**Лабораторная работа 4**

Вам предоставлен набор данных, который содержит информацию о клиентах банка:

* client\_id - ID клиента (неотрицательное целое число)
* Age - возраст (вещественный)
* Income - месячный доход (вещественный)
* BalanceToCreditLimit - отношение баланса на кредитной карте к лимту по кредиту (вещественный)
* DIR - Debt-to-income Ratio (вещественный)
* NumLoans - число заемов и кредитных линий
* NumRealEstateLoans - число ипотек и заемов, связанных с недвижимостью (натуральное число)
* NumDependents - число членов семьи, которых содержит клиент, исключая самого клиента (натуральное число)
* Num30-59Delinquencies - число просрочек выплат по кредиту от 30 до 59 дней (натуральное число)
* Num60-89Delinquencies - число просрочек выплат по кредиту от 60 до 89 дней (натуральное число)
* Delinquent90 - Таргет, были ли просрочки выплат по кредиту более 90 дней (бинарный)

**Задание:**

1. Анализ и предобработка.
   1. Проанализировать данные (EDA).
   2. Предобработать данные.
   3. Скалировать/нормализовать данные.
   4. Подготовить данные для обучения моделей.
2. Дерево решений.
   1. Обучить дерево решений, визуализировать решающее дерево
   2. Рассмотреть, как изменение глубины дерева влияет на качество модели. Обучите решающее дерево с различной глубиной (от 1 до 11), оцените его точность и определите, когда начинается переобучение и какая глубина дерева является оптимальной.
   3. Подберите оптимальные параметры дерева с помощью GridSearhCV.
3. Ансамблевые методы.
   1. Выберите любую классическую модель (не ансамблевую) и обучите ансамбль с помощью бэггинга (BaggingClassifier). Подберите оптимальные параметры как бэггинга так и модели с помощью GridSearhCV.
   2. Выберите любые классические модели (не ансамблевые) и обучите ансамбль с помощью стекинга (StackingClassifier). Подберите оптимальные параметры как стекинга так и модели с помощью GridSearhCV.
   3. Обучите cлучайный лес (RandomForestClassifier), настройте гиперпараметры случайного леса с помощью GridSearhCV.
   4. Инициализируете параметры и обучите классификатор XGBoost (обратите внимание, что XGBoost работает только с числовыми признаками). Вычислите F test (не F score!) для оценки важности признаков в XGBoost.