Модуль 2  
Машинное обучение

короткая линия

# **1. Как установить платформу H2O?**

## Ответ 1 (для windows)

1. Нажимаем win-R, запускаем интерпретатор командной строки cmd
2. Проверяем наличие java с помощью команды java -version
3. Если все в порядке, переходим в пункту 6.
4. Переходим по ссылке <https://java.com>
5. Нажимаем на кнопку [Загрузить java бесплатно], соглашаемся с соглашением и скачав инсталлятор, устанавливаем виртуальную машину (запускаем скачанный файл)
6. Переходим по ссылке <https://h2o-release.s3.amazonaws.com/h2o/rel-zizler/4/index.html>
7. Нажимаем на кнопку [DOWNLOAD H2O], скачиваем архив
8. Извлекаем папку из архива, переходим в нее и запускаем файл h2o.jar
9. Переходим по ссылке [http://localhost:54321](http://localhost:54321/)
10. При работе не используем кириллицу в именах папок и файлов

# 2. Как оформить задание?

## Ответ

Добавить в блокнот текстовые ячейки с описанием датасета, комментариями, выводом. Примеры описаний:

1. <https://ansmirnov.ru/kaggle-titanic-dataset/>
2. <https://neurohive.io/ru/tutorial/primer-reshenija-realnoj-zadachi-po-mashinnomu-obucheniju-na-python/>

Полезные ссылки:

1. <https://www.rea.ru/ru/org/managements/orgnirupr/Documents/gost_7.32-2017.pdf>

# 3. Как бороться с несбалансированными классами в данных?

## Ответ

1. <https://loginom.ru/blog/imbalance-class>
2. <https://dyakonov.org/2021/05/27/imbalance/>
3. <https://techarks.ru/qa/python/borba-s-disbalansom-klassa-ES/>
4. <https://www.machinelearningmastery.ru/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/>

# 4. Какую выбрать задачу(модель) для моих данных?

## Ответ

1. <https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html>
2. <https://learn.microsoft.com/ru-ru/dotnet/machine-learning/resources/tasks>

короткая черта