# ЗАДАНИЕ

В задании второго тура нужно решить задачу классификации типа стекол. Данные для обучения моделей размещены в папке «ИИ_Олимпиада им. Н.С. Стрелецкого» на рабочем столе.

Целевая переменная – тип стекла «Type». Остальные признаки описывают химические элементы в составе материала. Датасет нужно исследовать на наличие выбросов, провести EDA.

**Этапы работы:**

1. Получите данные и загрузите их в рабочую среду (Jupyter Notebook или другую)
2. Проведите первичный анализ. Проверьте количество записей для каждого класса. Сделайте вывод.
3. Разделите выборку на обучающее и тестовое подмножество. 80% данных оставить на обучающее множество, 20% на тестовое. Обучите модель дерева решений RandomForestClassifier на обучающем множестве.
4. Для тестового множества предскажите тип стекла и сравните с истинным значением, посчитав точность предсказания модели (accuracy).
5. Обработайте выбросы в данных:
   a) Визуализируйте распределение значений для каждой переменной. Можно использовать функции sns.boxplot, sns.distplot. Есть ли признаки с нормальным распределением?
   b) Исследуйте признаки на выбросы несколькими способами.
   c) Удалите выбросы. *Посчитайте процент удаленных записей от общего числа записей для каждого класса.
6. Повторите п. 4, п. 5.
7. Сформулируйте выводы по проделанной работе:
   a) Кратко опишите, какие преобразования были сделаны с данными.
   b) Сравните точность двух моделей.
   c) Напишите свое мнение, нужно ли исследовать данные на выбросы, для чего это делается, плюсы и минусы подхода.
8. Форма выполнения (представления результатов работы):
   - ссылка на Jupyter Notebook, загруженный на GitHub;
   - ссылка на Google Colab;
   - файл с расширением .ipynb .

Инструменты:

- Jupyter Notebook/Google Colab;
- GitHub;
- данные для обучения моделей;
- модель дерева решений RandomForestClassifier.

*Рекомендации к выполнению:*

- Текст оформляйте в отдельной ячейке Jupyter Notebook/Google Colab в формате markdown.
- У графиков должен быть заголовок, подписи осей, легенда (опционально). Делайте графики бОльшего размера, чем стандартный вывод, чтобы увеличить читабельность.
- Убедитесь, что по ссылкам есть доступ на чтение/просмотр.
- Убедитесь, что все ячейки в работе выполнены и можно увидеть их вывод без повторного запуска.