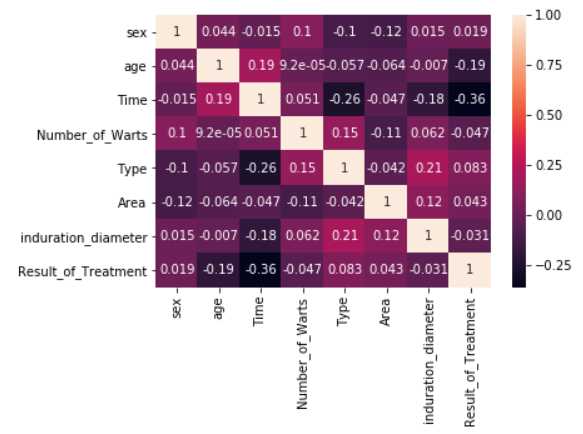
# Отчёт по лабораторной работе №1 Базовые алгоритмы классификации

Выполнил студент 5 курса ФТИ  
Группы ФИ-83

Малышко Антон

Dataset: <https://archive.ics.uci.edu/ml/datasets/Immunotherapy+Dataset>  
В нём описан результат лечения бородавок 90 пациентов с помощью иммунотерапии. Отслеживалось 7 параметров (плюс результат лечения).

Наибольшую корреляцию с разметкой имеют параметры Time (-0.36) и age (-0.19).

DecisionTreeClassifier также отдал предпочтение в первую очередь этим атрибутам.

Среди рассмотренных классификаторов, наиболее стабильный результат показал DecisionTree(criterion="entropy", max\_depth=1). Для него:  
Значения каппа для 10 разных разделений train\_test\_split(test\_size = 0.3):

0.577 0.509 0.539 0.786 0.229 0.509 0.609 0.430 0.502 0.658

Логистическая регрессия научилась любой вход классифицировать как «1»

KNN показывает наилучший результат при k = 1, который в среднем хуже чем для DTree (0.3 – 0.5 против 0.4 – 0.6).

SVM при разном разделении данных на обучающую и тестовую части имеет разные оптимальные гиперпараметры (поиск с пом. GridSearchCV), результат значительно хуже чем у KNN и DTree.

Поскольку датасет небольшой, для оценки качества классификатора он был обучен и протестирован 10 раз для разных разбиений выборки на обучающую и тестовую части.