

## **ЗАДАНИЕ**

### **Основные модели машинного обучения**

**Тема практического занятия:** Обучение моделей машинного обучения.

**Постановка задачи:** разработка моделей и определение метрик

**Предполагаемый результат деятельности:** практическая работа выполнена

**Предполагаемая форма результата деятельности:** Colab- ноутбук, python ноутбук

**Перечень инструментов, необходимых для реализации деятельности:**  
лекционный материал, презентация

**Критерии оценки деятельности:**

- результат деятельности полностью или частично соответствует поставленной задаче - зачтено
- результат деятельности не соответствует поставленной задаче или отсутствует - не зачтено


**Характер деятельности:** индивидуальный

## **ОПИСАНИЕ ЗАДАНИЯ**

**Практическое задание в Python:**

Задача разделения апельсинов и грейпфрутов довольно очевидна для человека, но даже при ручном наблюдении все равно есть небольшая погрешность. Набор данных включает в себя цвет, вес и диаметр "среднего" апельсина и грейпфрута и генерирует большой набор данных, содержащий широкий спектр значений и представляющий собой "апельсины" и "грейпфрут". Постройте на основе данных с [kaggle](https://www.kaggle.com) модель классификации, выполнив действия ниже:

- 1) Загрузите данные по ссылке с [kaggle](https://www.kaggle.com) при помощи csv или kaggle api;

- 
- 2) Изучите загруженные данные. При необходимости, визуализируйте их, постройте описательную статистику. Закодируйте целевую переменную при помощи функции Label Encoder из sklearn. Проведите нормализацию переменных для модели;
  - 3) Сделайте разбиение выборки на Train и Test при помощи функции train\_test\_split из sklearn;
  - 4) Импортируйте метрики precision, recall, roc-auc, accuracy, confusion matrix, f1\_score и pr-auc из sklearn, а также алгоритмы логистической регрессии, решающих деревьев и опорных векторов.
  - 5) Постройте три модели на обучающих данных и сравните их по метрикам между собой. Какая модель является наилучшей и почему?

**\*Задания повышенной сложности:**

- 1) Подберите оптимальные гиперпараметры для получения наилучшей метрики ROC-AUC для трех моделей при помощи функции GridSearchCV из sklearn.
- 2) Какая модель является наилучшей теперь?