Московский государственный технический университет им. Н.Э. Баумана

Кафедра «Системы обработки информации и управления»

****

Лабораторная работа №2

по дисциплине

«Методы машинного обучения»

на тему

## **«Обработка признаков (часть 1)»**

Выполнил:

студент Люй Чжэ

группы ИУ5И-23М

Москва — 2024 г.

**1. Цель лабораторной работы**

изучение продвинутых способов предварительной обработки данных для дальнейшего формирования моделей.

**2. Задание**

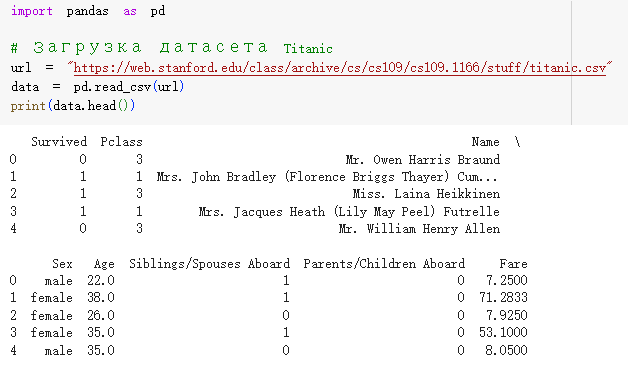
* Выбрать набор данных (датасет), содержащий категориальные и числовые признаки и пропуски в данных. Для выполнения следующих пунктов можно использовать несколько различных наборов данных (один для обработки пропусков, другой для категориальных признаков и т.д.) Просьба не использовать датасет, на котором данная задача решалась в лекции.
* Для выбранного датасета (датасетов) на основе материалов лекций решить следующие задачи:

1. устранение пропусков в данных;
2. кодирование категориальных признаков;
3. нормализация числовых признаков.

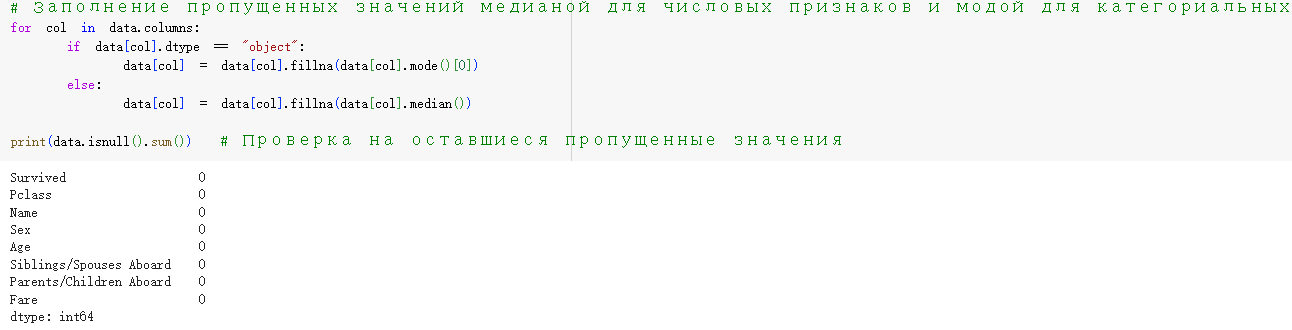
**3. Ход выполнения работы**

Для выполнения задания мы сначала выберем подходящий датасет, затем выполним обработку данных, включая устранение пропусков, кодирование категориальных признаков и нормализацию числовых признаков. В качестве примера давайте используем популярный датасет "Titanic", который содержит как категориальные, так и числовые признаки и имеет пропущенные значения.

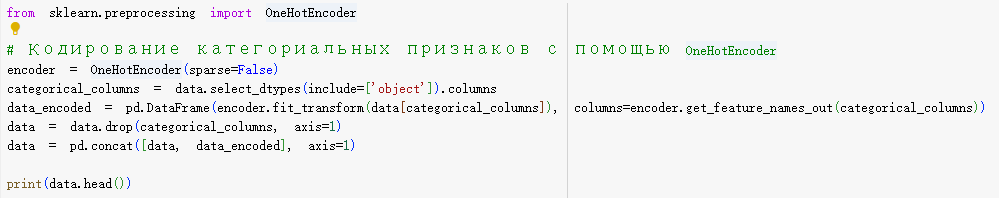
**Шаг 1: Загрузка датасета**

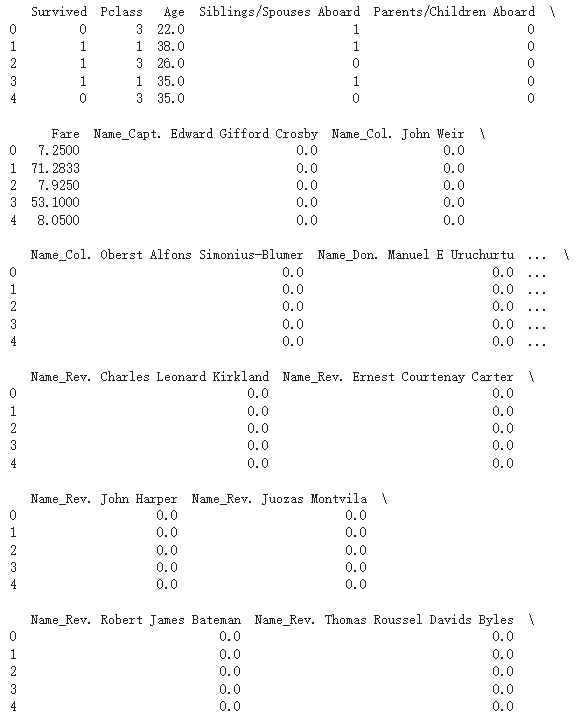


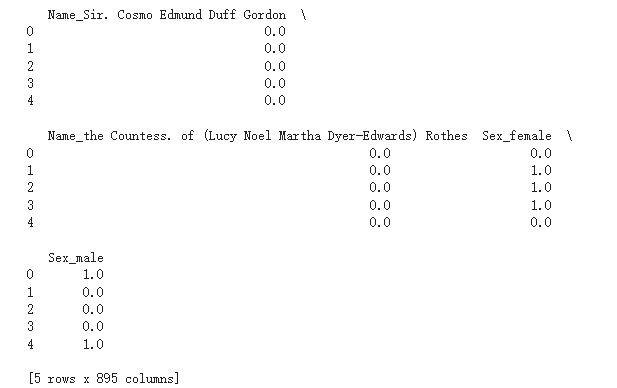
**Шаг 2: Устранение пропусков в данных**



**Шаг 3: Кодирование категориальных признаков**







**Шаг 4: Нормализация числовых признаков**

