

# Отчёт по экспериментам

**Студент:** Клевцов Арсений Иванович

**Логин:** student 2

**MLflow experiment:** homework\_Klevtsov

## Серия экспериментов № 1

### Влияние размера обучающей выборки

**Гипотеза:** увеличение размера обучающей выборки приводит к росту качества модели.

**Исследуемый параметр:** train\_size

**Сетка изменения:** 0 . 1 → 0 . 5 → 1 . 0 (доля от данной обучающей выборки)

**Модели:** LogisticRegression, RandomForestClassifier, DecisionTreeClassifier

### Результаты

Согласно графикам и таблицам в MLflow UI, увеличение train\_size не привело к существенному росту качества ни для одной из моделей.

### Вывод

Обратная гипотеза не подтвердилась: в данной задаче даже десятикратное уменьшение размера тренировочной выборки значимо не влияет на качество обученных моделей.

---

## Серия экспериментов № 2

### Влияние параметра регуляризации C (LogisticRegression)

**Гипотеза:** увеличение параметра C к ухудшению качества логистической регрессии.

**Исследуемый параметр:** C

**Сетка изменения:** 0 . 0 1 → 0 . 1 → 1 → 1 0 → 1 0 0

train\_size = 1 . 0 (зафиксирован)

### Результаты

По данным MLflow UI наблюдается рост качества модели при увеличении параметра C.

### Вывод

Гипотеза снова не подтвердилась: с уменьшением силы регуляризации качество логистической регрессии растёт на тесте

---

# **Серия экспериментов № 3**

## **Влияние количества деревьев в RandomForest**

**Гипотеза:** увеличение числа деревьев приводит к улучшению качества на небольшой обучающей выборке.

**Исследуемый параметр:** n\_estimators

**Сетка изменения:** 2 → 10 → 50 → 100

train\_size = 0.1 (задано)

### **Результаты**

При увеличении количества деревьев наблюдается устойчивый рост качества модели по метрикам.

### **Вывод**

Обратная гипотеза опровергнута. Увеличение числа деревьев в ансамбле улучшает качество RandomForest даже при небольшой обучающей выборке.

---

## **Итоговые выводы**

По видимому задача достаточно простая, поэтому размер обучающей выборки особо не влияет на качество моделей. Та же логика относится и к регуляризации логрега. Но качество случайного леса действительно зависит от числа деревьев в нем, что вполне ожидаемо