Национальный исследовательский Томский государственный университет

Отчёт по лабораторной работе

«Регрессионные модели»

Вариант 9

Выполнил студент гр. 932001

Андрюшина М. А.

Преподаватель

Аксёнов С. В.

2023 г.

**Работа с данными**

Для работы с бинарными классификаторами использовался набор данных Bike Sharing Dataset (Аренда велосипедов).

Набор данных не содержал пропущенных значений. Для анализа и обучения были использованы все признаки из набора данных, кроме целевого, registered и casual, так как они являются производными от таргета и не входят в основной набор данных. Для данных в столбцах season, hr, weathersit применили функцию get\_dummies. Данные были преобразованы через MinMaxScaler.

Датасет был разделен на 2 части: тренировочную (70%), тестовую (30%).

**Linear Regression**

При создании модели LinearRegression использовались следующие гиперпараметры: n\_jobs = -1.

**Метрики качества:**

R2-score: 0.669

MSE: 10725.116

MAE: 75.819

**Polynomial Features**

Для PolynomialFeatures возьмём degree, равное 2, так как модель обучается на 42 признаках и при большем значении degree может не хватить вычислительных мощностей.

При создании модели LinearRegression с PolynomialFeatures использовались следующие гиперпараметры: fit\_intercept=False, n\_jobs=-1, positive=True.

**Метрики качества:**

R2-score: 0.813

MSE: 6067.548

MAE: 53.043

**Random Forest**

При создании модели RandomForestRegressor использовались следующие гиперпараметры: max\_depth = 12, max\_features = 30, n\_estimators = 300, n\_jobs = -1, random\_state = 77.

При анализе значимости признаков мы выяснили, что наибольшее влияние на предсказание таргета оказывают следующие признаки: atemp, hr\_17, hr\_18, hum, yr.

**Метрики качества:**

R2-score: 0.831

MSE: 5475.327

MAE: 53.644.

**Gradient Boosting**

При создании модели GradientBoostingRegressor использовались следующие гиперпараметры: learning\_rate=0.5, max\_features=40, n\_estimators=400, random\_state=77.

При анализе значимости признаков мы выяснили, что наибольшее влияние на предсказание таргета оказывают следующие признаки: atemp, workingday, hr\_17, yr, hr\_18.

**Метрики качества:**

R2-score: 0.931

MSE: 2248.334

MAE: 31.384.

**Вывод**

Лучше всех показала себя модель GradientBoostingRegressor с R2-score = 0.931, MSE = 2248.334, MAE = 31.384.

В моделях RandomForestRegressor и GradientBoostingRegressor были схожие самые значимые признаки (atemp, hr\_17, yr, hr\_18), но они имели разные веса в каждой модели.

**Программный код**

Программный код доступен по ссылке:

<https://github.com/AnMari24/MachineLearning/tree/main/Lab3>