



МАШИННОЕ ОБУЧЕНИЕ И БОЛЬШИЕ ДАННЫЕ

СЕССИЯ 3

СОДЕРЖАНИЕ

Сессия 3 данного Конкурсного задания состоит из следующей документации / файлов:

- Инструкция к третьей сессии
- Файл, содержащий целевые переменные
- Архив, содержащий информацию с датчиков по каждому пациенту
- Файл с описанием признаков в датасете

ВВЕДЕНИЕ

На этой сессии, вы будете продолжать исследования, опираясь на то, что вы уже разработали. **Если вы не выполнили полностью задание в предыдущей сессии, не выполняйте его сейчас.** В этой сессии вы будете разрабатывать прогнозную модель для оценки продолжительности жизни пациентов. Вам предстоит подобрать алгоритмы прогнозирования продолжительности жизни пациента и классификации выживаемости пациента в результате сравнения нескольких вариантов и оценить точность их работы.

ИНСТРУКЦИЯ УЧАСТНИКУ

Убедитесь, что вы сохранили отчет о проделанной работе. К концу этой сессии у вас должны быть достигнуты следующие практические результаты:

ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

3.1. Отбор признаков
Необходимо определить, какие признаки имеют наибольшее влияние на продолжительность жизни и выживаемость пациента и оставить только их для обучения. Необходимо обосновать выбор признаков, оставленных для обучения (например, на основе оценки их информативности).
3.2. Подбор и реализация алгоритмов выявления зависимостей параметров данных
Подберите алгоритмы (не менее 3-х) для выявления зависимостей признаков, отобранных в предыдущем задании. Обоснуйте выбор. Реализуйте эти алгоритмы для оценки связи между входными признаками. Интерпретируйте результат. Визуализируйте полученные зависимости
3.3. Разбиение обработанных данных на обучающую, валидационную и тестирующую выборки
Необходимо определить принцип разделения данных на выборки. Обоснуйте выбор. Даже если Вы выбираете случайное разделение данных, необходимо обосновать выбор данного принципа.
3.4. Выбор модели прогнозирования и настройка гиперпараметров
Необходимо построить регрессионную модель для прогноза продолжительности жизни пациента (рассмотреть не менее трех алгоритмов построения регрессионной модели). Настроить гиперпараметры всех моделей, подобранных с помощью валидационной выборки. Выбрать модель для дальнейшего прогнозирования и обосновать выбор.
3.5. Прогноз значения целевой переменной
Обучить выбранную модель на обучающей выборке, предоставить оценку качества модели на тестовой выборке.



МАШИННОЕ ОБУЧЕНИЕ И БОЛЬШИЕ ДАННЫЕ

3.6. Выбор модели классификации и настройка гиперпараметров
Необходимо подобрать модель классификации выживаемости пациента (рассмотреть не менее трех моделей). Настроить гиперпараметры всех моделей, подобранных с помощью валидационной выборки. Выбрать модель для дальнейшей классификации и обосновать выбор.
3.7. Прогноз значения целевой переменной, качество модели
Обучите выбранную модель на обучающей выборке. Оцените качество работы построенного классификатора на тестовой выборке. Рассчитайте метрики оценки качества классификации: weighted accuracy, F1-score. Постройте ROC-кривые. Интерпретируйте результат.
3.8. Подготовка отчета
Подготовьте отчет, содержащий результаты, полученные в пп. 3.1-3.7. Загрузите для проверки отчет, а также исходный программный код.