Университет ИТМО

Практическая работа №5

по дисциплине «Визуализация и моделирование»

Автор: Никулина Ксения Григорьевна
Поток: ВИМ 1.1
Группа: К3221
Факультет: ИКТ
Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

1 Heart Failure Prediction

Сердечно-сосудистые заболевания являются причиной смерти номер 1 во всем мире, забирая в среднем 17.9 миллионов жизней в год, что составляет 31% от всех смертей по всему миру. Остановка сердца - результат сердечно-сосудистых заболеваний и данный датасет содержит 12 показателей, которые можно использовать для предсказания смертности от остановки сердца.

Название столбца	Описание данных	Тип данных	Шкала
Возраст	Возраст пациента	Integer	Интерв
Анемия	Снижение гемоглобина	Binary	Номин
Креатинкиназа	Уровень фермента	Integer	Относит
Диабет	Наличие диабета	Binary	Номин
Фракция выброса	% крови при сокращении	Integer	Относит
Высокое кровяное давление	Гипертония	Binary	Номин
Тромбоциты	Тромбоциты в крови	Float	Номин
Креатинин	Уровень креатинина в крови	Float	Номин
Натрий	Уровень натрия в крови	Integer	Относит
Пол	Пол пациента	Binary	Номин
Курение	Курит ли пациент	Binary	Номин
Время	Кол-во дней наблюдения	Integer	Относит
Смерть	Умер ли пациент	Binary	Номин

2 Анализ датасета

Обучение с учителем (supervised learning) предполагает наличие полного набора размеченных данных для тренировки модели на всех этапах ее построения.

Наличие полностью размеченного датасета означает, что каждому примеру в обучающем наборе соответствует ответ, который алгоритм и должен получить. Таким образом, размеченный датасет из фотографий цветов обучит нейронную сеть, где изображены розы, ромашки или нарциссы. Когда сеть получит новое фото, она сравнит его с примерами из обучающего датасета, чтобы предсказать ответ.

Я считаю, мой датасет подходит для обучения с учителем, так как на вход он будет получать входные данные - показатели пациента, а на выход его шансы на смерть, основываясь на предыдущем опыте осматриваемых пациентов. Таким образом, можно будет предсказать шансы на выживание у поступившего на осмотр пациента. Данная задача является задачей регрессии. Отклик - действительное число 0 или 1 (0 - жив, 1 - смерть)

3 CRISP-DM анализ

1. Понимание бизнеса

   - Формулировка глобальной задачи: предсказание смерти на этапе обследования пациента, предотвращение ранней смерти пациента
   - Задача: предотвращение ранней смерти пациента

2. Понимание данных

   - Данные: датасет Heart Failure Prediction
   - Описание данных: таблица
   - Исследование и обработка данных: см. лаб 2-4

3. Подготовка данных

   - Обработка: см. лаб 4

4. Моделирование

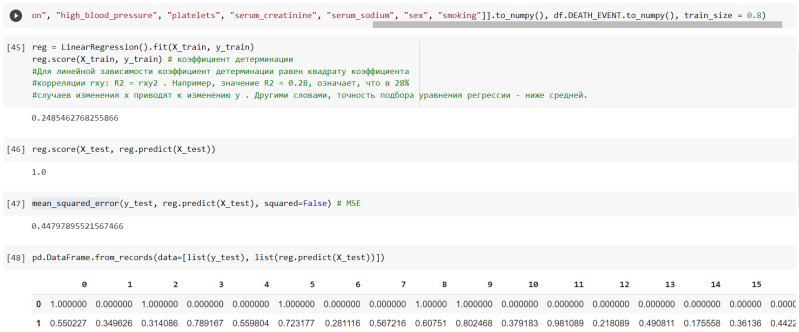

Рис. 1 Моделирование

5. Оценка

   - Благодаря анализу при помощи линейной регрессии мы заметили, что изменение наших данных практически не влияют на изменение результата. Это может быть связано с тем, что количество разнообразных данных недостаточно, так как в датасете всего около 200 строк и для более точной информации требует более расширенный датасет.