

Лабораторная работа

Решите пожалуйста следующие задачи.

1. Примените алгоритм KNN (k-ближайших соседей) для данных по заболеваемости сахарным диабетом. Датасет представлен по [ссылке](#).

Выходной столбец "outcome" принимает значения 0 или 1 и является классификатором текущего набора данных.

- имеется диабет (значение 1)
- заболевание отсутствует (значение 0)

Используйте все остальные столбцы как data features (входные числовые данные) для адаптации модели kNN.

Рассчитайте модель kNN для $k = 5, 10, 15, 20, 25$. Укажите объем тестового набора данных равным 25% от общего количества данных.

Укажите score модели и постройте визуализацию обученной классификации для указанных k.

2. Постройте модель множественной линейной регрессии для атрибутов $X = \{\text{SkinThickness}, \text{BMI}\}$ и $Y = \{\text{Insulin}\}$ из датасета предыдущего задания.

Укажите score модели и предоставьте значения b, a_1, a_2 формулы:

$$y = b + a_1 \cdot x_1 + a_2 \cdot x_2$$

3. Примените к указанным данным любую одну модель (решите сами - для задачи классификации или задачи регрессии) из списка ниже
 - SVM
 - DecisionTree
 - Полиномиальная регрессия