

Защита персональных данных клиентов

Поздравляем! Вы прошли курс в тренажёре. Самое время проверить знания и решить новую задачу машинного обучения. Выполнять работу будете самостоятельно.

Когда закончите, отправьте её на проверку ревьюеру: он пришлёт комментарии в течение суток. После этого нужно доработать проект и пройти повторную проверку.

Скорее всего, вы будете дорабатывать кейс по комментариям ещё несколько раз. Это нормально.

Проект завершён, когда ревьюер одобрит все доработки.

Описание проекта

Вам нужно защитить данные клиентов страховой компании «Хоть потоп». Разработайте такой метод преобразования данных, чтобы по ним было сложно восстановить персональную информацию. Обоснуйте корректность его работы.

Нужно защитить данные, чтобы при преобразовании качество моделей машинного обучения не ухудшилось. Подбирать наилучшую модель не требуется.

Инструкция по выполнению проекта

1. Загрузите и изучите данные.
2. Ответьте на вопрос и обоснуйте решение.
Признаки умножают на обратимую матрицу. Изменится ли качество линейной регрессии? (Её можно обучить заново.)
 - a. Изменится. Приведите примеры матриц.
 - b. Не изменится. Укажите, как связаны параметры линейной регрессии в исходной задаче и в преобразованной.
3. Предложите алгоритм преобразования данных для решения задачи. Обоснуйте, почему качество линейной регрессии не поменяется.
4. Запрограммируйте этот алгоритм, применив матричные операции. Проверьте, что качество линейной регрессии из *sklearn* не отличается до и после преобразования. Примените метрику R^2 .

Описание данных

Набор данных находится в файле `/datasets/insurance.csv`. [Скачать датасет](#).

- **Признаки:** пол, возраст и зарплата застрахованного, количество членов его семьи.
- **Целевой признак:** количество страховых выплат клиенту за последние 5 лет.