

Московский Авиационный Институт
(Национальный Исследовательский Университет)
Факультет: «Информационные технологии и прикладная математика»

Кафедра вычислительной математики и программирования

Лабораторная работа №1 По курсу «Машинное обучение»:

Студент: Кудинов С.В.
Группа: М8О-306Б-18
Преподаватель: Ахмед Саир Халид
Дата:
Оценка:
Подпись:

Москва, 2021

Задача:

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 Кближайших соседа с использованием весов и Наивный Байесовский классификатор и сравнить с реализацией библиотеки sklearn.

Метод решения:

Я нашел датасет с медицинскими данными (давление, уровень глюкозы в крови, инсулин, возраст). В зависимости от медицинских данных нужно было предсказать, есть ли диабет у пациента или нет

В этом датасете была проблема: отсутствовало давление, инсулин у некоторых пациентов. Чтобы решить эту проблему, я заменил отсутствующие значения в столбцах средними.

Далее я реализовал свой KNN и наивный байесовский классификатор с распределением Гаусса и сравнил результаты с библиотекой `skLearn`. Подробнее можно посмотреть в файле `Jupyter Notebook`

Результаты:

Точность KNN(sklearn) : 0.7532467532467533

Точность собственного KNN: 0.7597402597402597

Точность Наивного Байесовского классификатора(skLearn): 0.7532467532467533

Точность своего Байесовского классификатора: 0.7467532467532467