

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. Н.Э. Баумана

Факультет «Информатика и системы управления»
Кафедра «Систем обработки информации и управления»

ОТЧЕТ

Рубежный Контроль № 1
по дисциплине «Методы машинного обучения»

ИСПОЛНИТЕЛЬ:

группа ИУ5-21М__

__ Лю Бэйбэй ____
ФИО

подпись

"__" _____ 2023__ г.

ПРЕПОДАВАТЕЛЬ:

ФИО

подпись

"__" _____ 202__ г.

Москва - 2023

Вариант:19

Задачи:

Задача №19.

Для набора данных проведите масштабирование данных для одного (произвольного) числового признака с использованием метода "Mean Normalisation"

Набор Данных: /kaggle/input/customers-dataset/Customers.csv

```
df = pd.read_csv("E:\liu\lab\pk\Customers.csv")
```

```
df.head()
```

	CustomerID	Gender	Age	Annual Income (\$)	Spending Score (1-100)	Profession	Work Experience	Family Size
0	1	Male	19	15000	39	Healthcare	1	4
1	2	Male	21	35000	81	Engineer	3	3
2	3	Female	20	86000	6	Engineer	1	1
3	4	Female	23	59000	77	Lawyer	0	2
4	5	Female	31	38000	40	Entertainment	2	6

```
mean_of_Income=np.mean(df['Annual Income ($)'])  
print('mean_of_Income=', mean_of_Income)
```

```
mean_of_Income= 110731.8215
```

```
maximum=max(df['Annual Income ($)'])  
minimum=min(df['Annual Income ($)'])  
l=maximum-minimum  
new_Income=(df['Annual Income ($)']-mean_of_Income)/l  
print(new_Income)
```

```
0    -0.503921  
1    -0.398643  
2    -0.130185  
3    -0.272310  
4    -0.382851  
...  
1995    0.387712  
1996   -0.197784  
1997   -0.104071  
1998    0.375721  
1999   -0.000641  
Name: Annual Income ($), Length: 2000, dtype: float64
```

Задача №39.

Для набора данных проведите процедуру отбора признаков (feature selection). Используйте класс SelectPercentile для 10% лучших признаков, и метод, основанный на взаимной информации.

Набор Данных: sklearn.datasets.load_digits

```
X, y=load_digits(return_X_y=True)  
X.shape
```

```
(1797, 64)
```

```
X_new=SelectPercentile(mutual_info_classif,percentile=10).fit_transform(X, y)  
X_new.shape
```

```
(1797, 7)
```

Дополнительное требование: для пары произвольных колонок данных построить график "Диаграмма рассеяния"

Набор Данных: /kaggle/input/customers-dataset/Customers.csv

```
plt.figure(figsize=(8,4))
plt.scatter(df['Age'],df['Annual Income ($)'])
plt.xlabel='Age'
plt.ylabel='Annual Income ($)'
plt.show()
```

