Лабораторная работ 1.
Стандартизация данных.

Цель: познакомиться с методами стандартизации данных из библиотеки Scikit Learn.

Задачи:

1. Стандартизировать данные используя sklearn;
2. Привести данные к диапазону;
3. Привести данные к равномерному распределению.

Пример выполнения лабораторной работы.

Для выполнения лабораторной работы подойдет датасет из примера lab1.csv или можно использовать любой общедоступный.

Стандартизируем данные из колонки «age», методом sklearn с помощью normalize(), предварительно преобразовав его в массив.

```
1.import pandas as pd
2.import numpy as np
3.from sklearn import preprocessing
4.df = pd.read_csv('lab1.csv',encoding='cp1251')
5.array = np.array(df['age'])
6.normal_arr = preprocessing.normalize([array])
7.print(normal_arr)
```

Результат:

```
[[0.03280067 0.03625337 0.03452702 0.03970608 0.05351689 0.05524324
  0.06042229 0.06905405 0.11048648 0.04315878 0.11566553 0.06042229
  0.06042229 0.04143243 0.06905405 0.03797973 0.06042229 0.03797973
  0.03797973 0.06042229 0.04315878 0.07941215 0.05351689 0.09322296
  0.05006418 0.0776858  0.06042229 0.06905405 0.03970608 0.10358107
  0.03625337 0.09149661 0.03107432 0.08459121 0.03625337 0.07250675
```

А теперь стандартизируем все данные по столбцам axis=0. Для стандартизации по строка необходимо указать axis=1.

```
1.import pandas as pd
2.import numpy as np
3.from sklearn import preprocessing
4.df = pd.read_csv('lab1.csv',encoding='cp1251')
5.normal = preprocessing.normalize(df, axis=0)
6.print(normal)
```

Результат:

```
[[0.02178034 0.4138265  0.32670513 0.84943334]
 [0.02351965 0.2469563  0.17639736 0.95254574]
 [0.11330844 0.75538957 0.60431166 0.22661687]
 [0.04875874 0.28036275 0.19503495 0.93860571]
 [0.09325048 0.57815298 0.31705163 0.74600385]
 [0.07108187 0.37910328 0.20139862 0.90037029]
 [0.09634345 0.48171727 0.86709108 0.0825801 ]
 [0.07689464 0.38447322 0.17301295 0.90351208]
```

Приведем данные к стандартному диапазону (0, 1) с помощью Sklearn методом MinMaxScaler. Для того чтобы изменить диапазон, например (0,3) необходимо указать feature_range=(0,3).

```
import pandas as pd
import numpy as np
from sklearn import preprocessing
df = pd.read_csv('lab1.csv',encoding='cp1251')
scaler = preprocessing.MinMaxScaler()
names = df.columns
d = scaler.fit_transform(df)
scal_df = pd.DataFrame(d, columns=names)
print(scaled_df.head(5))
```

Результат:

	id	age	income	spending_rating
0	0.000000	0.019231	0.000000	0.387755
1	0.005025	0.057692	0.000000	0.816327
2	0.010050	0.038462	0.008197	0.051020
3	0.015075	0.096154	0.008197	0.775510
4	0.020101	0.250000	0.016393	0.397959

В ходе лабораторной работы мы научились стандартизировать данные с помощью sklearn, приводить к данные к диапазону используя MinMaxScaler, MaxAbsScaler, RobustScaler и приводить данные к равномерному распределению используя QuantileTransformer.

Задание.

1. Стандартизировать колонку из своего датасета используя sklearn предварительно преобразовав его в массив.
2. Стандартизировать все данные из датасета по строкам и по столбцам.
3. Привести данные к диапазону (0,3) используя MinMaxScaler.
4. Привести данные к диапазону (-1,1) используя MaxAbsScaler.
5. Привести данные к диапазону (25, 75) используя RobustScaler.
6. Привести данные к равномерному распределению используя QuantileTransformer.

Формат отчета.

Протокол лабораторной работы в формате PDF, который должен содержать, поэтапное выполнение всех задач с текстовым описанием ваших действий и экранными формами, отображающими данные действия.