

# Применение алгоритмов Decision Tree и Random Forest

## Цель:

Освоить процесс подготовки данных, применения алгоритмов машинного обучения (Decision Tree и Random Forest), визуализации моделей и оценки их эффективности.

---

## Задание:

### 1. Выбор датасета:

- Выберите один датасет из открытых источников, таких как Kaggle, [UCI Machine Learning Repository](#) или [OpenML](#).
- Убедитесь, что датасет содержит как минимум одну целевую переменную для задачи классификации.

### 2. Предварительная обработка данных:

- **Очистка данных:**
  - Обработайте пропущенные значения (удаление, заполнение средними/медианными значениями и т.д.).
  - Удалите или исправьте выбросы, если это необходимо.
  - Преобразуйте категориальные признаки в числовые (например, с помощью one-hot encoding).
- **Нормализация/Стандартизация:**
  - Примените методы нормализации или стандартизации к числовым признакам для улучшения работы алгоритмов.

### 3. Разделение данных:

- Разделите датасет на обучающую и тестовую выборки (например, 80% на обучение и 20% на тестирование).

### 4. Применение алгоритмов:

- **Decision Tree:**
  - Обучите модель Decision Tree на обучающей выборке.
- **Random Forest:**
  - Обучите модель Random Forest на той же обучающей выборке.

### 5. Визуализация модели Decision Tree:

- Постройте графическое представление дерева решений.
- Объясните структуру дерева: количество уровней, признаки, использованные для разбиения, и т.д.

### 6. Анализ моделей:

- Сравните результаты моделей Decision Tree и Random Forest.
- Обсудите преимущества и недостатки каждого алгоритма на вашем датасете.
- Проанализируйте важность признаков в модели Random Forest.

### 7. Оценка качества моделей:

- Рассчитайте следующие метрики для обеих моделей:
  - **Для задачи классификации:**
    - Точность (Accuracy)
    - Полнота (Recall)

- Точность (Precision)
  - F1-Score
  - Матрица ошибок (Confusion Matrix)
- Представьте результаты в виде таблицы и визуализаций (графики, диаграммы).