

# Абуховский Иван Александрович, ИУ5-61Б

## РК2 ТМО, Вариант 1

### Задание

Для заданного набора данных (по варианту - sklearn iris) постройте модели классификации или регрессии (в зависимости от конкретной задачи, рассматриваемой в наборе данных). Для построения моделей используйте методы 1 и 2 (по группе - Линейная/логистическая регрессия и Случайный лес). Оцените качество моделей на основе подходящих метрик качества (не менее двух метрик).

### Выполнение

#### Загрузим датасет и необходимые библиотеки

```
In [1]: from sklearn.datasets import load_iris
iris = load_iris()
x = iris.data
y = iris.target

In [2]: from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error, r2_score
```

#### Масштабируем его с помощью StandardScaler

```
In [3]: from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
x = scaler.fit_transform(x)
```

#### Разделим датасет на обучающую и тестовую выборки

```
In [4]: x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3)
```

#### Создадим и обучим модели Случайного леса и Линейной регрессии

```
In [5]: rand_for = RandomForestClassifier(n_estimators = 100)
rand_for.fit(x_train, y_train)
rand_for_pred = rand_for.predict(x_test)

lin_regr = LinearRegression()
lin_regr.fit(x_train, y_train)
lin_regr_pred = lin_regr.predict(x_test)
```

#### Сравним метрики качества средней абсолютной ошибки и коэффициент детерминации

```
In [6]: lin_regr_mae = mean_absolute_error(y_test, lin_regr_pred)
rand_for_mae = mean_absolute_error(y_test, rand_for_pred)

lin_regr_r2 = r2_score(y_test, lin_regr_pred)
rand_for_r2 = r2_score(y_test, rand_for_pred)

In [8]: print("Линейная регрессия")
print("MAE : ", lin_regr_mae)
print("R2 : ", lin_regr_r2)

print("Случайный лес")
print("MAE : ", rand_for_mae)
print("R2 : ", rand_for_r2)
```

Линейная регрессия  
MAE : 0.17848925260935825  
R2 : 0.9073203947869973  
Случайный лес  
MAE : 0.06666666666666667  
R2 : 0.8897058823529411

Делаем вывод, что линейная регрессия показывает лучшие показатели коэффициента детерминации, а случайный лес - средней абсолютной ошибки