



Министерство науки и высшего образования Российской
Федерации Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Московский государственный технический
университет имени Н.Э. Баумана
(национальный исследовательский
университет)» (МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ _____ Информатика и системы управления

КАФЕДРА _____ Системы обработки информации и управления

Рубежный контроль №1
По курсу «Технологии машинного обучения»
Вариант 20

Подготовила:

Студентка группы ИУ5-64Б

Тахтамышева А.А.

Проверил:

Преподаватель кафедры ИУ5

Гапанюк Ю.Е.

Москва, 2022 г.

Тема: Технологии разведочного анализа и обработки данных.

Номер варианта	Номер задачи	Номер набора данных, указанного в задаче
20	3	4

Дополнительные требования по группам:

- Для студентов группы ИУ5-64Б - для набора данных построить "Скрипичная диаграмма (violin plot)".

Задача №3.

Для заданного набора данных произведите масштабирование данных (для одного признака) и преобразование категориальных признаков в количественные двумя способами (label encoding, one hot encoding) для одного признака. Какие методы Вы использовали для решения задачи и почему?

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import MinMaxScaler, StandardScaler, Normalizer

data = pd.read_csv('states_all.csv', sep=",")
data.shape
data.isnull().sum()
data.head()
data.dtypes
#данный признак является единственным категориальным в данном датасете,
#поэтому кодирование будет проводиться относительно данного признака
state_df = pd.DataFrame(data, columns=['STATE'])
#Выполним преобразование категориальных признаков в количественные
le = LabelEncoder()
cat_enc_le = le.fit_transform(state_df['STATE'])
# Наименования категорий в соответствии с порядковыми номерами

# Свойство называется classes, потому что предполагается что мы решаем
# задачу классификации и каждое значение категории соответствует
# какому-либо классу целевого признака

le.classes_
cat_enc_le
ohe = OneHotEncoder()
cat_enc_ohe = ohe.fit_transform(state_df[['STATE']])
cat_enc_ohe
cat_enc_ohe.todense()[0:50]
state_df.head(50)
```

```
#выполним MinMax масштабирование
sc1 = MinMaxScaler()
sc1_data = sc1.fit_transform(data[['GRADES_ALL_G']])
plt.hist(data['GRADES_ALL_G'], 50)
plt.show()
plt.hist(sc1_data, 50)
plt.show()

#выполним масштабирование на основе Z-оценки
sc2 = StandardScaler()
sc2_data = sc2.fit_transform(data[['GRADES_ALL_G']])
plt.hist(sc2_data, 50)
plt.show()

#Построение скрипичной диаграммы (violin plot)
sns.violinplot(x=data['ENROLL']) #диаграмма о количестве учащихся
```