

Лабораторная работа № 1 по машинному обучению

Постановка задачи:

Найти себе набор данных (датасет), для следующей лабораторной работы, и проанализировать его. Выявить проблемы набора данных, устранить их. Визуализировать зависимости, показать распределения некоторых признаков. Реализовать алгоритмы К ближайших соседа с использованием весов и Наивный Байесовский классификатор и сравнить с реализацией библиотеки sklearn.

Подсказки:

Наборы данных можно найти здесь:

- 1) [Dataset Search \(google.com\)](https://www.google.com/datasetsearch),
- 2) [Topics - World and regional statistics, national data, maps, rankings \(knoema.com\)](https://knoema.com),
- 3) [UCI Machine Learning Repository](https://mlproject.github.io/)
- 4) Quandl
- 5) Kaggle
- 6) <https://scikit-learn.org/stable/modules/preprocessing.html>

Про проблемы в данных можно почитать здесь:

- 1) <https://www.kaggle.com/learn/data-cleaning>
- 2) <https://habr.com/ru/post/470650/>
- 3) <https://habr.com/ru/post/173049/>

Про KNN можно прочитать:

- 1) http://www.machinelearning.ru/wiki/index.php?title=Метод_ближайших_соседей
- 2) <https://scikit-learn.org/stable/modules/neighbors.html>
- 3) <https://nlp.stanford.edu/IR-book/html/htmledition/k-nearest-neighbor-1.html>

Про Байесовский классификатор можно прочитать:

- 1) http://www.machinelearning.ru/wiki/index.php?title=Байесовский_классификатор
- 2) https://neerc.ifmo.ru/wiki/index.php?title=Байесовская_классификация

- 3) <https://habr.com/ru/post/120194/>
- 4) <https://ru.coursera.org/lecture/trendy-klassifikatsii/4-3-baiiesovskii-klassifikator-JlMmI>