

Московский государственный технический университет им. Н.Э. Баумана

Факультет «Информатика и системы управления»
Кафедра ИУ5 «Системы обработки информации и
управления»

Курс

«Технологии машинного обучения»

Отчет по лабораторной работе №6

«Разведочный анализ данных. Исследование и визуализация
данных.»

Выполнил:
студент группы ИУ5-63Б
Воронова О. А.

Проверил:
преподаватель каф. ИУ5
Гапанюк Ю.Е.

Москва, 2022 г.

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split, KFold,
cross_val_score
from sklearn.neighbors import KNeighborsRegressor
from sklearn.metrics import mean_absolute_error,
median_absolute_error, r2_score
data = pd.read_csv("Train.csv")
data.head()

```

	ID	Warehouse_block	Mode_of_Shipment	Customer_care_calls	Customer_rating	Cost_of_the_Product
0	1	D	Flight	4	2	177
1	2	F	Flight	4	5	216
2	3	A	Flight	2	2	183
3	4	B	Flight	3	3	176
4	5	C	Flight	2	2	184

```

data.shape
(10999, 12)

```

```

data.dtypes

```

```

ID                int64
Warehouse_block   object
Mode_of_Shipment  object
Customer_care_calls    int64
Customer_rating      int64
Cost_of_the_Product    int64
Prior_purchases      int64
Product_importance    object
Gender              object
Discount_offered      int64
Weight_in_gms         int64
Reached.on.Time_Y.N    int64

```

```

dtype: object

```

```

LE = LabelEncoder()

```

```

for col in data.columns:

```

```

    if data[col].dtype == "object":

```

```

        data[col] = LE.fit_transform(data[col])

```

```

fig, ax = plt.subplots(figsize=(15,9))

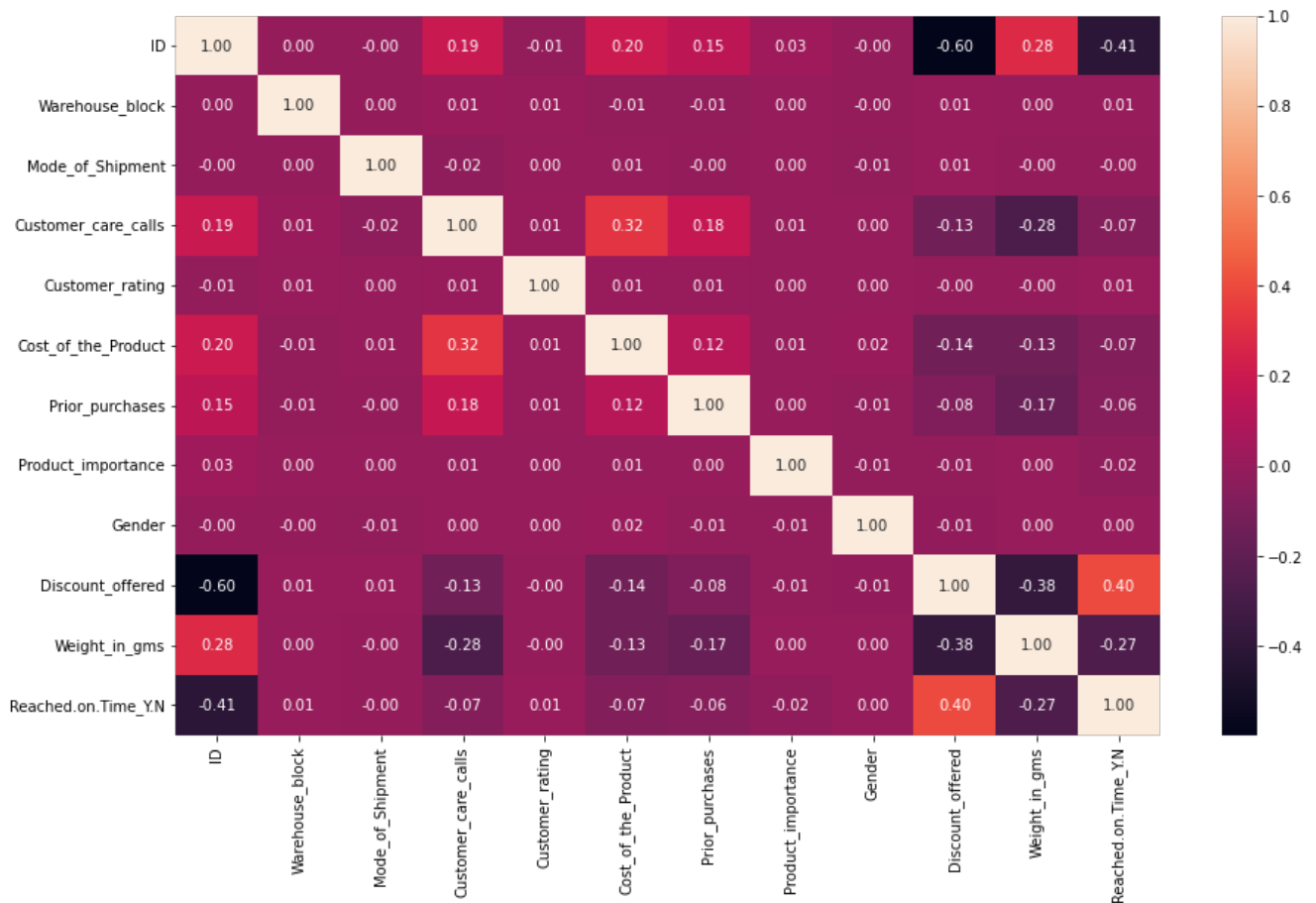
```

```

sns.heatmap(data.corr(method="pearson"), ax=ax,annot=True,
fmt=".2f")

```

<AxesSubplot:>



```
target = "Customer_care_calls"
data[target] = data[target].astype("float")
xArray = data.drop(target, axis=1)
yArray = data[target]
trainX, testX, trainY, testY = train_test_split(xArray, yArray,
test_size=0.2, random_state=1)
trainX
```

	ID	Warehouse_block	Mode_of_Shipment	Customer_rating	Cost_of_the_Product
8854	8855	2	2	1	138
887	888	4	2	3	150
2477	2478	4	0	2	209
89	90	4	2	1	186
3956	3957	0	1	1	157
...

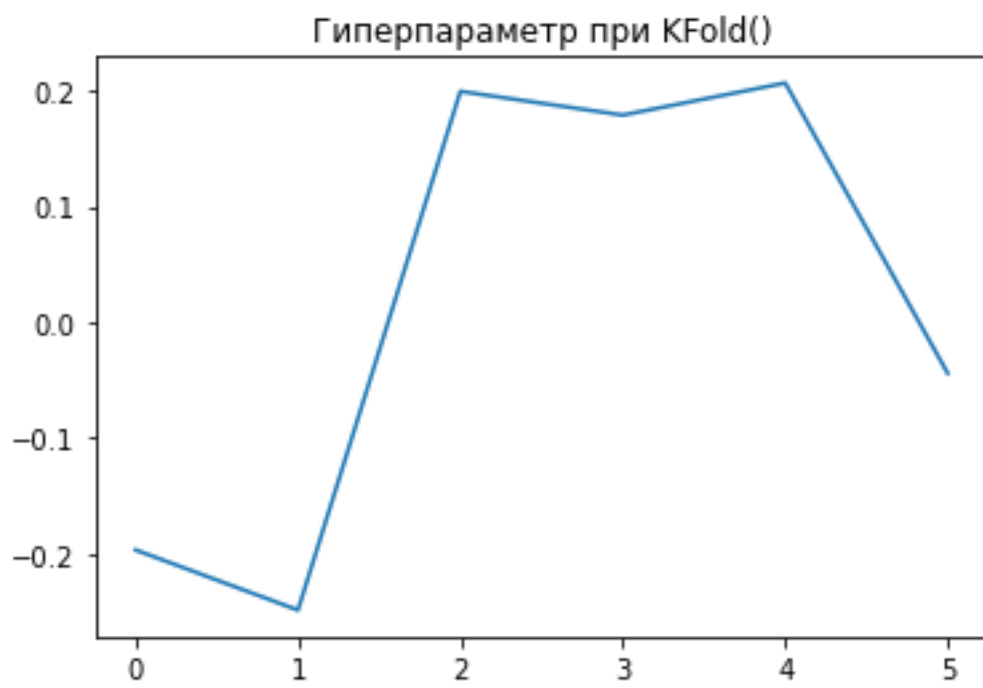
	ID	Warehouse_block	Mode_of_Shipment	Customer_rating	Cost_of_the_Product
7813	7814	4	0	1	240
10955	10956	4	0	5	125
905	906	4	2	4	147
5192	5193	0	1	4	233
235	236	4	2	2	134

8799 rows × 11 columns

```

KNN = KNeighborsRegressor(n_neighbors=10)
KNN.fit(trainX, trainY)
KNeighborsRegressor(n_neighbors=10)
testY.shape
(2200,)
KNN.predict(testX).shape
(2200,)
kf = KFold(n_splits=10)
scores = cross_val_score(KNeighborsRegressor(n_neighbors=5),
xArray, yArray, scoring='r2', cv=6)
scores
array([-0.19649137, -0.24845836,  0.19951572,  0.1790049 ,
 0.20675991,
        -0.04392033])
plt.plot(range(len(scores)), scores)
plt.title("Гиперпараметр при KFold()")
Text(0.5, 1.0, 'Гиперпараметр при KFold()')

```



Результаты(метрики)

Гиперпараметры

Модель