Задание 3. Исследование алгоритмов классификации

Для выполнения данного задания вам необходимо сначала оценить, насколько набор

данных, выбранный вами ранее, подходит для решения задачи. Для этого следует:

1) Оценить какой атрибут выступит в роли целевого класса (метки)

2) Оценить сбалансированность классов

В случае, если невозможно выбрать атрибут, выступающего в роли класса, необходимо

подобрать подходящий датасет.

Задание состоит из последовательного выполнения следующих подзадач.

1. Необходимо оценить и сравнить результаты классификации, используя следующие

алгоритмы классификации:

\* kNN

\* дерево решений

2. Сравните полученные результаты с помощью различных метрик оценки качества:

Accuracy

Presicion, Recall, F-measure

ROC

3. Объяснить полученные результаты

Отчет должен включать описания выполнения каждой подзадачи.

Дополнительная литература

Оценка качества в задачах классификации и регрессии

http://neerc.ifmo.ru/wiki/index.php?title=%D0%9E%D1%86%D0%B5%D0%BD%D0%BA%D0%

B0\_%D0%BA%D0%B0%D1%87%D0%B5%D1%81%D1%82%D0%B2%D0%B0\_%D0%B2\_%

D0%B7%D0%B0%D0%B4%D0%B0%D1%87%D0%B0%D1%85\_%D0%BA%D0%BB%D0%B0

%D1%81%D1%81%D0%B8%D1%84%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D0%B8

\_%D0%B8\_%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D0%

B8

Очень хорошая лекция с примерами

https://vk.com/video-158557357\_456239019

Та же лекция на хабре https://habr.com/ru/company/ods/blog/322534/

Ссылка на ноутбук из лекции https://github.com/Yorko/mlcourse.ai/blob/master/jupyter\_russian/topic03\_decision\_trees\_knn/topic3\_trees\_knn.ipynb