

Домашнее задание 6

Дедлайн сдачи работы: **23.02.2026, 23:59**

Для выполнения заданий используйте датасет `diabetes_data.csv`

Описание датасета можно найти по ссылке: Кликабельная ссылка на описание датасета

1. Загрузите датасет, проведите описательную статистику переменных и сделайте выводы о данных
2. Выберите 3 категориальные переменные и не менее 3 количественных переменных, чтобы различать пациентов с диабетом и без него (переменная `Diagnosis`). Выбор переменных может основываться на ваших предположениях о предикторах диабета или на основе чтения литературы по этой теме.
3. Изучите связь между выбранными переменными и зависимой переменной.
4. Изучите связь между выбранными предикторами.
5. Постройте модель логистической регрессии со всеми выбранными предикторами и включите в нее взаимодействие между двумя категориальными переменными.
6. Проверьте допущения и проинтерпретируйте результаты.
7. Проинтерпретируйте значимость модели в целом (работает ли она лучше случайного угадывания?)
8. Содержательно проинтерпретируйте каждый значимый предиктор.
9. Попробуйте убрать из модели незначимые предикторы (используйте функцию `drop1()`, чтобы выбрать предиктор для удаления). Стала ли модель от этого лучше?

10. Исследуйте качество построенной модели (подсказка: посмотрите на accuracy, precision, recall, ROC-AUC).