

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 № 1

По дисципление

«Методы Машинного Обучения»

Группа ИУ5И-23М

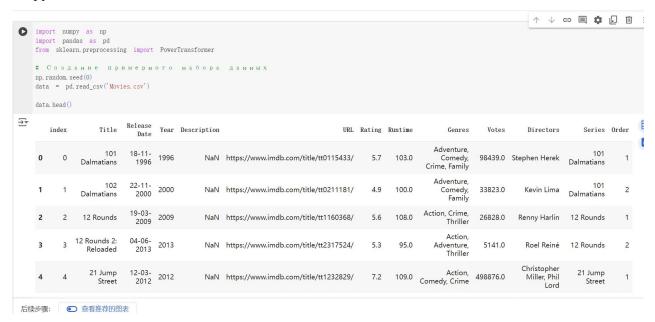
Ли Хао

Номер варианта: 17

Номер задачи №1: 17

Для набора данных проведите нормализацию для одного (произвольного) числового признака с использованием преобразования Йео-Джонсона (Yeo-Johnson transformation).

Загрузить данные



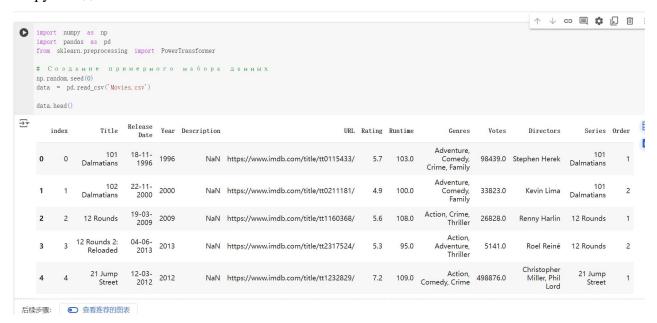
Yeo-Johnson transformation



Номер задачи №2: 37

Для набора данных проведите процедуру отбора признаков (feature selection). Используйте класс SelectPercentile для 5% лучших признаков, и метод, основанный на взаимной информации.

Загрузить данные



Использование класса SelectPercentile для 5% лучших признаков, и метод, основанный на взаимной информации.

```
import numpy as np import pandas as pd from sklearn. feature_selection import SelectPercentile, mutual_info_regression

# Cosganue примерного набора данных np. random. seed(0) data = {
        'feature1': np. random. normal(loc=0, scale=1, size=100), # нормальное распределение 'feature2': np. random. exponential(scale=1, size=100), # экспоненциальное распределение 'feature2': np. random. uniform(low=-1, high+1, size=100), # равномерное распределение 'target': np. random. normal(loc=0, scale=1, size=100) # целева и переменная
}

df = pd. DataFrame(data)

# Определение признаков и целевой переменной X = df. drop(columns='target')

y = df['target']

# Применение SelectPercentile с использованием mutual_info_regression selector = SelectPercentile(mutual_info_regression. percentile=5)

X_selected = selector.fit_transform(X, y)

# Получение маски отобранных признаков selecte_features = X. columns[selector.get_support()]

selected_features

Index(['feature3'], dtype='object')
```