**Постановка задачи и описание проекта**

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

Нужно спрогнозировать, уйдёт клиент из банка в ближайшее время или нет. Вам предоставлены исторические данные о поведении клиентов и расторжении договоров с банком.

Постройте модель с предельно большим значением F1-меры. Чтобы сдать проект успешно, нужно довести метрику до 0.59. Проверьте F1-меру на тестовой выборке самостоятельно.

Дополнительно измеряйте AUC-ROC, сравнивайте её значение с F1-мерой.

(источник данных: <https://www.kaggle.com/barelydedicated/bank-customer-churn-modeling>)

**Решение поставленной задачи**

Мы применили несколько методов борьбы с дисбалансом классов. Метод взвешивания классов дал следующие результаты:

* логистическая регрессия: F1-мера = 0.488, AUC-ROC = 0.763
* решающее дерево: F1-мера = 0.596 AUC-ROC = 0.831, при max\_depth = 5
* случайный лес: максимальная F1-мера: 0.59, при n\_estimators = 8 (AUC-ROC=0.806), максимальная метрика AUC-ROC = 0.838, при n\_estimators = 84(F1-мера = 0.573)

Метод изменения порога классификации дал следующие результаты:

* логистическая регрессия: максимальная F1-мера = 0.502, AUC-ROC = 0.758 при пороге = 0.26
* решающее дерево: F1-мера = 0.6082, AUC-ROC = 0.831 при пороге 0.58
* случайный лес: при изменении порога показатели не улучшились и остались на уровне метода взвешивания классов.

В качестве лучшей модели была выбрана модель решающего дерева при значении гиперпараметра max\_depth = 5, указании атрибута class\_weight='balanced' и изменении порога, равным 0.58.

На валидационной выборке модель следующие результаты:

* F1-мера: 0.608
* AUC-ROC: 0.831

На тестовой выборке модель следующие результаты:

* F1-мера: 0.60 (уменьшилась на 0.008)
* AUC-ROC: 0.836 (увеличилась на 0.005)

Проверку на адекватность модель выдержала. Лучшие (F1-мера:0.29, AUC-ROC: 0.517) показатели фиктивного классификатора значительно ниже.

Требование касательно значения F1-меры (нужно довести метрику до 0.59) выполнено - **значение F1-меры нашей модели равно 0.6 на тестовой выборке.**