



**Московский Государственный Технический Университет имени  
Н.Э.Баумана**

**Факультет Информатика и системы управления**

**Кафедра ИУ-5**

**«Системы обработки информации и управления»**

**Отчёт по Рубежному контролю No 1**

**Методы обработки данных**

**Выполнили студенты группы ИУ-5И 21М**

**Фань Лицзе**

**Москва 2024г.**

## Задача №5.

Для набора данных проведите кодирование одного (произвольного) категориального признака с использованием метода "one-hot encoding".

Выберите библиотеку для загрузки встроенных наборов данных. В качестве примера используйте один из встроенных наборов данных библиотеки seaborn. Библиотека seaborn имеет несколько встроенных наборов данных, например набор данных «Титаник», который содержит некоторую информацию о пассажирах, включая класс их салона (класс), пол (пол), возраст и другую информацию. Здесь мы будем использовать столбец «пол» для «горячего кодирования».

```
pip install pandas seaborn
```

Реализовано с использованием библиотеки pandas.

```
import pandas as pd
import seaborn as sns

# 加载数据集
data = sns.load_dataset('titanic')

# 选择'embarked'作为要编码的列（表示乘客登船的地点）
one_hot_encoded = pd.get_dummies(data['embarked'], prefix='embarked')

print(one_hot_encoded.head())
```

Приведенный выше код выполняет горячее кодирование «погруженного» столбца набора данных «Титаник». Каждая категория будет преобразована в новый столбец, и 0 и 1 будут использоваться для указания того, принадлежит ли она к этой категории.

	embarked_C	embarked_Q	embarked_S
0	False	False	True
1	True	False	False
2	False	False	True
3	False	False	True
4	False	False	True