

## Разработка курса «Основы машинного обучения» для студентов

Ищук М. А.

*mai.ima@yandex.ru*

*Саратовский государственный университет имени Н.Г. Чернышевского»*

Данная статья описывает важность машинного обучения в настоящее время, а также обзор учебного курса по машинному обучению.

**Ключевые слова:** машинное обучение, большие данные, python.

С развитием информационных технологий, в частности, социальных сетей, интернет-магазинов и других онлайн-сервисов, в современном мире возрастает потребность в анализе и обработке больших объемов информации, извлечении из нее нужных данных. В связи с этим, возникает желание автоматизировать различные способы и методы решения однотипных задач. Поэтому, сейчас так возрастает популярность машинного обучения.

Машинное обучение является одним из разделов искусственного интеллекта. Основной принцип заключается в том, что компьютер получает данные и пытается выявить закономерности, чтобы использовать эти закономерности, например, для прогнозирования характеристик новых данных. Системы машинного обучения используются для решения таких задач как распознавание речи, распознавание объектов, перевод, построение рекомендательных систем, прогнозов и многих других.

Этапы разработки алгоритмов машинного обучения:

1. Сбор данных. Для того, чтобы обучить алгоритм чему-то, нужно достаточное количество данных, на котором он будет обучаться. Собирать их можно с помощью парсера сайтов, используя RSS или API. Можно также воспользоваться готовыми учебными наборами данных.

2. Подготовка входных данных. Для того, чтобы собранную информацию можно было обрабатывать, сначала нужно привести ее в пригодный для использования вид. Например, представить данные в виде таблицы.

3. Анализ входных данных. Перед тем, как передавать данные на обучение, следует их проанализировать, проследить некоторые закономерности.

4. Обучение алгоритма. На данном этапе собранные и подготовленные данные передаются компьютеру, и он проходит обучение.

5. Тестирование алгоритма. Информация, полученная на предыдущем этапе, проверяется на то, насколько хорошо алгоритм справился со своей задачей. В случае обучения с учителем, имеются некоторые данные, которые можно использовать для проверки. В случае обучения без учителя, возможно использовать другие методики тестирования качества алгоритма. Если результаты неудовлетворительны, можно вернуться к шагу 4, что-то изменить и попробовать снова.

6. Использование результата. На этом шаге алгоритм преобразуется в работающую программу, если его работа корректна и удовлетворяет всем

требованиям.

При изучении машинного обучения необходимо использование какого-либо языка программирования. Наиболее приспособленный для этого язык – Python.

Если данный язык не изучался ранее, то имеет смысл начинать обучение с него, чтобы потом все рассмотренные примеры и методы иллюстрировать наглядно с использованием программ и библиотек данного языка.

При создании курса необходимо учитывать теоретическую и практическую подготовленность обучающихся. Создание курса, который подойдет для любого обучающегося, начинающего изучать машинное обучение, заключается в том, чтобы предоставить всю информацию, начиная с самых основ. Таким образом, начиная с обучения языку программирования Python, к концу такого курса, слушатель будет уметь решать практические задачи машинного обучения.

Рассмотрим самостоятельно разрабатываемый курс по машинному обучению, размещенный на портале school.sgu.ru.

Данный курс может использоваться как при дистанционном обучении, так и для поддержки очного обучения.

Цель курса – изучить теоретические основы машинного обучения и научиться применять теоретические знания при решении практических задач, связанных с машинным обучением, используя язык программирования Python.

Данный курс может быть использован в ВУЗе в рамках дисциплин и практик по данной теме, а отдельные его элементы – в рамках спецкурсов по информатике для школьников старших классов. Кроме того, студенты и школьники могут использовать его для самостоятельного обучения.

Освоивший данный курс, студент будет:

– знать основные алгоритмы машинного обучения, такие как: регрессия, алгоритмы кластеризации, деревья решений, метод опорных векторов, коллаборативная фильтрация, наивный байесовский метод.

– уметь решать практические задачи, связанные с машинным обучением;

– владеть: навыками программирования на языке Python; навыками применения инструментов языка Python для решения задач машинного обучения.

Таблица. Содержание курса

№ пп	Наименование темы	Всего, час.	В том числе:			Формы текущего контроля успеваемости, форма промежуточной аттестации
			Теоретические занятия	Практические занятия	Самостоятельная	
1	2	3	4	5	6	7
1.	Изучение языка программирования Python	16	6	6	18	Практические задания, тест контроля знаний

2.	Машинное обучение	28	15	15	27	Практические задания, тест контроля знаний
3.	Практические задачи машинного обучения	28	15	15	27	Практические задания, тест контроля знаний
Итого:		144	36	36	72	

Первый раздел «Изучение языка программирования Python» включает в себя основную информацию о языке программирования Python, а также практические задания, которые помогают закрепить теорию. В качестве практических заданий присутствуют: тесты, проверяющие теоретические знания; задачи, позволяющие применить полученные знания на практике.

Второй раздел «Машинное обучение» посвящен основам теории о машинном обучении. Он включает в себя теорию о больших данных, так как эта тема непосредственно связана с машинным обучением и является его частью. Помимо этого, раздел содержит в себе теоретическое описание алгоритмов, и задачи, связанные с написанием и использованием алгоритмов.

Третий раздел «Практические задачи машинного обучения» включает в себя конкретные задачи машинного обучения, рассмотренные на примерах и способы решения этих задач.

Ниже представлен полный план курса.

1. Изучение языка программирования Python. Возможности языка. Установка Python и библиотек, используемых в машинном обучении. Операторы и циклы. Встроенные функции языка программирования Python. Числа в Python. Строки в Python. Списки, кортежи, словари. Множества. Функции. Исключения и их обработка. Работа с файлами. Библиотеки в Python.

2. Машинное обучение. Введение. Большие данные. Регрессия. Алгоритмы кластеризации. Деревья решений. Метод опорных векторов. Коллаборативная фильтрация. Наивный байесовский метод.

3. Практические задачи машинного обучения. Задачи кластеризации. Задачи классификации. Рекомендательные системы. Компьютерное зрение.

Курс расположен по адресу <http://school.sgu.ru/course/view.php?id=200>.

На данный момент разработан первый раздел курса, посвященный изучению языка программирования Python. Предполагается внедрение курса в учебный процесс на факультете КНиИТ СГУ.

#### Список литературы

- [1] Коэльо Л.П., Ричарт В. Построение систем машинного обучения на языке Python. – 2-е издание, пер. с англ. Слинкин А.А. - М.: ДМК Пресс, 2016. - 302 с.: ил.
- [2] Witten I.H., Frank E. Data mining: practical machine learning tools and techniques. – 2nd ed. p. cm. – (Morgan Kaufmann series in data management systems), 2005.
- [3] D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young. Machine Learning: The High-Interest Credit Card of Technical

Debt, 2014 [4] Березина Л. Ю. Графы и их применение : пособие для учителей. М : Просвещение, 1979. 143 с.