



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 и системы управления \_\_\_\_\_

КАФЕДРА \_\_\_\_\_ Системы обработки информации и управления \_\_\_\_\_

Отчет по лабораторной работе № 5  
«Ансамбли моделей машинного обучения. Часть 1.»  
по дисциплине «Технологии машинного обучения»

Студент ИУ5-65Б  
(Группа)

\_\_\_\_\_  
(Подпись, дата)

Д.А. Шиленок  
(И.О.Фамилия)

Преподаватель

\_\_\_\_\_  
(Подпись, дата)

Ю.Е. Гапанюк  
(И.О.Фамилия)

Москва

2025

**Цель лабораторной работы:** изучение ансамблей моделей машинного обучения.

### **Задание**

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:
  - две модели группы бэггинга (бэггинг или случайный лес или сверхслучайные деревья);
  - AdaBoost;
  - градиентный бустинг.
5. Оцените качество моделей с помощью одной из подходящих для задачи метрик. Сравните качество полученных моделей.

## Текст программы и результаты

# Лабораторная работа №5

## Ансамбли моделей машинного обучения. Часть 1.

Шиленок Даниил ИУ5-65Б

```
from sklearn.datasets import fetch_california_housing
from sklearn.model_selection import train_test_split

data = fetch_california_housing(as_frame=True)
X = data.data
y = data.target
df = data.frame

print(data.DESCR)
print(df.head())
print(df.describe())
```

.. \_california\_housing\_dataset:

California Housing dataset

-----

**\*\*Data Set Characteristics:\*\***

:Number of Instances: 20640

:Number of Attributes: 8 numeric, predictive attributes and the target

:Attribute Information:

- MedInc median income in block group
- HouseAge median house age in block group
- AveRooms average number of rooms per household
- AveBedrms average number of bedrooms per household
- Population block group population
- AveOccup average number of household members
- Latitude block group latitude
- Longitude block group longitude

:Missing Attribute Values: None

This dataset was obtained from the StatLib repository.

[https://www.dcc.fc.up.pt/~ltorgo/Regression/cal\\_housing.html](https://www.dcc.fc.up.pt/~ltorgo/Regression/cal_housing.html)

...

25%	2.429741	33.930000	-121.800000	1.196000
50%	2.818116	34.260000	-118.490000	1.797000
75%	3.282261	37.710000	-118.010000	2.647250
max	1243.333333	41.950000	-114.310000	5.000010

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
```

## Бэггинг

```
from sklearn.ensemble import BaggingRegressor
from sklearn.metrics import mean_squared_error

model = BaggingRegressor(random_state=1)
model.fit(X_train, y_train)
preds = model.predict(X_test)
mse = mean_squared_error(y_test, preds)
print(f"Бэггинг: MSE = {mse:.4f}")
```

## Случайный лес

```
from sklearn.ensemble import RandomForestRegressor

model = RandomForestRegressor(random_state=1)
model.fit(X_train, y_train)
preds = model.predict(X_test)
mse = mean_squared_error(y_test, preds)
print(f"Случайный лес: MSE = {mse:.4f}")
```

## AdaBoost

```
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import AdaBoostRegressor

model = AdaBoostRegressor(estimator=DecisionTreeRegressor(max_depth=10, random_state=1), n_estimators=200, learning_rate=0.05, random_state=1)
model.fit(X_train, y_train)
preds = model.predict(X_test)
mse = mean_squared_error(y_test, preds)
print(f"AdaBoost: MSE = {mse:.4f}")
```

## Градиентный бустинг

```
from sklearn.ensemble import GradientBoostingRegressor

model = GradientBoostingRegressor(random_state=1)
model.fit(X_train, y_train)
preds = model.predict(X_test)
mse = mean_squared_error(y_test, preds)
print(f"Градиентный бустинг: MSE = {mse:.4f}")
```

---

Бэггинг:  $MSE = 0.2682$

Случайный лес:  $MSE = 0.2524$

AdaBoost:  $MSE = 0.2695$

Градиентный бустинг:  $MSE = 0.2940$