

Загрузите датасет из файла `diabetes.csv` и ответьте на следующие вопросы:

1. У какого количества пациентов из датасета присутствует нарушенная толерантность к глюкозе (НТГ)? Критерий НТГ – это уровень повышения гликемии ≥ 7.8 ммоль/л на 120 минуте теста на толерантность к глюкозе. Сколько в датасете пациентов, у которых нет НТГ?
2. Как выглядит ROC-кривая для предсказания сахарного диабета по переменной, характеризующей уровень гликемии? Постройте график.
3. Чему равна площадь под ROC-кривой, которую вы построили в вопросе 2?
4. Чему равен 95% двусторонний ДИ для площади под ROC-кривой, которую вы построили в вопросе 2?
5. Постройте ROC-кривую и определите, какое пороговое значение является оптимальным для предсказания сахарного диабета по уровню инсулина? Какой чувствительностью и специфичностью обладает данный порог?
6. Какая из количественных переменных в датасете обладает наибольшей площадью под ROC-кривой? Как вы можете интерпретировать это знание? Какая количественная переменная имеет наименьшую площадь?

Форма представления ответов – документ word, pdf или html. Документ можете как составлять вручную, так и сгенерировать с помощью `markdown` или `quarto`.

Подсказки для работы:

Описание датасета `pima` (`diabetes.csv`):

- `Pregnancies` – количество беременностей в анамнезе (кол-во);
- `Glucose` – уровень глюкозы на 120 минуте орального глюкозотолерантного теста (мг/дл);
- `BloodPressure` – диастолическое артериальное давление (мм рт.ст.);
- `SkinThickness` – толщина кожной складки на трицепсе (мм),
- `Insulin` – уровень инсулина (мЕ/мл);
- `BMI` – индекс массы тела (кг/м^2);
- `DiabetesPedigreeFunction` – индекс, отражающий вероятность наличия диабета на основании наследственного анамнеза;
- `Age` – возраст (лет);
- `Outcome` – наличие сахарного диабета (0 – нет, 1 – да);

Оригинальная статья с подробностями об этих данных также выложена в папке диска.

1. Сахарный диабет — это хроническое эндокринное заболевание, сопровождающееся повышенным уровнем глюкозы;
2. Для уровня глюкозы $1 \text{ ммоль/л} = 18 \text{ мг/дл}$;
3. Не забудьте отфильтровать или заменить на пропуски некорректные нулевые значения в количественных переменных.