

**Московский государственный технический
университет им. Н.Э. Баумана.**

Факультет «Информатика и управление»

Кафедра «Системы обработки информации и управления»

Курс «Теория машинного обучения»

Отчет по лабораторной работе №5

Выполнила:
студентка группы ИУ5-64
Бредня Елизавета

Подпись и дата:

Проверил:
преподаватель каф. ИУ5
Гапанюк Ю.Е.

Подпись и дата:

Москва, 2022 г.

Описание задания

1. Выберите набор данных (датасет) для решения задачи классификации или регрессии.
2. В случае необходимости проведите удаление или заполнение пропусков и кодирование категориальных признаков.
3. С использованием метода `train_test_split` разделите выборку на обучающую и тестовую.
4. Обучите следующие ансамблевые модели:
 - одну из моделей группы бэггинга (бэггинг или случайный лес или сверхслучайные деревья);
 - одну из моделей группы бустинга;
 - одну из моделей группы стекинга.
5. Оцените качество моделей с помощью одной из подходящих для задачи метрик. Сравните качество полученных моделей.

Текст программы и её результаты

```
▶ from sklearn.datasets import load_wine
import pandas as pd
from sklearn.ensemble import BaggingClassifier, AdaBoostClassifier
from sklearn.ensemble import StackingClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
import numpy as np
import matplotlib.pyplot as plt
import warnings
%matplotlib inline

warnings.filterwarnings("ignore")
wine = pd.DataFrame(load_wine().data, columns=load_wine().feature_names)
target = pd.DataFrame(load_wine().target, columns=[“type”])
wine.head()
```

▶

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavanoid
0	14.23		1.71	2.43		15.6	127.0	2.80
1	13.20		1.78	2.14		11.2	100.0	2.65
2	13.16		2.36	2.67		18.6	101.0	2.80
3	14.37		1.95	2.50		16.8	113.0	3.85
4	13.24		2.59	2.87		21.0	118.0	2.80

+ Код

+ Текст

Пропуски отсутствуют:

```
[ ] wine.isna().sum()

alcohol          0
malic_acid      0
ash              0
alcalinity_of_ash 0
magnesium        0
total_phenols    0
flavanoids       0
nonflavanoid_phenols 0
proanthocyanins 0
color_intensity   0
hue              0
od280/od315_of_diluted_wines 0
proline          0
dtype: int64
```

```
▶ X_train, X_test, y_train, y_test = train_test_split(wine, target, test_size=0.3, random_state=42)
```

Бэггинг

```
[ ] bc1 = BaggingClassifier(n_estimators=5, oob_score=True, random_state=10)
bc1.fit(X_train, y_train)
bc1_pred = bc1.predict(X_test)
accuracy_score(y_test, bc1_pred)
```

```
▶ 0.8703703703703703
```

Бустинг

```
[ ] ab1 = AdaBoostClassifier(n_estimators=5, algorithm='SAMME', random_state=10)
ab1.fit(X_train, y_train)
ab1_pred = ab1.predict(X_test)
accuracy_score(y_test, ab1_pred)

0.9074074074074074
```

Стекинг

```
▶ stacking = StackingClassifier(
    [
        ('lr', LogisticRegression()),
        ('dt', DecisionTreeClassifier()),
        ('kn', KNeighborsClassifier())
    ],
    LogisticRegression()
)
stacking.fit(X_train, y_train)

👤 StackingClassifier(estimators=[('lr', LogisticRegression()),
                                    ('dt', DecisionTreeClassifier()),
                                    ('kn', KNeighborsClassifier())],
                        final_estimator=LogisticRegression())

▶ st_pred = stacking.predict(X_test)
accuracy_score(y_test, st_pred)

0.9814814814814815
```