

Задачи машинного обучения в бизнесе

Импорт класса классификатора

```
In from sklearn.tree import DecisionTreeClassifier
# создание объекта классификатора
model = DecisionTreeClassifier()
# обучение модели
model.fit(X, y)
# получение предсказаний
predictions = model.predict(X)
```

Разделение датафрейма на целевую переменную и обучающие признаки

```
In y = data['target']
X = data.drop(['target'], axis = 1)
```

Разделение выборки на тренировочную и тестовую (валидационную) выборку

```
In # test_size – процент от исходного датасета, который выделяется для тестовой выборки
# random_state – параметр для воспроизведения результата

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
```

Словарь

Машинное обучение

это процесс поиска машиной взаимосвязи между величинами на примере разных объектов

Модель

система взаимосвязей между признаками и целевой переменной или между наблюдениями, которая максимально близко отражает действительность

Целевая переменная

та, которую предсказывают, или прогнозируют по определённым **признакам, или фичам**

Наблюдения

уже известные примеры, для которых есть значения признаков и целевой переменной

Обучающая (тренировочная) выборка

данные, на которых обучается модель.

Валидационная выборка

данные, на которых проверяется качество модели во время обучения

Тестовая выборка

данные, на которых модель проверяется после обучения.

Обучение с учителем

построение взаимосвязи между признаками и целевыми переменными на основе известных размеченных данных

Классификация

задача обучения с учителем, в которой целевая переменная принимает значения из ограниченного числа возможных значений

Регрессия

задача обучения с учителем, в которой целевая переменная является непрерывной величиной

Обучение без учителя

построение взаимосвязи между объектами без известных целевых переменных

Кластеризация

задача обучения без учителя, в которой объекты разделяются на группы на основе взаимосвязи между ними

Снижение размерности

задача обучения без учителя, в которой признаковое описание объектов изменяется с уменьшением числа признаков