



онлайн
университет

Формирование портрета потенциального клиента

Хакатон
Школы программирования

Аннотация

Чтобы сделать потенциальную продажу реальной, необходимо четко знать потребности клиента. Для этих целей компании часто используют **модели машинного обучения**.

В системах машинного обучения или же системах нейросетей существуют входы и выходы. То, что подаётся на входы, принято называть **признаками** (англ. features). Когда признаки подаются на входы системы машинного обучения, эта система пытается найти совпадения, заметить закономерность между признаками.

На выходе генерируется результат этой работы. Этот результат принято называть **меткой** (англ. label), поскольку у выходов есть некая пометка, выданная им системой, т. е. предположение (прогноз) о том, в какую категорию попадает выход после классификации.

Процесс обучения модели — это подача данных для нейросети, которая в результате должна вывести определённые шаблоны для данных. В процессе обучения модели с учителем на вход подаются признаки и метки, а при прогнозировании на вход классификатора подаются только признаки.

Принимаемые сетью данные делятся на две группы: **набор данных для обучения и набор для тестирования**. Не стоит проверять сеть на том же наборе данных, на которых она обучалась, т. к. модель уже будет «заточена» под этот набор.



Методы машинного обучения

Задача **классификации** машинного обучения — это любая задача, где нужно определить тип объекта из двух и более существующих классов.

В задаче хакатона классификация сводится к определению, является ли данный человек потенциальным клиентом или нет.

Основные методы машинного обучения:

- [Метод k-ближайших соседей \(K-Nearest Neighbors\);](#)
- [Метод опорных векторов \(Support Vector Machines\);](#)
- [Классификатор дерева решений \(Decision Tree Classifier\) / Случайный лес \(Random Forests\);](#)
- [Наивный байесовский метод \(Naive Bayes\);](#)
- [Линейный дискриминантный анализ \(Linear Discriminant Analysis\);](#)
- [Логистическая регрессия \(Logistic Regression\);](#)

По мере накопления опыта вам будет проще выбирать подходящий тип классификатора. Однако хорошей практикой является реализация нескольких подходящих классификаторов и выбор наиболее оптимального и производительного.

Процесс содержит в себе следующие этапы: подготовка данных, создание обучающих наборов, создание классификатора, обучение классификатора, составление прогнозов, оценка производительности классификатора и настройка параметров.

Вероятнее всего, вам нужно будет «корректировать» параметры классификатора, пока вы не достигнете желаемой точности (т. к. маловероятно, что классификатор будет соответствовать всем вашим требованиям с первого же запуска).

Пример работы машинного обучения от обработки данных и до оценки можно посмотреть [здесь](#).

Полезные ссылки

Статьи, в которых можно почерпнуть идеи для решения задачи:

- [Анализ данных клиентов и прогнозирование оттока пользователей на Python](#)
- [Как снизить отток клиентов с помощью алгоритмов машинного обучения](#)

Ходовые **примеры** алгоритмов для решения задач регрессии / классификации:

- [Линейная Регрессия](#)
- [Полиномиальная Регрессия](#)
- [Дерево решений \(классификация\)](#)
- [K-NN \(классификация\)](#)
- [Множественная линейная Регрессия](#)

Желаем вам успехов в решении задачи! 😊