**Задача:**

Разработать и оценить классификатор, прогнозирующий вероятность прекращения сотрудничества клиента с банком. Основное требование — добиться F1-меры не ниже 0.59 на тестовом наборе. Для дополнительного анализа качества использовать метрику ROC AUC.

В распоряжении имеются данные с информацией о клиентах финансовой компании, включающие:

* *Number —* индекс строки в данных
* *CustomerId* — уникальный идентификатор клиента
* *Surname —* фамилия
* *CreditScore —* кредитный рейтинг
* *Geography —* страна проживания
* *Gender —* пол
* *Age —* возраст
* *Tenure —* сколько лет человек является клиентом банка
* *Balance —* баланс на счёте
* *NumOfProducts —* количество продуктов банка, используемых клиентом
* *HasCrCard —* наличие кредитной карты
* *IsActiveMember —* активность клиента
* *EstimatedSalary —* предполагаемая зарплата
* *Exited* — факт ухода клиента

## 

## 

## 

## 

## 

## 

## 

## **Этапы выполнения**

1. Работа с данными
   * Загрузить набор, изучить формат и виды переменных.
   * Провести очистку — убрать или скорректировать дубликаты, аномалии и пропуски.
   * Проанализировать распределение целевой переменной для выявления дисбаланса.
2. Обучение эталонных моделей
   * Выполнить разбиение исходных данных на обучающий и контрольный наборы.
   * Обучить базовую модель без учета дисбаланса, например, логистическую регрессию либо дерево решений.
   * Рассчитать базовые метрики качества: F1-меру и ROC AUC, проанализировать результаты.
3. Оптимизация с учетом дисбаланса
   * Применить техники балансировки классов, такие как взвешивание, изменения в обучающем наборе (oversampling/undersampling).
   * Обучить несколько алгоритмов (2 и более).
   * Найти и зафиксировать модель с максимальным значением F1-меры.
4. Итоговое тестирование и сравнение
   * Оцените финальную модель на независимом тестовом наборе.
   * Сравните F1-меру с ROC AUC, обсудите взаимосвязь и практическую значимость.
   * Сделайте рекомендации по использованию модели в бизнес-процессах.

## **Оформление**

Включить следующие элементы:

* Подробное описание этапов очистки данных
* Использованные подходы к работе с дисбалансом
* Таблицы и графики, иллюстрирующие результаты обучения и сравнения метрик
* Краткий анализ и рекомендации на основе полученных показателей