Как начать изучение Machine Learning

Екатерина Антакова,
инженер в Intel

О чем рассказ

1. Кто и что изучает: обо мне и о machine learning
2. Пошаговый план
3. Курсы и другие способы обучения
4. Советы и выводы

Мои исходные данные

- Образование – математика и информатика
- 9 лет в разработке программного обеспечения
- 2 года декрета:
  - потеря квалификации
  - низкий доход
  - нет свободного времени
  - кризис поисков смысла и призвания
- Интерес к популярной теме machine learning & big data

Что такое Big Data и Machine Learning?

- **Big data** - данные, слишком большие для обычных программ (баз данных)
  - В широком смысле – это сбор, хранение, анализ, поиск, передача, визуализация, обновление и безопасность данных
  - В узком смысле – анализ и предсказания по большим наборам данных
- **Machine Learning** – методы обработки, анализа и стуктурирования больших данных, которые позволяют компьютеру обучаться без задания конкретной программы выдачи ответов.

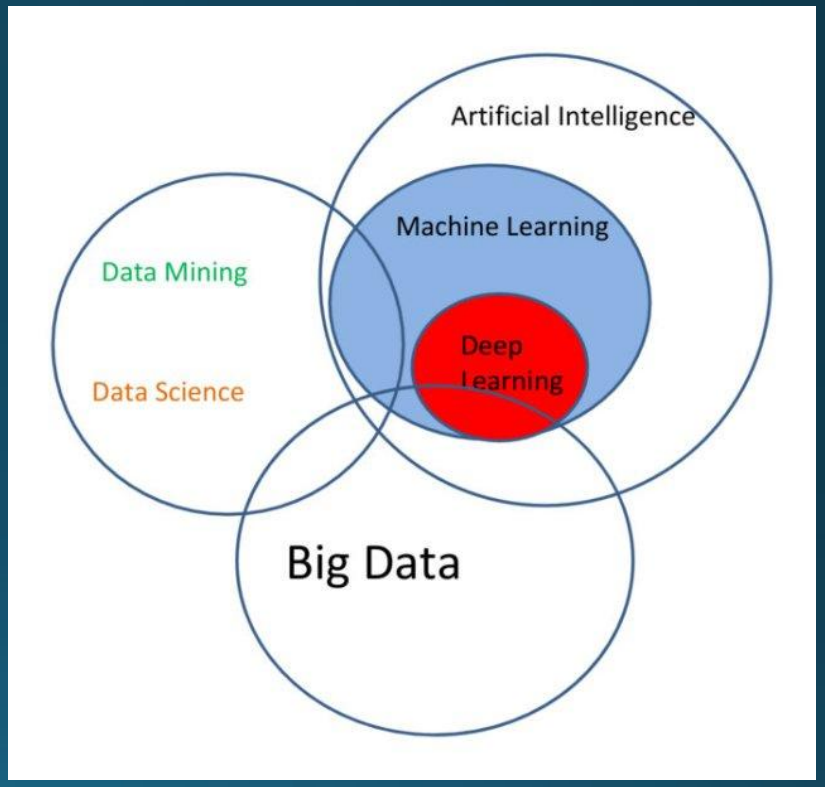


Диаграмма с www.linkedin.com
из статьи Srikanth Victory

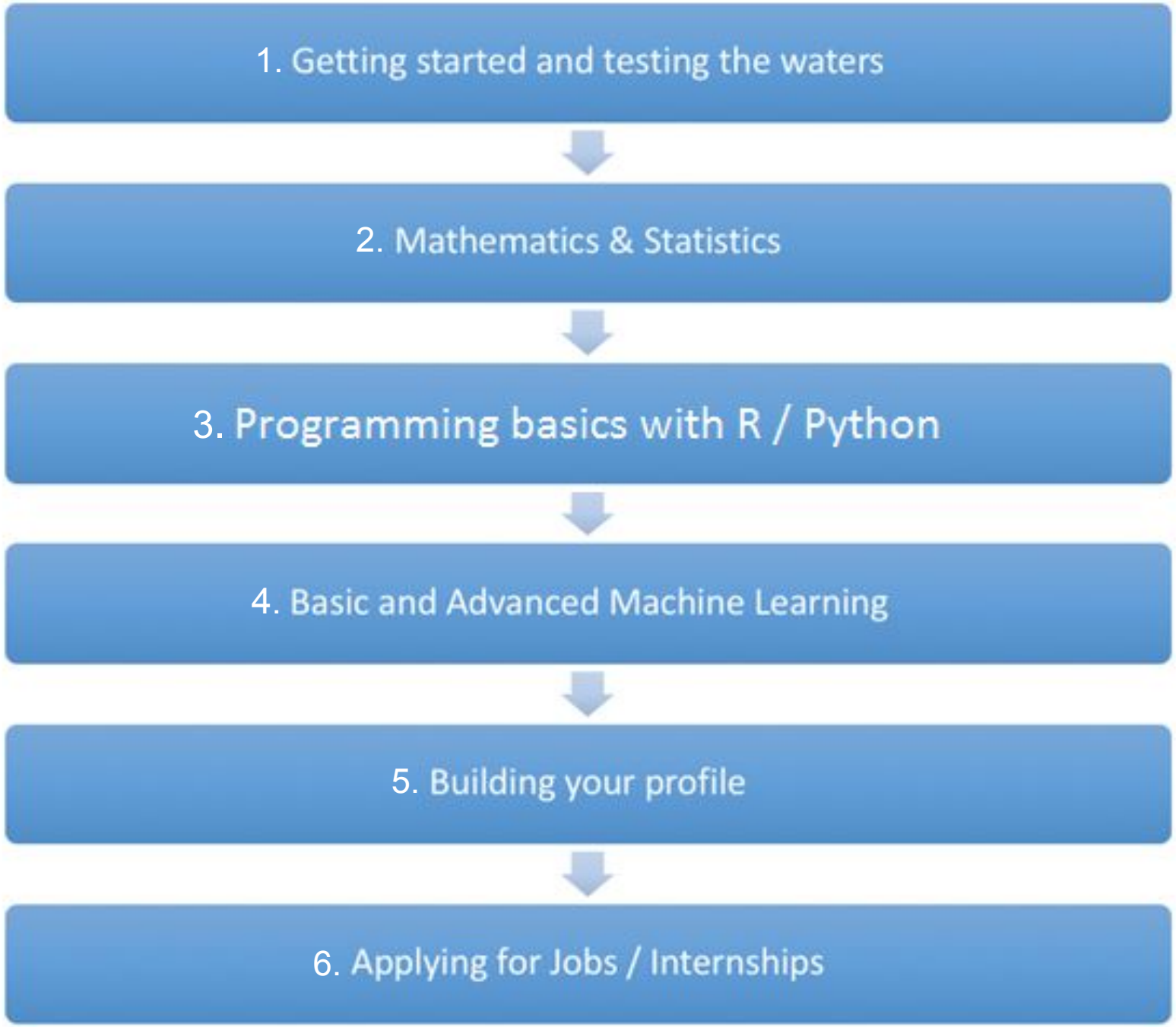
Зачем это изучать?

План для начинающих

+ Английский язык

+ Навыки поиска в интернете

+ Свободное время



Картинка с analyticsvidhya.com

Coursera: Machine Learning by Andrew Ng

+ Подробные, чёткие объяснения
+ Средний уровень детализации
+ Много примеров и диаграмм
+ Большой круг задач и алгоритмов
+ Преподаватель - всемирно известный учёный, основатель платформы Coursera
+ Есть русские субтитры
+ №2 по популярности и №4 по рейтингу среди всех курсов Coursera

Возможные трудности:
- Много программирования
- Аудио, слайды, задания и материалы на английском
- Есть неточности, поправки нужно искать на форумах

Хороший курс для начала, если есть база в английском и программировании

«Excellent machine learning course for beginners like me. Professor Andrew Ng has explained details of machine learning in very systematic, simple and clear terms.»

Позволит разобраться в базовых алгоритмах и концепциях

Специализация «Машинное обучение и анализ данных» на Coursera

+ Русский язык
+ Богатая учебная программа
+ Создатели курса признанные специалисты в обработке данных
+ Активное сообщество в Slack
+ Много практики на Python

- Много математического анализа и теории вероятностей
- Не все темы одинаково подробно изложены
- Часть заданий закрыта при бесплатном просмотре, полная стоимость ~15 000 р.
- Специализация из 6 курсов – это долго**, но круто, если пройдёте**

«Много практики. Нагрузка довольно высокая, особенно для работающего человека. Некоторые моменты в лекциях освещаются очень поверхностно, а порой на слушателя просто вываливаются большие формулы, а пояснение дано к ним в двух словах. В этом плане классический ml class с Andrew Ng мне понравился больше, т.к. теория там дана была более подробно, но зато было меньше практики (а тут плюс текущему курсу).»

Поможет найти первую работу, связанную с данными

Udemi Data Science A-Z™: Real-Life Data Science Exercises Included

+ Примеры на Python и R
+ Много задач из реального бизнеса
+ Большой обзор методов анализа данных
+ Различные представления данных и программы для их обработки
+ Хороший курс для новичка

- Немного теории (математики, алгоритмов)
- Немного и практических заданий
- Цена = 10-15$

This course is the clear winner in terms of breadth and depth of coverage of the data science process of the 20+ courses I qualified.

Well in terms of flow of topics and covering the fundamentals. It does add great value to beginners.

Помогает сориентироваться в методах и областях анализа данных

Ресурсы и новости на русском

- Видео с конкурсов Avito contest. Сложно, но интересно
- http://ods.ai/ – Slack-канал про анализ данных в России. Встречи, тренинги, совместное решение задач, помощь коллег, вакансии
- Открытый курс по машинному обучению на Хабрахабре. Для продвинутых - много математики и программирования
- https://habrahabr.ru/ – самый полезный ресурс по IT
- https://www.it52.info/ – курсы и встречи программистов и тестировщиков в Нижнем Новгороде. Telegram-канал @it52info

Как ещё учиться

- Чемпионаты по анализу данных на платформе Kaggle
- Брать известные наборы данных, анализировать, описывать на своём сайте или блоге, выкладывать модели на github
- Использовать анализ данных в повседневной работе
- Пройти программу типа DataQuest (~1800 р. / мес)
- Читать бесплатный e-mail курс по основам машинного обучения (на английском)
- Проходить курсы Online masters degree in data science (на английском)

Советы

- Мотивация возникает только в процессе деятельности при правильной наргузке. Трудно начать, но надо.

  Оптимальная нагрузка = умеренно сложно + интересно
- Английский язык – важен. Для этой области придётся его знать и улучшать.
- Не жалеть себя, не опираться на мужчин (или родителей, друзей) в серьёзных делах. Но трудиться сообща, помогать взаимно
- Найти единомышленников и держать связь: блоги, чаты, обеденная компания на работе, участники одного курса
- Писать код

Пишите на почту
kate@antakova.ru