

# Качество модели

## Задание параметров модели для воспроизводимости эксперимента

```
In model = DecisionTreeClassifier(random_state=12345)

# random_state – случайное число
```

## Вычисление доли правильных ответов

```
In from sklearn.metrics import accuracy_score

accuracy = accuracy_score(target, predictions)
```

## Указание параметра дерева решений - максимальной глубины дерева

```
In model = DecisionTreeClassifier(random_state=12345, max_depth=3)

model.fit(features, target)
```

## Сохранение обученной модели

```
In from joblib import dump

joblib.dump(model, 'model.joblib')

# первый аргумент – модель
# второй аргумент – путь к файлу
```

## Загрузка обученной модели

```
In from joblib import load

model = joblib.load('model.joblib')
```

# Словарь

### Тестовый набор данных, или тестовая выборка

набор данных, на котором будет проверяться качество обученной модели

### Метрики качества

способы оценки качества работы модели машинного обучения

### Доля правильных ответов (accuracy)

отношение числа правильных ответов к размеру тестовой выборки

### Точность (precision)

метрика качества, которая показывает, какая доля объектов, для которых модель предсказала ответ «1», действительно имеют ответ «1»

### Полнота (recall)

метрика качества, которая выявляет, какую часть объектов, имеющих ответ «1», выделила модель

### Проверка на вменяемость (sanity check)

сравнение модели со случайной для оценки адекватности модели

### Переобучение

проблема, когда метрика качества модели даёт хорошие показатели на тренировочной выборке, но плохие на тестовой

### Недообучение

проблема, когда метрика качества модели даёт плохие показатели на тренировочной и на тестовой выборке