с F1-мерой.

**Результаты работы:**

Для прогноза мы использовали модели машинного обучения такие, как дерево решений, случайный лес и логистическая регрессия. Сначала модели были построены без учета дисбаланса классов. F1-мера случайного леса показала результат 0.59, AUC-ROC – 0.83; дерево решений – 0.55, AUC-ROC – 0.82; логистическая регрессия – 0.32, AUC-ROC – 0.78.

Затем была проведена работа с дисбалансом классов методами upsampling и downsampling. После upsamplingа показатель F1-меры у случайного леса вырос до 0.6, AUC-ROC – 0.83; у дерева решений – до 0.56, AUC-ROC – 0.82; у логистической регрессии – до 0.51, AUC-ROC – 0.78. После downsamplingа показатель F1-меры у случайного леса 0.6, AUC-ROC – 0.83; у дерева решений – 0.54, AUC-ROC – 0.8; у логистической регрессии – 0.51, AUC-ROC –0.78.

Увеличение выборки у модели случайного леса показало наилучший результат F1-меры. Именно ее мы и применили на тестовой выборке, что показало результат F1-меры 0.59, AUC-ROC – 0.84.