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, искусственный и системы управления»

Кафедра «Системы обработки информации и управления»

Отчет по Лабораторной работе №3 «*Подготовка обучающей и тестовой выборки, кросс-валидация и подбор гиперпараметров на примере метода ближайших соседей.*» по дисциплине «Технология машинного обучения»

Выполнил:
студент группы ИУ5-61Б
И.А. Абуховский

Проверил:
Ю.Е. Гапанюк

2023 г.

In [100…

```
from operator import itemgetter
import matplotlib.pyplot as plt
import matplotlib.ticker as ticker
import numpy as np
import math
import pandas as pd
```

In [101…

```
from enum import Enum
class PredictionType(Enum):
    CLASSIFICATION = 1
    REGRESSION = 2
```

In [102…

```
data = pd.read_csv('KNNAlgorithmDataset.csv')
```

In [103…

```
data['diagnosis'] = data['diagnosis'].replace('B',0)
data['diagnosis'] = data['diagnosis'].replace('M',1)
data['diagnosis']
data = data.fillna(0)
```

In [104…

```
from sklearn.model_selection import train_test_split
```

In [105…

```
x_train, x_test, y_train, y_test = train_test_split(data,
                                                    data['diagnosis'],
                                                    random_state=2) # random_state - для воспроизводимости
```

In [106…

```
y_train
```

Out[106]:

```
489    1
194    1
188    0
285    0
362    0
      ..
299    0
534    0
493    0
527    0
168    1
Name: diagnosis, Length: 426, dtype: int64
```

In [107…

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=5)
```

In [108…

```
knn_model = knn.fit(x_train, y_train)
```

In [109…

```
knn_predictions = knn.predict(x_test)
knn_predictions
```

```
D:\anaconda\lib\site-packages\sklearn\neighbors\_classification.py:228: FutureWarning: Unlike other reduction f
unctions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically preserves the axis it acts along.
In SciPy 1.11.0, this behavior will change: the default value of `keepdims` will become False, the `axis` over
which the statistic is taken will be eliminated, and the value None will no longer be accepted. Set `keepdims`
to True or False to avoid this warning.
  mode, _ = stats.mode(_y[neigh_ind, k], axis=1)
```

Out[109]:

```
array([0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0,
       1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0], dtype=int64)
```

In [110…

```
from sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_test, knn_predictions)
print(f'Accuracy: {accuracy}')
```

```
Accuracy: 0.7762237762237763
```

In [111…

```
from sklearn.ensemble import GradientBoostingRegressor
```

In [112…

```
### Тренируем
gbr = GradientBoostingRegressor()
gbr.fit(x_train,y_train)
```

Out[112]:

```
GradientBoostingRegressor()
```

In [113…

```
train_accuracy_score=gbr.score(x_train,y_train)
print(train_accuracy_score)

test_accuracy_score=gbr.score(x_test,y_test)
print(test_accuracy_score)
```

```
0.9999999992944921
0.99999999929258
```

In [114...

```
from sklearn import tree

clf = tree.DecisionTreeClassifier()
clf = clf.fit(x_train, y_train)
```

In [115...

```
test_accuracy_score=clf.score(x_test,y_test)
print(test_accuracy_score)
```

```
1.0
```

In [116...

```
train_accuracy_score=clf.score(x_test,y_test)
print(train_accuracy_score)
```

```
1.0
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js