

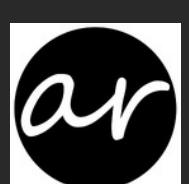


# DS/AI Self-Starter Handbook

BUILD YOUR OWN ROADMAP

---

Ankit Rathi



From a time around when DS/AI field started picking up, every other day I get at least 8–10 messages from DS/AI starters & enthusiasts on ‘How can I get into DS/AI field?’. Over a while, I have improvised my response based on the follow-up questions they ask like:

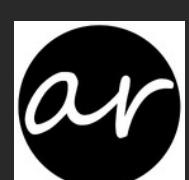
1. What is the difference between DS, ML, DL, AI, DM?
2. What are the roles in DS/AI, who does what?
3. What concepts, processes & tools they need to learn?
4. Which books, courses, etc they need to refer to?
5. How to build a DS/AI portfolio?
6. How to write a resume for DS/AI?
7. How to build a helpful network?
8. How to search for the job?
9. How to prepare for the interview?
10. How to stay up to date in this still-evolving field?

You can notice that these questions are not conceptual ones and there is no dedicated material to address these roadblocks. I thought why not to build a framework or a road-map for DS/AI starters and enthusiasts so that I need not to answer the same type of questions again and again. And that is when I started documenting what a starter or enthusiast need to do step by step in order to reach a level when he is ready to tackle any challenge thrown to him. My answer to the above questions in a structured way to help DS/AI starters & enthusiasts is this book. This book covers the framework to launch your DS/AI career in 8 chapters.



Ankit Rathi provides unique combination of Data Engineering (DB/ETL/DWH/BI)/Architecture (Data Management & Governance) & Data Science (ML/DL/AI) with more than a decade of demonstrated history of working in IT industry using Data & Analytics. His interest lies primarily in building end to end DS/AI applications/products following best practices of Data Engineering and Architecture.

In his free time, he blogs about various topics on DS/AI field & tries to simplify it for starters & enthusiasts.



# DS/AI Self-Starter Handbook

Build Your Own Roadmap

---

Ankit Rathi



[ankitrathi.com](http://ankitrathi.com)

[ankitratthi.com/youtube](http://ankitratthi.com/youtube)

*To my wife, Divya, who's always accepted me the way I am and supported my hustle, drive & ambition.*

*To my children, Aarsh & Driti, who are the reason to wake up every morning and work as hard as I can.*

[ankitratthi.com/youtube](http://ankitratthi.com/youtube)

DS/AI Self-Starter handbook is a great resource for aspirants starting in the space of Data Science. It covers approach and useful resources that can help in your learning journey and written by one who himself is an Data Science practitioner. I recommend this to anyone who are aspiring to get into Data Science and are looking for insights on how and where to get started.

**Srivatsan Srinivasan**

Chief Data Scientist (Cognizant)

Wow, this is very impressive! It has taken some time to review, but WOW!

I should have had you as a co-author next time!!!

**T. Scott Clendaniel**

Chief Data Scientist (Legg Mason)

To be great data scientist you should emphasis on skillset and mindset. Where a lot of book that give you skill set, this is the first book I read that dedicating to shape data scientist mindset.

**Nabih Ibrahim Bawazir**

Data Science Head (Datanest)

Extremely laudable & heroic attempt to put all your thoughts and experience together to help people.

**Sumit Pal**

Big Data Architect (Qcentive)

Ankit has done a great job summarizing what is possibly one of the toughest and most frequently asked questions, "How to get started with data science?". Packed with information, this book will definitely be helpful for people from both academia and industry looking to get started on their own Data Science and AI journey.

**Dipanjan Sarkar**

Data Scientist (Rad Hat)

I think it is a brilliant book for starting Career in Data Science as New Entrants to Data Science often deviate from Path to reach End Goal and this Book tries to solve that Problem in a easy way. I would really like to Congratulate Ankit for Providing Data Science Career Steps in this useful manner.

**Yatin Bhatia**

Data Scientist (RxLogix)

An indispensable guide and a valuable resource for anyone seeking to enter the field of Data Science. Replete with great advice directly from the author's personal experience.

**Parul Pandey**

Data Science Evangelist (H2O.ai)

This book kicks you into the right direction definitely worth reading for the beginners trying to break into DS/AI.

**Avik Jain**

Machine Learning Intern (EMA Solutions)

If you are one among people struggling to identify the right book for data science, this book would probably help to understand where to start, how to prepare, how to develop the habit of continuous learning.

**Vishnu Durgha Prasaad**

Data Science Practitioner

[ankitratthi.com/youtube](http://ankitratthi.com/youtube)

---

## About the Author



Ankit Rathi is currently working as a Lead Architect-DS/AI at SITA aero. He is a Data Science (ML/DL/AI) practitioner with more than a decade of demonstrated history of working in IT industry using Data & Analytics. His interest lies primarily in the theory & application of artificial intelligence, particularly in developing business applications for machine learning and deep learning. Ankit's work at SITA aero has revolved around designing FlightPredictor product & building the CoE capability. During his tenure as a Principal Consultant at Genpact HCM, Ankit architected and deployed machine learning pipelines for various clients across different industries like Insurance, F&A. He was previously a Tech Lead at RBS IDC where he designed and developed various data intensive applications in AML & Mortgages area. Ankit is a well-known author for various publications (Towards Data Science, Analytics Vidhya etc) on Medium where he actively contributes by writing blog-posts on concepts & latest trends in Data Science. His blog-series on 'Probability & Statistics for Data Science' has been well received by Data Science community in 2018. He is followed by around 30K data science practitioners & enthusiasts on LinkedIn.

U0.1: Webpage: <https://www.ankitrathi.com/>

[ankitratthi.com/youtube](http://ankitratthi.com/youtube)

# Table of Contents

<b>About the Author.....</b>	<b>9</b>
<b>Preface .....</b>	<b>13</b>
<b>1 Understanding the Big Picture .....</b>	<b>19</b>
<b>1.1 Navigating the landscape .....</b>	<b>22</b>
<b>1.2 Working on building blocks .....</b>	<b>23</b>
<b>1.3 Utilizing the resources .....</b>	<b>24</b>
<b>1.4 Building your portfolio.....</b>	<b>25</b>
<b>1.5 Networking &amp; landing the job .....</b>	<b>25</b>
<b>1.6 Making Career Future-Proof.....</b>	<b>26</b>
<b>1.7 Putting it all together.....</b>	<b>27</b>
<b>2 Navigating the Landscape .....</b>	<b>31</b>
<b>2.1 Multi-disciplinary field.....</b>	<b>33</b>
<b>2.2 Various Terminologies .....</b>	<b>35</b>
<b>2.3 Interconnection.....</b>	<b>37</b>
<b>2.4 DS/AI Roles.....</b>	<b>38</b>
<b>2.5 Academics Vs Industry .....</b>	<b>41</b>
<b>3 Working on Building Blocks .....</b>	<b>47</b>
<b>3.1 Concepts to Learn .....</b>	<b>49</b>
<b>3.2 Process to Follow .....</b>	<b>54</b>
<b>3.3 Tools to Master .....</b>	<b>57</b>
<b>4 Utilizing the Resources .....</b>	<b>61</b>
<b>4.1 Books to refer.....</b>	<b>63</b>

<b>4.2 Courses to Attend .....</b>	<b>67</b>
<b>4.3 Blogs to follow .....</b>	<b>70</b>
<b>4.4 Podcasts to listen.....</b>	<b>74</b>
<b>4.5 YouTube Channels .....</b>	<b>77</b>
<b>5 Building your Portfolio .....</b>	<b>87</b>
<b>5.1 Work on Public data-sets.....</b>	<b>88</b>
<b>5.2 Participate in Hackathons.....</b>	<b>95</b>
<b>5.3 Publish on Git-hub .....</b>	<b>99</b>
<b>5.4 Write a Blog.....</b>	<b>101</b>
<b>6 Networking &amp; Landing the Job .....</b>	<b>105</b>
<b>6.1 Network on LinkedIn .....</b>	<b>106</b>
<b>6.2 Prepare the CV .....</b>	<b>110</b>
<b>6.4 Crack the interview.....</b>	<b>116</b>
<b>7 Making Career Future-Proof .....</b>	<b>121</b>
<b>7.1 Build ‘Evolve’ Mindset .....</b>	<b>123</b>
<b>7.2 Hone Core Skills .....</b>	<b>124</b>
<b>7.3 Develop Soft Skills.....</b>	<b>125</b>
<b>7.4 Maintain Digital Portfolio .....</b>	<b>126</b>
<b>7.5 Expand Network Globally .....</b>	<b>127</b>
<b>7.6 Monitor Industry Trends .....</b>	<b>128</b>
<b>8 Putting It All Together .....</b>	<b>131</b>
<b>8.1 Avoid Being a Junkie .....</b>	<b>133</b>
<b>8.2 Learn to Solve.....</b>	<b>135</b>
<b>8.3 Just Enough Approach .....</b>	<b>137</b>
<b>8.4 Conclusion .....</b>	<b>138</b>

---

## Preface

It is probably not a secret that data science/artificial intelligence (DS/AI) has become one of the most exciting fields of this age. Although it may seem the buzzword of our time, it is certainly not the hype. This exciting field opens the way to new possibilities and is becoming indispensable to our daily lives.

Organizations, big or small, are heavily investing in DS/AI research and applications these days. And hence, it has become the hottest career. If you want to become a DS/AI practitioner, there is no better time than this.

Aspirants are taking different approaches to get into the field, some are fortunate enough to be put into projects as freshers, but most aspirants are building their capabilities by learning theory and applying them on public data-sets.

*While there is no dearth of free & paid material, too much of information has only confused the current crop of DS/AI aspirants.*

Based on the questions that I am asked by them on day to day basis, I can see how perplexed they are. Not to say, taking advantage of the situation, most of the training institutes are minting money like anything, bundling even non-relevant courses as well in data science ones.

## Why this book?

From a time around when DS/AI field started picking up, every other day I get at least 8–10 messages from DS/AI starters & enthusiasts on ‘How can I get into DS/AI field?’. Over a while, I have improvised my response based on the follow-up questions they ask like:

1. What is the difference between DS, ML, DL, AI, DM?
2. What are the roles in DS/AI, who does what?
3. What concepts, processes & tools they need to learn?
4. Which books, courses, etc they need to refer to?
5. How to build a DS/AI portfolio?
6. How to write a resume for DS/AI?
7. How to build a helpful network?
8. How to search for the job?
9. How to prepare for the interview?
10. How to stay up to date in this still-evolving field?

You can notice that these questions are not conceptual ones and there is no dedicated material to address these roadblocks.

I thought why not to build a framework or a road-map for DS/AI starters and enthusiasts so that I need not to answer the same type of questions again and again. And that is when I started documenting what a starter or enthusiast need to do step by step in order to reach a level when he is ready to tackle any challenge thrown to him.

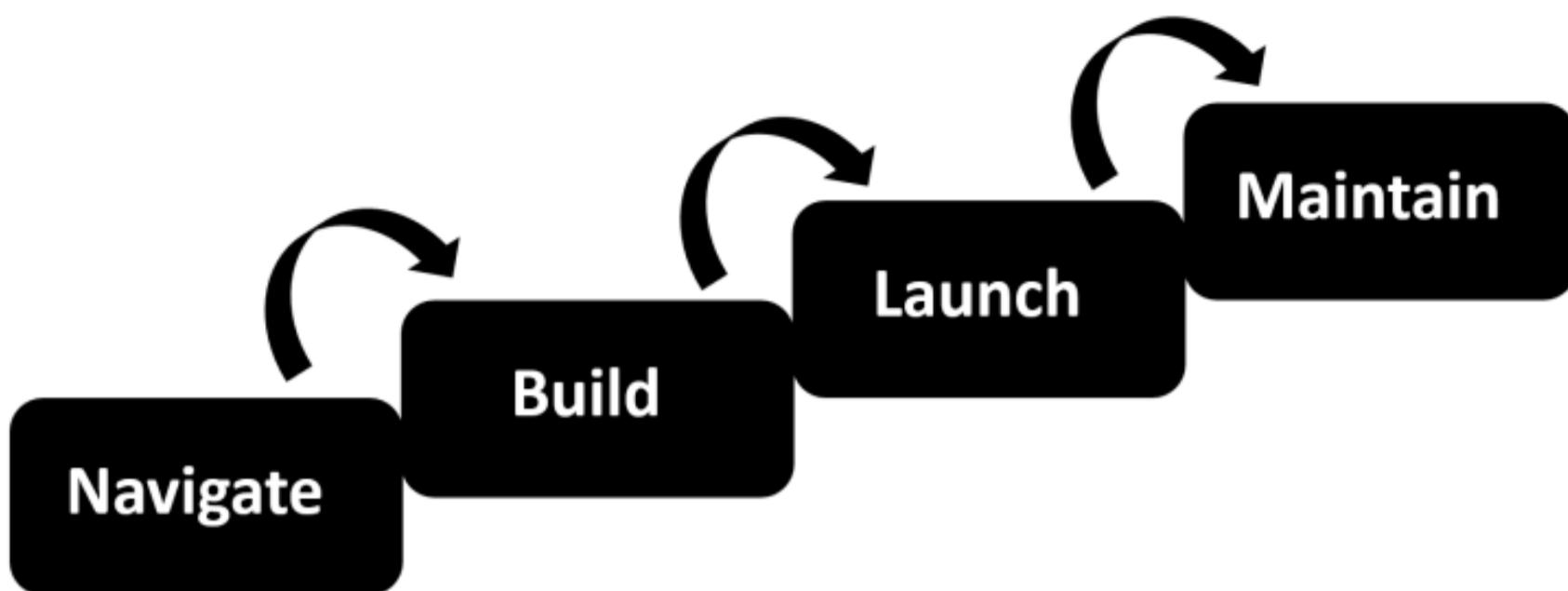
My answer to the above questions in a structured way to help DS/AI starters & enthusiasts is this book.

## What this book covers?

This book covers the framework to launch your DS/AI career in 8 chapters. Lets learn about the framework first:

*DS/AI field is not a rocket science, still you can relate launching your career with launching a satellite.*

On an abstract level, it looks fairly simple, overall process can be divided into four logical phases: Navigate, Build, Launch & Maintain.



## DS/AI: Self-Starter Kit

Build Your Own Roadmap

Chapter 1 covers the Big Picture of this book, the phases in above framework is discussed in more detail.

**Navigate** (Chapter 2) is the very first phase of your DS/AI journey, where you need to understand the overall landscape before diving deep. This step covers Q1 & Q2 from previous section.

**Build** (Chapter 3 & 4) phase covers all the concepts, processes, tools you need to learn and the resources you need to refer to gain required knowledge. This step responses Q3 & Q4 from previous section.

**Launch** (Chapter 5 & 6) is the phase where you build your portfolio, network with like-minded professionals and start looking for job. This step elaborates from Q5 to Q9 from previous section.

**Maintain** (Chapter 7) phase details out how you can stay up to date in this ever-evolving field. This step answers Q10 from previous section.

Chapter 8 covers puts all that we have learnt in this book together, details out 3 more roadblocks in your DS/AI journey and how to tackle them.

## Who this book is for?

This book can be useful for a variety of readers, but I wrote it with two main target audiences in mind. One of these target audiences is university students learning about DS/AI field, including those who are starting a career in data science and artificial intelligence. The other target audience is professionals working in other fields, who do not have a data science, statistics or programming background but want to rapidly acquire one and expand their career.

*There is no assumption or prerequisites for a reader to cover before reading this book.*

Looks interesting? Let's start with your journey in DS/AI field. But before that I would like to highlight one thing.

*Please note that this book does not cover Data Science & Artificial Intelligence concepts.*

Consider it as a personal coach to launch your career in DS/AI field who gives you enough exposure to the field so that you can yourself prepare a roadmap of your DS/AI journey.

There is a GitHub companion notebook for this book where you can find addendum & latest updates.

U0.2: GitHub Companion Notebook: <https://github.com/ankitrathi169/Self-Starter-Handbook>

ankitrathi.com/youtube

## Chapter 1

---

### Understanding the Big Picture



When you start learning a new skill, the first thing is to look at the big picture and where your would-be skills fit in the field. It gives you a context of what role you can play or are expected to play. And when the skill and field are evolving & overwhelmingly large, you are so engrossed in the details that most probably you tend to miss the purpose.



*In my view, to understand the big picture, ask yourself 'why' more often and start with the end in your mind.*

The following analogy is not particularly about DS/AI but in general.

Job — Interview — Network — Portfolio — Core skills — Resources — Awareness

As you can see above, to land the job you want you need to crack the interview. To get the call for the interview, you need a network that can refer you for opportunities. To have a relevant & helpful network, you need to have impressive credentials or portfolio. To have a solid portfolio, you need to work on the core skills of the job role that you want. To build core skills, you need to have the right resources. To gather the right

resources, you need to have awareness of the field in which you want to build your career.

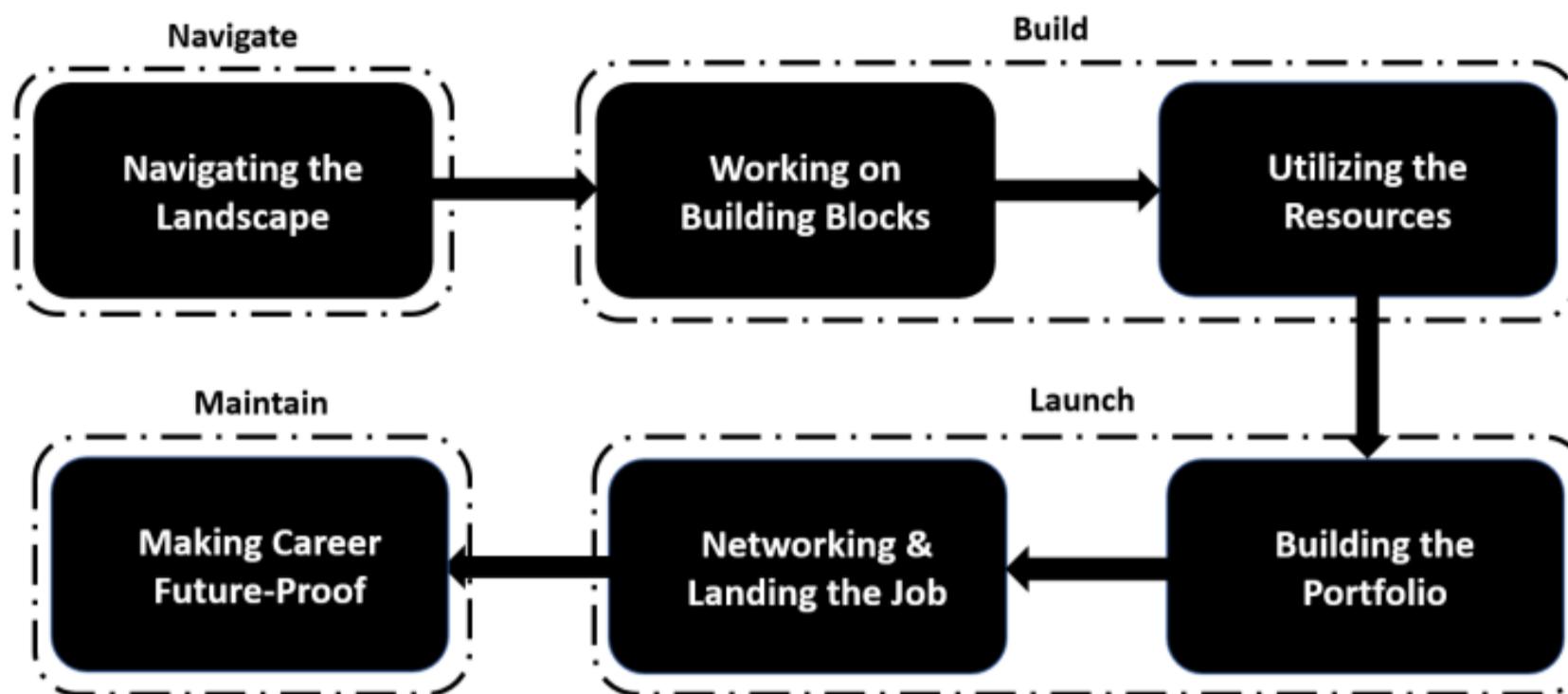


Now, reverse the order and assess at what stage you are and fill the gap in the coming months to build your career.

Based on the above approach, this chapter gives you a holistic view of what you need to know and do to build your road-map in DS/AI field.

*In upcoming sections, you will get to know about these steps and their details are covered in subsequent chapters of the book.*

Each section represents an upcoming chapter of this book.



## Understanding the Big Picture

DS/AI: Self-Starter Kit

### 1.1 Navigating the landscape

Before learning DS/AI, it is very much required that you understand the overall landscape and where all the buzzwords like data science, machine learning, deep learning fit in. In this chapter, you will learn different terminologies, their meaning and how these are interconnected.



### Navigating the Landscape

DS/AI: Self-Starter Kit

DS/AI is a vast field, not a single person has all the required knowledge or does all day to day tasks. Due to the variety of skills required in data science projects, specific roles are evolving so that individual can contribute according to their abilities. In this chapter, you will get to know almost all the DS/AI roles, which will give you a better idea which one suits you.

## 1.2 Working on building blocks

This is the core part of DS/AI, knowing the concept, process & tools are the most important part of any job. Intrinsic details of the building blocks are not in the scope of this book. This chapter will make you aware of what is in the scope of DS/AI field and which role need to have what kind of skills.



### Working on the Building Blocks

DS/AI: Self-Starter Kit

After completing this section you will get to know what is in the scope of DS/AI field. You can relate the concepts, process & tools based on the role you think you can fit in. Think of this chapter as the syllabus of DS/AI field with electives based on your role.

## 1.3 Utilizing the resources

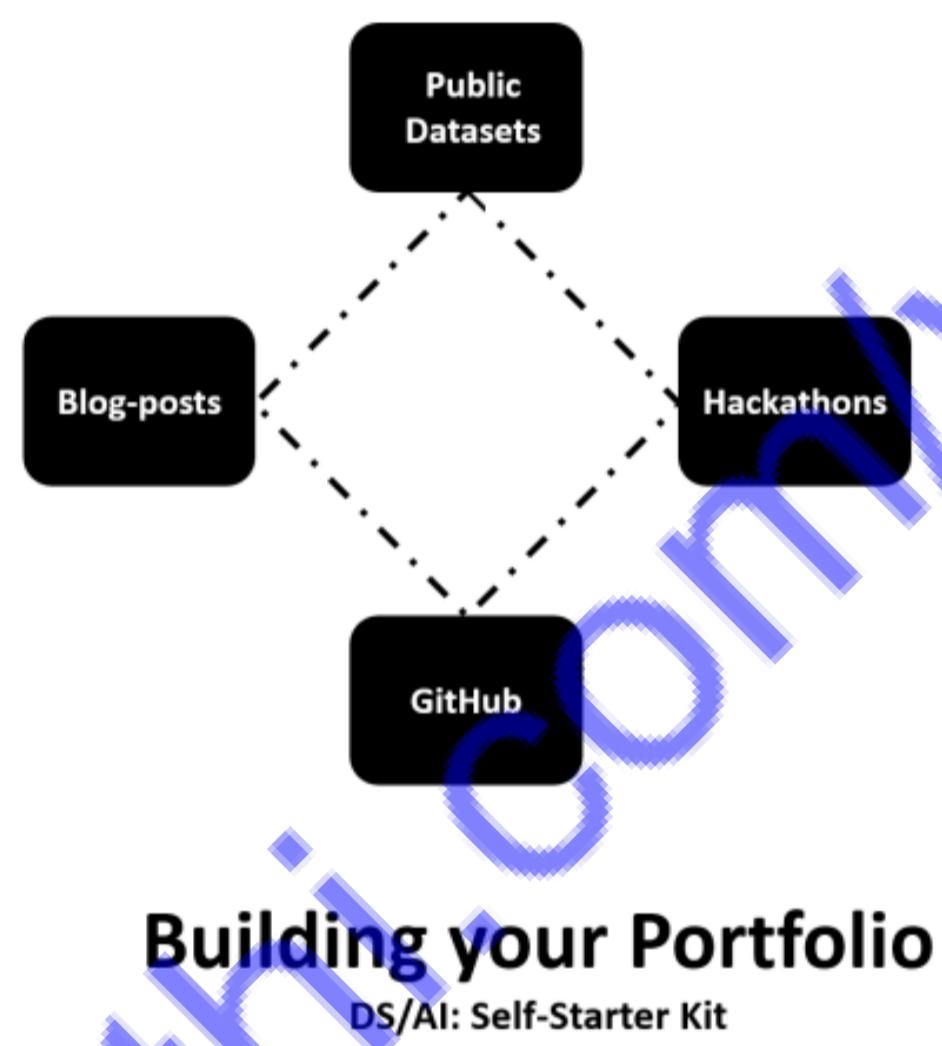
There is no dearth of learning resources but that has only confused the starters in the data science field. I get many questions related to what books to read, what courses to enrol, what blogs/portals to follow, what data-sets to work on.



In this chapter, I will provide critical reviews of all the prominent books, courses, blogs, portals, data-sets & podcasts etc. Please note that this list is based on my exposure to the field, there may be additional and better resources but I think you will get enough exposure after going through this chapter to evaluate other resources comparatively.

## 1.4 Building your portfolio

Refining & honing the skills required for a job is one part, showcasing what you can offer is another. What if you have all the required skills but nobody looking for the skill-set is aware that you do. Having a credible portfolio is an effective way to showcase your skill-set and talent.



After going through this chapter you will get to know what is needed to build an impressive portfolio even before entering the market looking for a relevant job.

## 1.5 Networking & landing the job

You know, you have filtered the best-suited role & honed your skill-set & built the portfolio as well. Your mission is not accomplished unless you

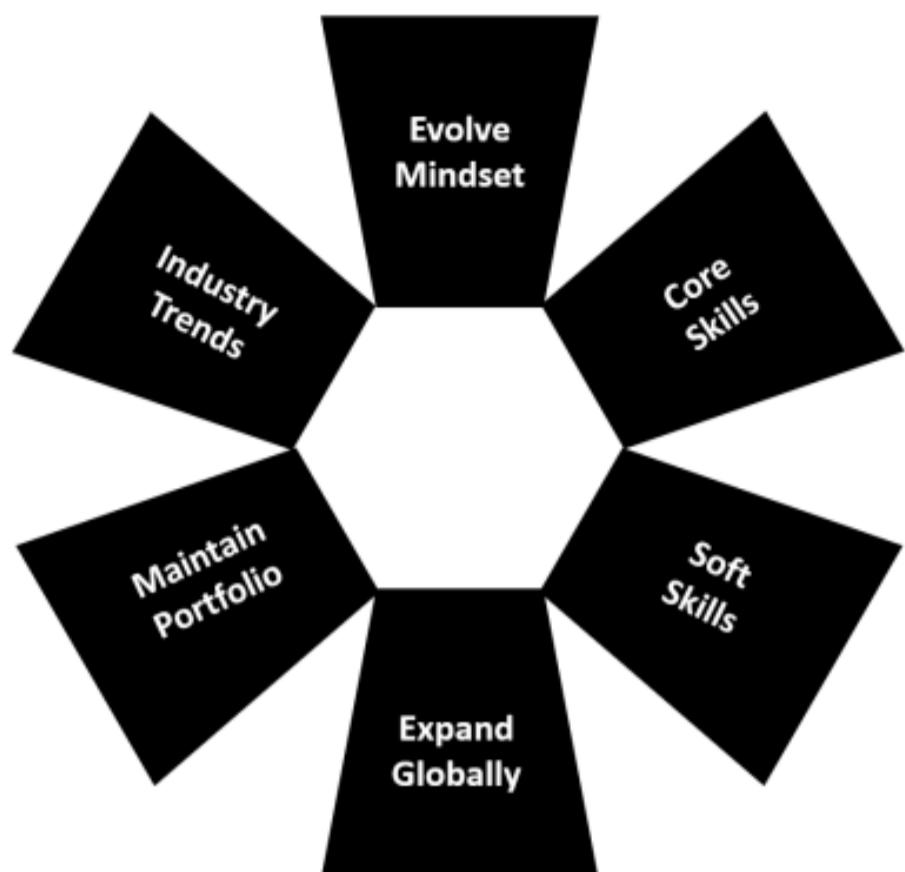
have like-minded professionals in your network who are aware of your capabilities and can refer you for the jobs that match your skills. Interview preparation is another task in itself.



This chapter will focus on the steps required to network & land the job in DS/AI field. How to network with like-minded professionals, how to search for relevant job openings, how to prepare & crack DS/AI interviews.

## 1.6 Making Career Future-Proof

Technology disruption & DS/AI evolution is changing the rule of the game over time. So rather than waiting for someone or technology to replace your labour, you'll take a proactive approach to put yourself in a position where potential employers can't afford to work without you.



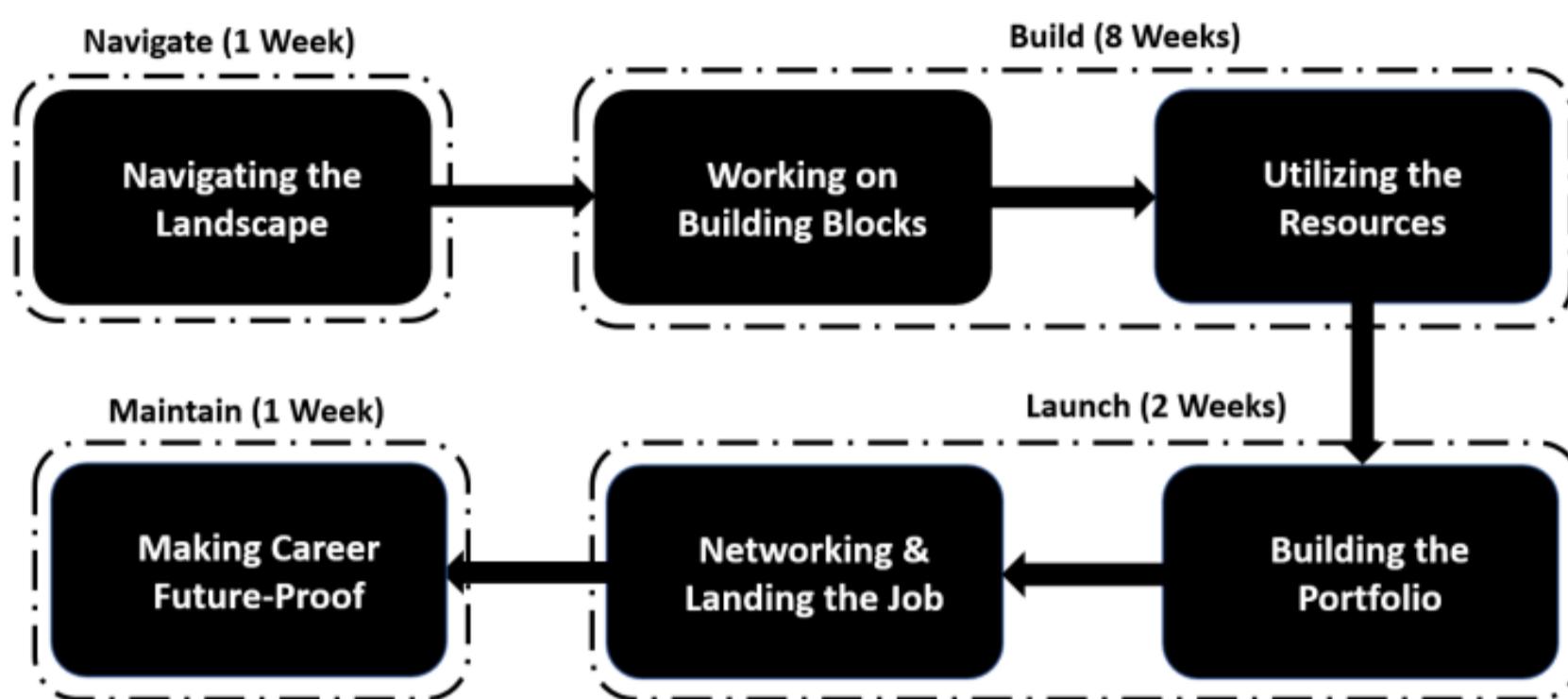
## Making Career Future-Proof

DS/AI: Self-Starter Kit

Future proofing your career is simply taking the extra steps to prepare yourself for constant technology disruption, one that's going to rely heavily on adaptability.

### 1.7 Putting it all together

Like in the current chapter, before diving into the details, you are looking at the big picture of what is be covered in this book and how that fits into your learning curve as a DS/AI starter/enthusiast.



## Putting it All Together

DS/AI: Self-Starter Kit

Similarly, after going through each step in detail in upcoming chapters, this chapter will revisit the high-level steps once again so that we can make sure that you have covered what we needed to.



**DS/AI: Self-Starter Kit**  
Build Your Own Roadmap

ankitrathi.com/youtube

[ankitratthi.com/youtube](http://ankitratthi.com/youtube)

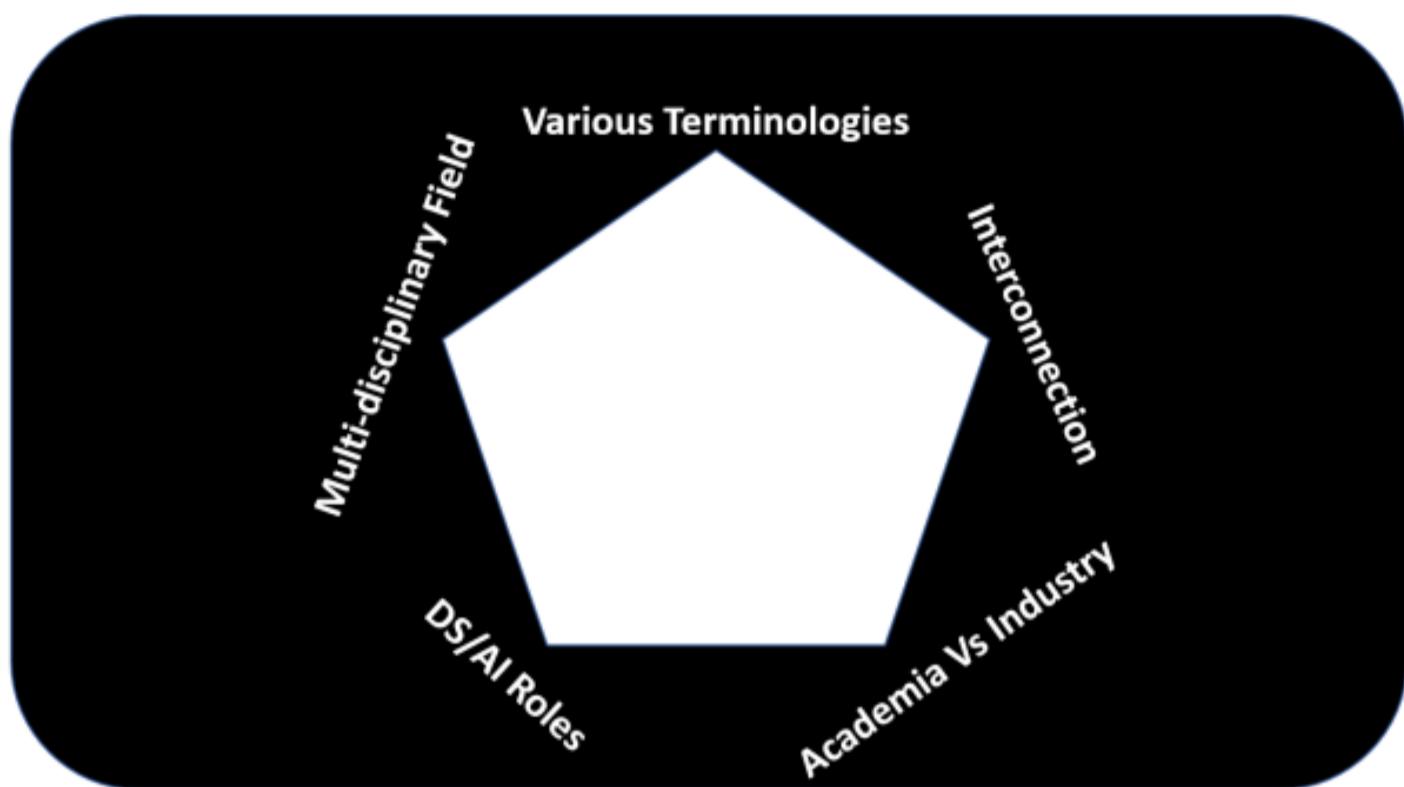
## Chapter 2

---

### Navigating the Landscape



DS/AI is a complex and evolving field. The first challenge a DS/AI aspirant faces is understanding the landscape and how he could navigate through it. Consider this, if you are travelling to a new city, and if you don't have the map, you will have trouble to navigate the city and you will need to ask a lot of random people during your travel without knowing how much they know about the place. Similarly, all the newcomers to data science have this trouble, and there are two ways to deal with this, arrange the map (or a guide) or travel yourself and learn with experience.



## Navigating the Landscape

DS/AI: Self-Starter Kit

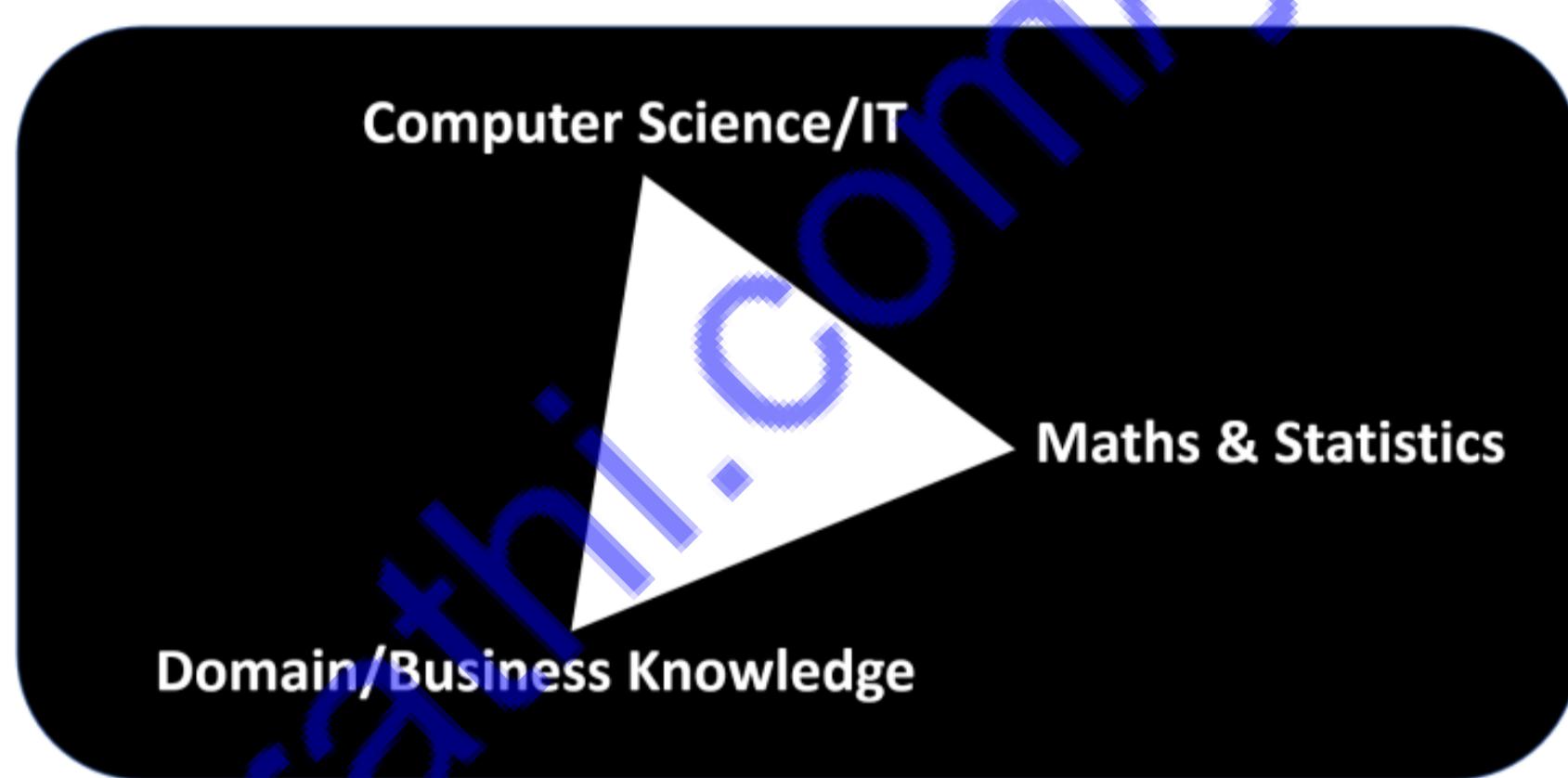
*This chapter intends to serve as a map of DS/AI field.*

You might have heard data science, machine learning, deep learning, artificial intelligence etc terminology but might not be fully aware of these terms, what to use when and how these topics are interconnected. After going through this chapter, you should be able to understand what is where in DS/AI field.

## 2.1 Multi-disciplinary field

*DS/AI is a multidisciplinary field with sub-fields of study in Math/Statistics, CS/IT & Business/Domain knowledge.*

Math/Statistics is required to understand the data and relationship between data elements. CS/IT skills are required to process the data to generate insights. And Business or domain knowledge is required to apply above to skills in the context of a business problem.



### Multi-disciplinary Field

Navigating the Landscape

#### Computer Science/IT

Programming is an essential skill to become a data scientist but one needs not be a hard-core programmer to learn DS/AI. Having familiarity with basic concepts of programming will ease the process of learning data science programming tools like Python/R. These basic concepts of

programming should help a candidate get a long way on the journey to pursue a career in DS/AI as it is all about writing efficient code to analyse big data and not being a master of programming. Individuals should learn the basics of programming in Python/R (or any relevant language) before they begin to work on DS/AI problems/projects.

## Maths & Statistics

Data science teams have people from diverse backgrounds like chemical engineering, physics, economics, statistics, mathematics, operations research, computer science, etc. You will find many data scientists with a bachelor's degree in statistics and machine learning but it is not a requirement to learn DS/AI. However, having familiarity with the basic concepts of Math and Statistics like Linear Algebra, Calculus, Probability, etc. is important to learn DS/AI.

## Domain/Business Knowledge

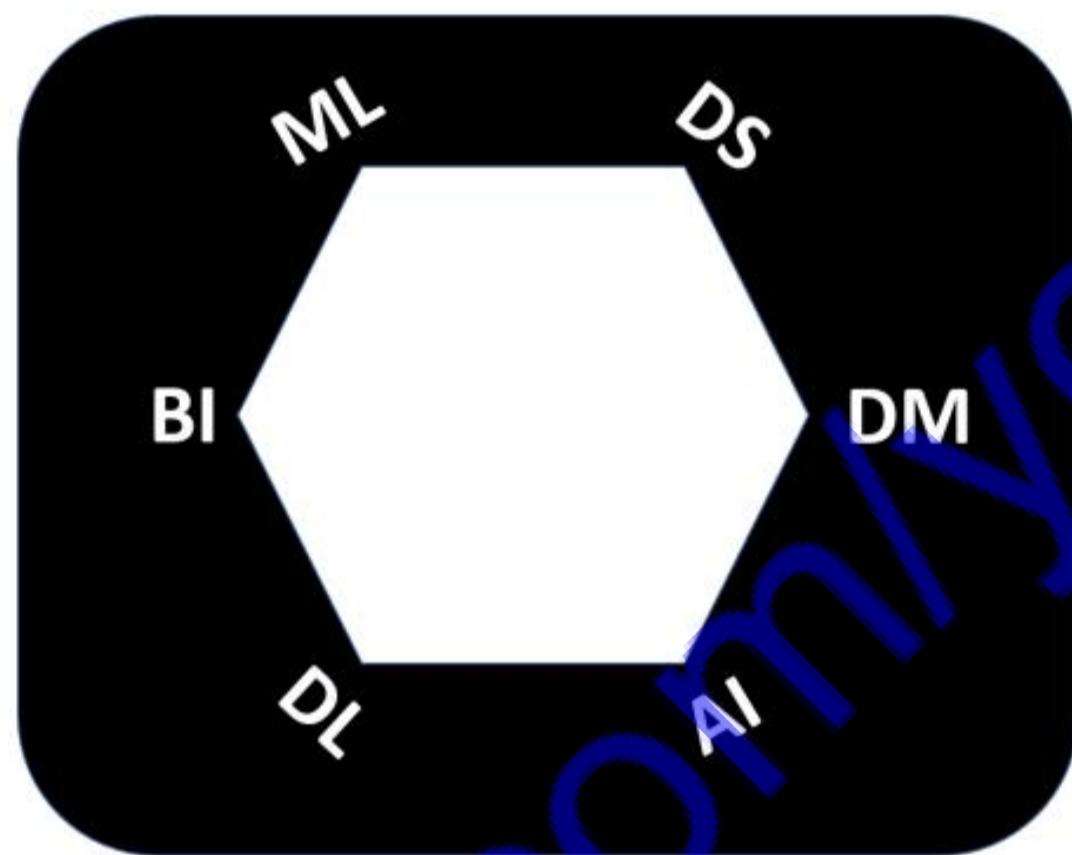
Subsequently, the business knowledge that the data scientists would need to have would be related to the domain that the project/analysis is in. For instance, if the data scientist is working for a credit card department in a bank, it will need to understand the specific business definitions, regulations, accounting policies & international standards, processes etc. This is the part that is more specific to the organization the data scientist is deployed in.

In my view, one thing to take care while hiring data scientists is not to give huge preference to domain knowledge. This may severely limit the supply of data science talents to the organization. You would have a better chance of getting more value from data science by looking for those that are strong in math & programming, being able to convert business objectives to mathematical models. Based on my observation, this is a much more difficult skill to find or train, as compared to domain knowledge.

U02.1: Multi-disciplinary field: <https://datajobs.com/what-is-data-science>

## 2.2 Various Terminologies

As a DS/AI starter, you will come across many similar terminologies. First thing you need to do is to understand what each term means and where each fits in the bigger picture. Data Science, Business Intelligence, Data Mining, Machine Learning, Deep Learning, Artificial Intelligence; let's have a look at Wikipedia definition for each term & later see how these are interconnected.



### Various Terminologies

Navigating the Landscape

#### Data Science

Data Science is an interdisciplinary field, where we try to solve a business problem by getting insights from available data after applying scientific methods on it.

## **Business Intelligence**

Business intelligence is a technology-driven process of converting raw data into meaningful information which can be used by business.

## **Data Mining**

Data mining is a process of collecting & discovering patterns in the data.

## **Machine Learning**

Machine learning is a field where we develop predictive models that are generic and can be applied to any domain related data problem.

## **Deep Learning**

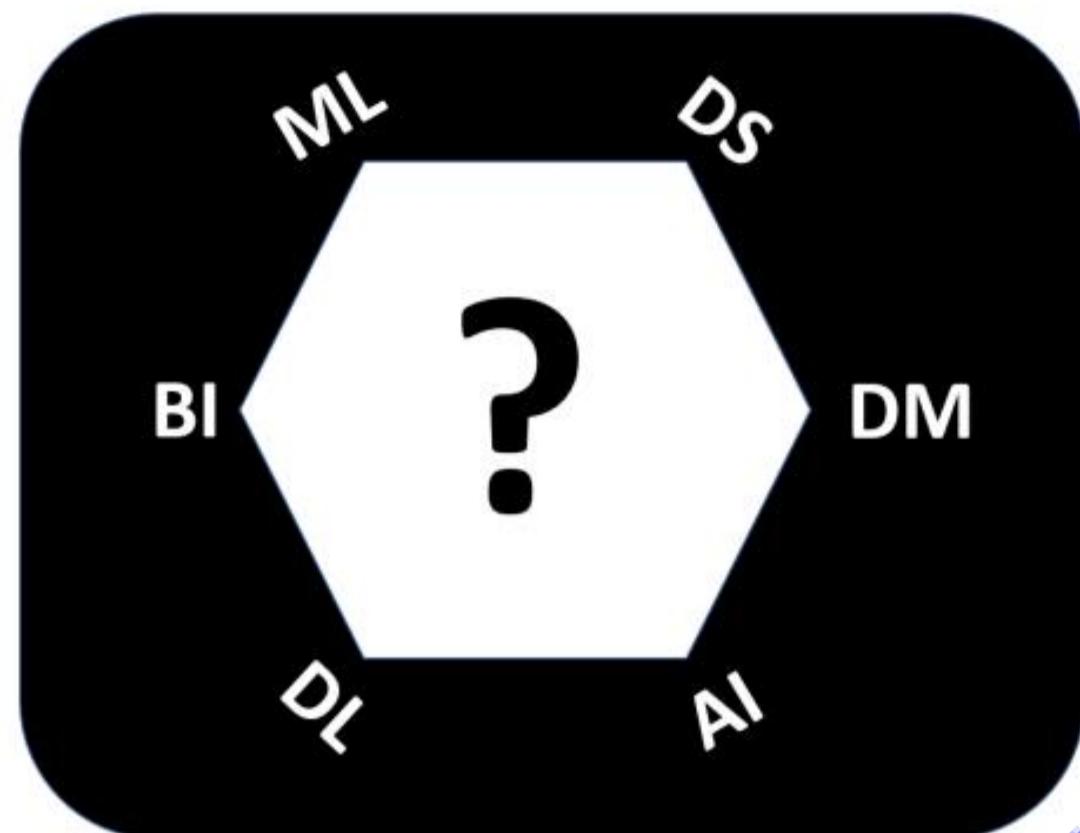
Deep learning is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms.

## **Artificial Intelligence**

Artificial intelligence is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals.

U02.2: Various Terminologies: <https://www.dezyre.com/article/data-science-compared-with-different-analytics-disciplines/175>

## 2.3 Interconnection



### Interconnection

Navigating the Landscape

Business Intelligence is a plain technology-driven process to analyse and provide actionable intelligence to executives/managers.

Data mining uses statistics and other programming languages to find hidden patterns in the data to explain a certain phenomenon. It helps in building a perception about the data using both math and programming.

Machine Learning deploys data mining techniques as well as other algorithms to develop models of what is happening behind some data to forecast future outcomes.

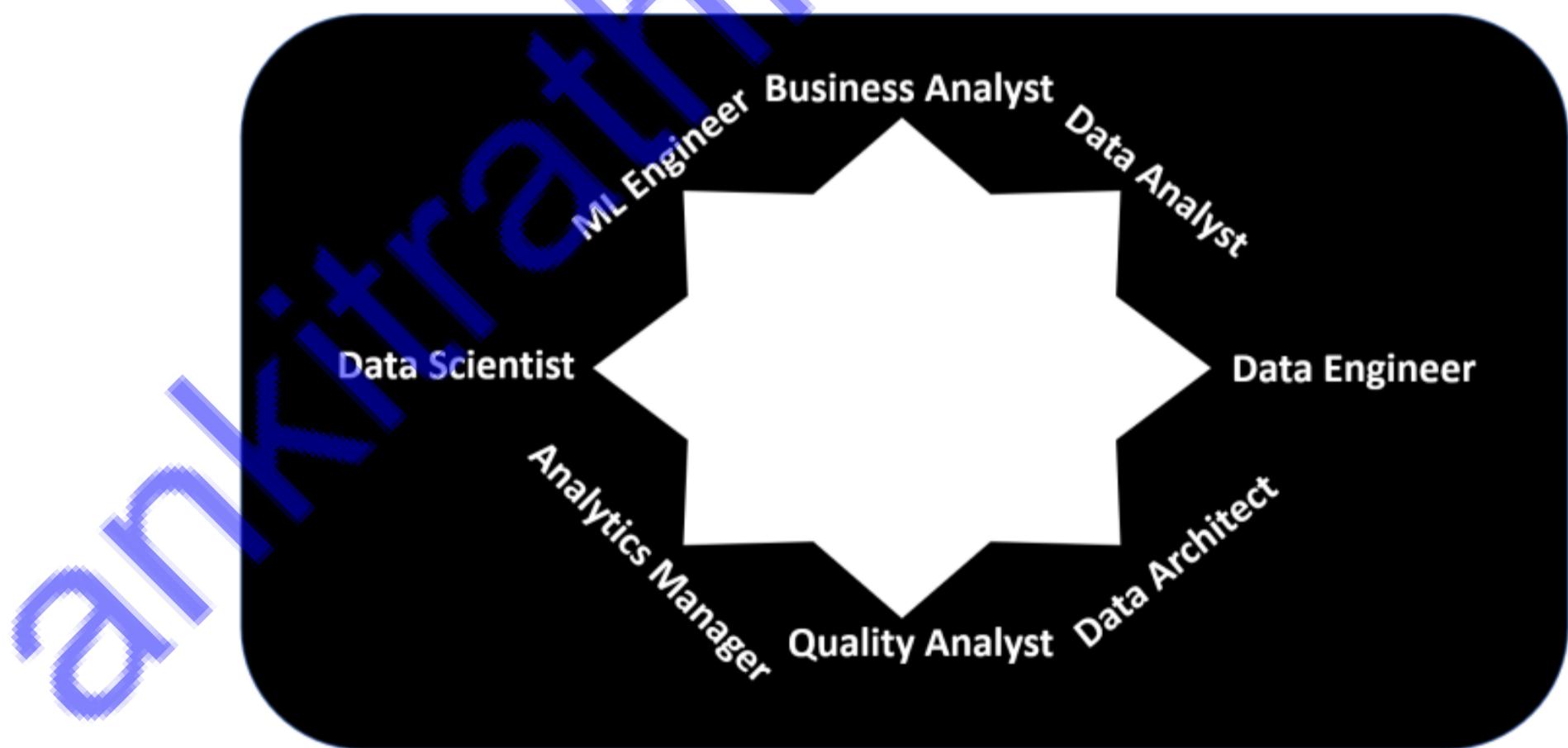
Artificial Intelligence uses models developed by Machine Learning and other algorithms to lead to intelligent behaviour. AI is very much programming based.

- Business Intelligence provide actionable insights
- Data Mining demonstrates patterns
- Machine Learning forecasts with models
- Artificial Intelligence shapes behaviours

*So you see that these terms are different but still inter-connected.*

U02.3: Interconnection: <https://www.quora.com/What-is-the-difference-between-artificial-intelligence-machine-learning-data-mining-and-business-intelligence-How-they-are-related>

## 2.4 DS/AI Roles



### DS/AI Roles

Navigating the Landscape

Before looking into the skill-set of a data scientist, let's have a look at various roles required to work and deliver a data science project, after all, it's a teamwork.

*Every role has its own skills that are critical to data science projects at various stages.*

## Data Analyst

Data analysts translate numbers into plain English. Every business collects data, whether it's sales figures, market research, logistics, or transportation costs. A data analyst's job is to take that data and use it to help companies make better business decisions. There are many different types of data analysts in the field, including operations analysts, marketing analysts, financial analysts, etc.

## Data Scientist

A data scientist is someone who knows how to extract meaning from and interpret data, which requires both tools and methods from statistics and machine learning. She spends a lot of time in the process of collecting, cleaning, and munging data. Domain knowledge is also an integral part of the skill.

## Data Engineer

Data Engineers are responsible for the creation and maintenance of analytics infrastructure that enables almost every other function in the data world. They are responsible for the development, construction, maintenance and testing of architectures, such as databases and large-scale processing systems.

## Data Architect

Data architects build complex computer database systems for companies, either for the general public or for individual companies. They work with a team that looks at the needs of the database, the data that is available and creates a blueprint for creating, testing and maintaining that data architecture.

## Analytics Manager

The data science manager coordinates the different tasks that must be completed by their team for a DS/AI project. Tasks may include researching and creating effective methods to collect data, analyzing information, and recommending solutions to business.

## Business Analyst

Data science business analyst converts the business problem statement to a DS/AI problem statement which means what data needs to be analyzed to arrive at the insights. The data would then be reviewed with the technology team and results would be delivered to the business team in the form of insights and data patterns. The business analyst should also be knowledgeable enough to apply various predictive modelling techniques and right model selection for generating insights for the problem at hand.

## Quality Analyst

The job of quality analyst includes checking the quality of the training data-set, preparing data-sets for testing, running statistics on human-labelled data-sets, evaluating precision and recall on the resulting ML model, reporting on unexpected patterns in outputs, and implementing necessary tools to automate repetitive parts of the work. Experience in software testing with data quality or DS/ML focus, understanding of

statistics, exposure to Data Science / Machine Learning techniques and coding proficiency in Python, are some of the skills required for the job.

*To work on DS/AI projects in any of the above mentioned roles, one needs to have an understanding of the core concepts at a high level but depth is required in the specific area you would be working in.*

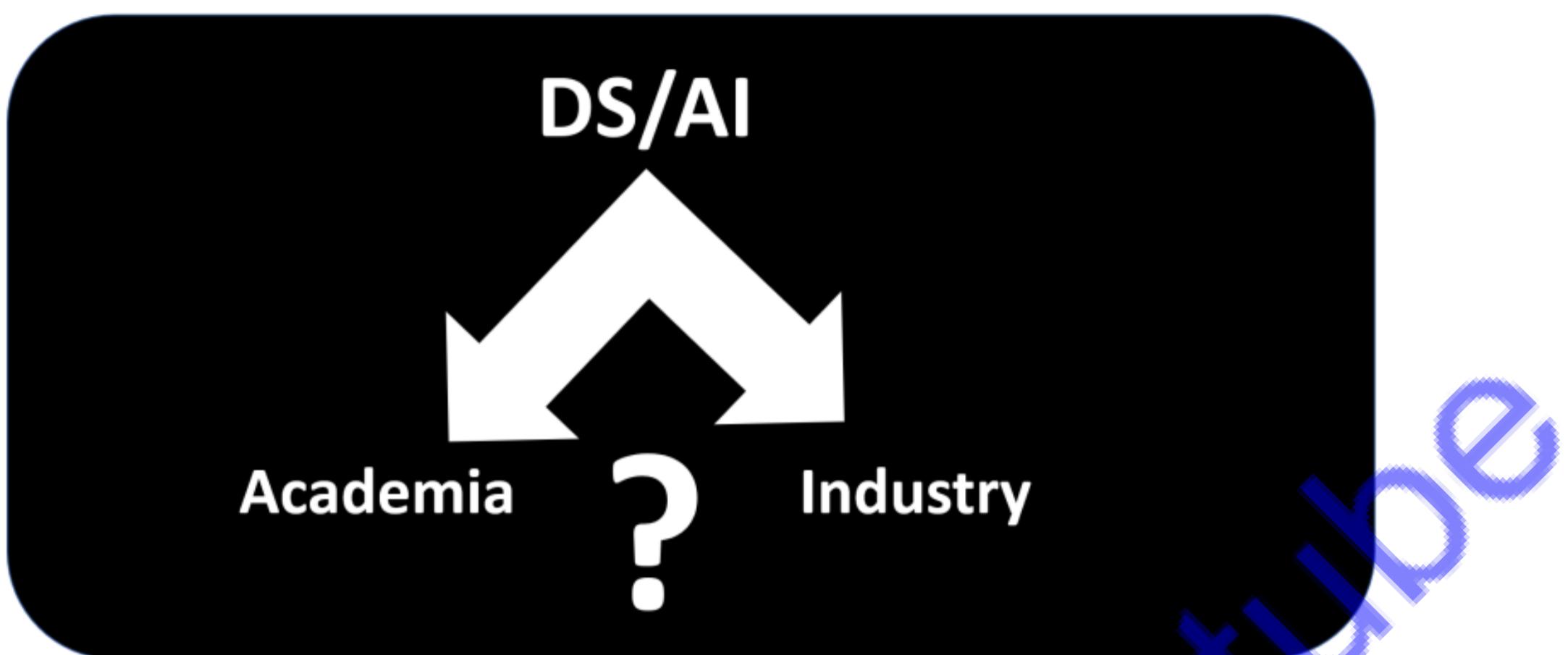
U02.4: DS/AI Roles: <https://hackernoon.com/top-10-roles-for-your-data-science-team-e7f05d90d961>

## 2.5 Academics Vs Industry

Academia and Industry are different fields with different people and culture. People working in Academia for longer tenure may find it difficult to adjust to industry culture and vice versa.

*There is also an academic trap when your career trajectory is so specialized for academia that you're unprepared for a job outside of it.*

The academic trap happens in all areas of study, but for this chapter, we focus only on DS/AI students who want to leave academia for data science positions.



## Academia Vs Industry

*Navigating the Landscape*

Further, companies are often hesitant to hire people coming straight from academia for various reasons like:

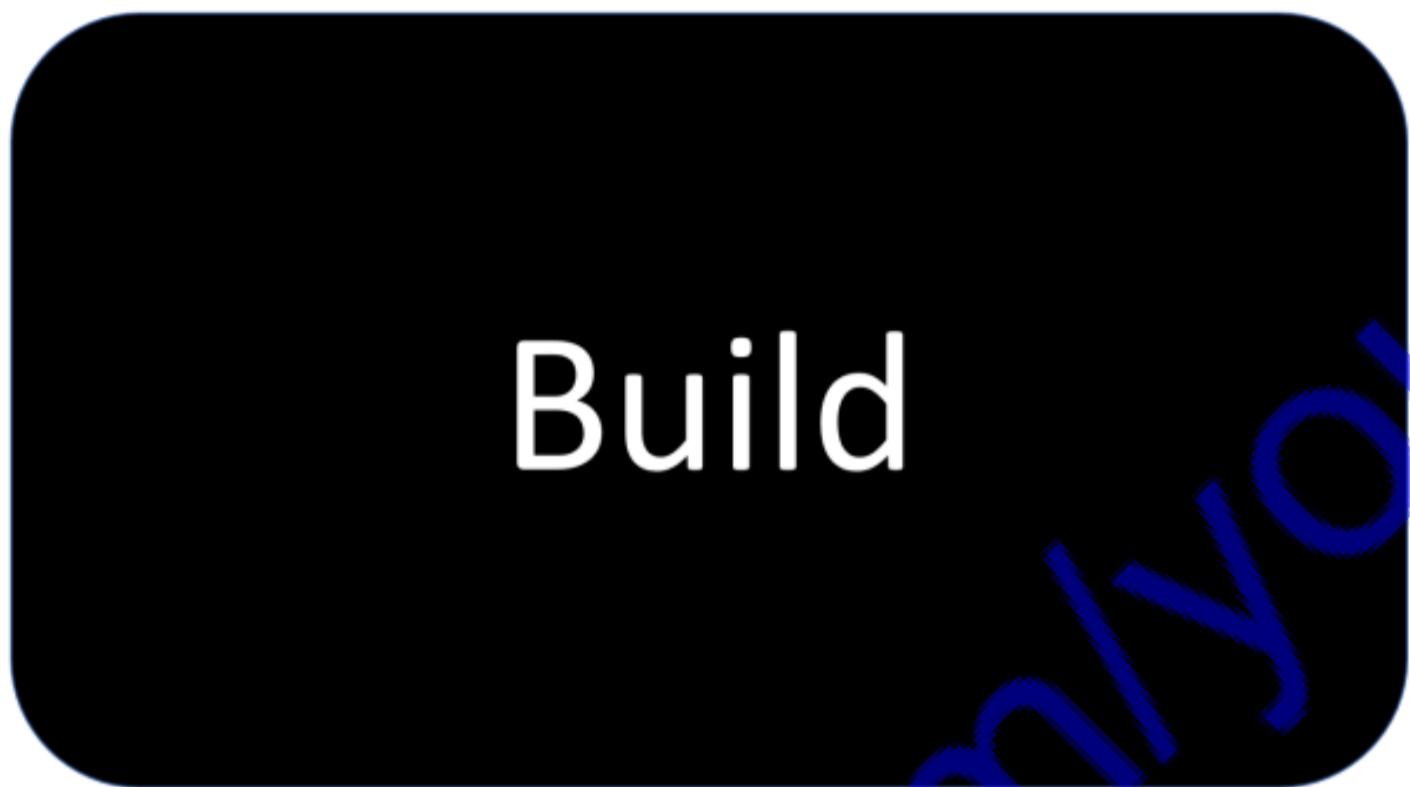
- In academia, individuals prefer writing papers over internships, making grants over learning programming languages, and not doing the things that could help you in the industry but not academia. The things that are important for academic hirings, such as papers, talks, and grants, are not as important in the industry.
- Working as a data scientist within a corporation requires an understanding of how the business world works, including how quickly deliverable need to be made, how to craft a good presentation, and how to word an email to make a request.
- In academia, you are encouraged to find the most innovative and elegant solution. In industry, you are encouraged to spend as little time as possible to find an analytical solution that just fits the need.
- Salary expectations for advanced degree holders are higher than someone with only an undergraduate degree. This also pushes away recruiters as the industry works in a different way, culture is simply different than the academic one. People coming from academia

need to learn these lessons at their first job, which means that there is a lot of risk for the hiring company.

U02.5: Academia Vs Industry: <https://www.northeastern.edu/graduate/blog/working-in-industry-vs-academia/>

ankitrathi.com/youtube

[ankitratthi.com/youtube](http://ankitratthi.com/youtube)



**DS/AI: Self-Starter Kit**  
Build Your Own Roadmap

ankitrathi.com/youtube

[ankitratthi.com/youtube](http://ankitratthi.com/youtube)

## Chapter 3

---

### Working on Building Blocks



There are few core skills of every job. To perform that job, you need to be aware of core concepts, you need to be aware of the end to end process and you need to learn how to use related tools to perform that job. Data science is no different job, it has its own core concepts, processes and tools.

This chapter covers the core concepts you need to learn, end-to-end process you need to be aware of & important tools you need to master to work as a data scientist.



## Working on the Building Blocks

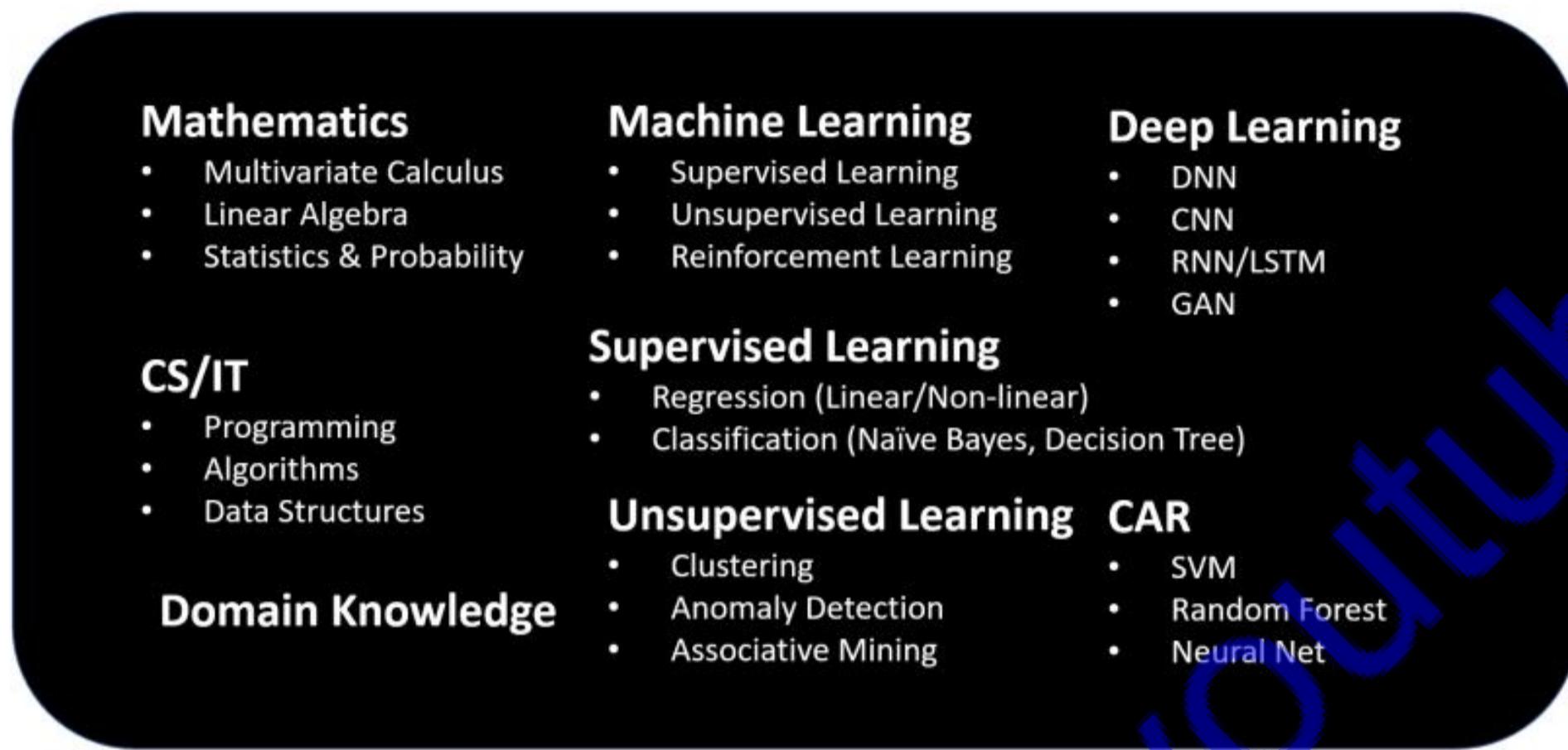
DS/AI: Self-Starter Kit

*Please note that this chapter only outlines the concepts, processes and tools used by data scientists. I have covered the resources (mostly free) for these topics in upcoming chapter.*

Still, if you want to build a quick understanding you can refer following link:

U03: DS/AI The Complete Reference: <https://medium.com/data-deft/data-science-the-complete-reference-series-3fb35077fc5a>

## 3.1 Concepts to Learn



### Concepts to Learn

Working on Building Blocks

## Mathematics

Data science contains math — no avoiding that! This section is for learners about basic math they need in order to be successful in almost any data science project/problem. So let's start:

### Multivariate Calculus

Calculus is a set of tools for analyzing the relationship between functions and their inputs. In Multivariate Calculus, we can take a function with multiple inputs and determine the influence of each of them separately.

In data science, we try to find the inputs which enable a function to best match the data. The slope or descent describes the rate of change off the output with respect to an input. Determining the influence of each input on the output is also one of the critical tasks. All this requires a solid understanding of Multivariate Calculus.

## Linear Algebra

The word *algebra* comes from the Arabi word “*al-jabr*” which means “*the reunion of broken parts*”. This is the collection of methods deriving unknowns from knowns in mathematics. *Linear Algebra* is the branch that deals with *linear equations* and *linear functions* which are represented through *matrices* and *vectors*. In simpler words, it helps us understand geometric terms such as planes, in higher dimensions, and perform mathematical operations on them. By definition, algebra deals primarily with scalars (one-dimensional entities), but Linear Algebra has vectors and matrices (entities which possess two or more dimensional components) to deal with linear equations and functions.

*Linear Algebra* is central to almost all areas of mathematics like *geometry* and *functional analysis*. Its concepts are a crucial prerequisite for understanding the theory behind *Data Science*. You don't need to understand *Linear Algebra* before getting started in *Data Science*, but at some point, you may want to gain a better understanding of how the *different algorithms* really work under the hood. So if you really want to be a professional in this field, you will have to master the parts of *Linear Algebra* that are important for *Data Science*.

## Statistics & Probability

*Statistics* is a mathematical body of science that pertains to the *collection, analysis, interpretation* or *explanation*, and *presentation* of data. Probability is the chance that something will happen — how likely it is that some event will happen.

Statistics help you to understand your data and is an initial & very important step of Data Science. This is due to the fact that Data Science is all about making predictions and you can't predict if you can't understand the patterns in existing data.

Uncertainty and randomness occur in many aspects of our daily life and having a good knowledge of probability help us make sense of these uncertainties. Learning about probability helps us make informed judgments on what is likely to happen, based on a pattern of data collected previously or an estimate.

Data science often uses statistical inferences to predict or analyze trends from data, while statistical inferences use probability distributions of data. Hence knowing probability & statistics and its applications are important to work effectively on data science problems.

## Programming

To execute the DS/AI pipeline, you need to learn algorithm design as well as fundamental programming concepts such as data selection, iteration and functional decomposition, data abstraction and organisation. In addition to this, you need to learn how to perform simple data visualizations using programming and embed your learning using problem-based assignments.

## Machine Learning Algorithms

Machine learning algorithms can be divided into 3 broad categories —

- Supervised learning,
- Unsupervised learning
- Reinforcement learning

Supervised learning is useful in cases where a property (*label*) is available for a certain dataset (*training set*) but is missing and needs to be predicted for other instances. Unsupervised learning is useful in cases where the challenge is to discover implicit relationships in a given *unlabelled* dataset (items are not pre-assigned). Reinforcement

learning falls between these 2 extremes — there is some form of feedback available for each predictive step or action, but no precise label or error message.

*Intrinsic details of various algorithms is not in scope of this book, you can refer the resources mentioned in the next chapter to learn them.*

Supervised learning can be further divided into Regression (Linear, Non-linear etc) & Classification (Logistics Regression, Decision Tree, Naïve Bayes etc) algorithms. Some algorithms can be used for regression as well as classification i.e. Random Forests, Support Vector Machines etc.

Unsupervised learning can also be further divided into Clustering, Anomaly Detection, Associative Mining.

Reinforcement learning is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward.

## Deep Learning Frameworks

Deep learning frameworks are a more advanced form of ML and solve specific problems where data is either unstructured or huge or both. Neural Nets, CNNs, RNNs & LSTM, GANs are the frameworks one needs to be aware of.

*If you want to get understanding of algorithms and frameworks mentioned here, I have covered the*

*reference material in upcoming chapter, or you can refer following link for top 10 algorithms.*

U03.1: Top 10 ML Algorithms: <https://towardsdatascience.com/top-10-machine-learning-algorithms-for-data-science-cdb0400a25f9>

## Domain Knowledge

This lack of domain knowledge, while perfectly understandable, can be a major barrier to data scientists. For one thing, it's difficult to come up with project ideas in a domain that you don't know much about. It can also be difficult to determine the type of data that may be helpful for a project — if you want to build a model to predict an outcome, you need to know what types of variables might be related to this outcome so you can make sure to gather the right data.

Knowing the domain is useful not only for figuring out projects and how to approach them but also for having rules of thumb for sanity checks on the data. Knowing how data is captured (is it hand-entered? Is it from machines that can give false readings for any number of reasons?) can help a data scientist with data cleaning and from going too far down the wrong path. It can also inform what true outliers are and which values might just be due to measurement error.

Often the most challenging part of building a machine learning model is feature engineering. Understanding variables and how they relate to an outcome is extremely important for this. Knowing the domain can help direct the data exploration and greatly speed (and enhance) the feature engineering process.

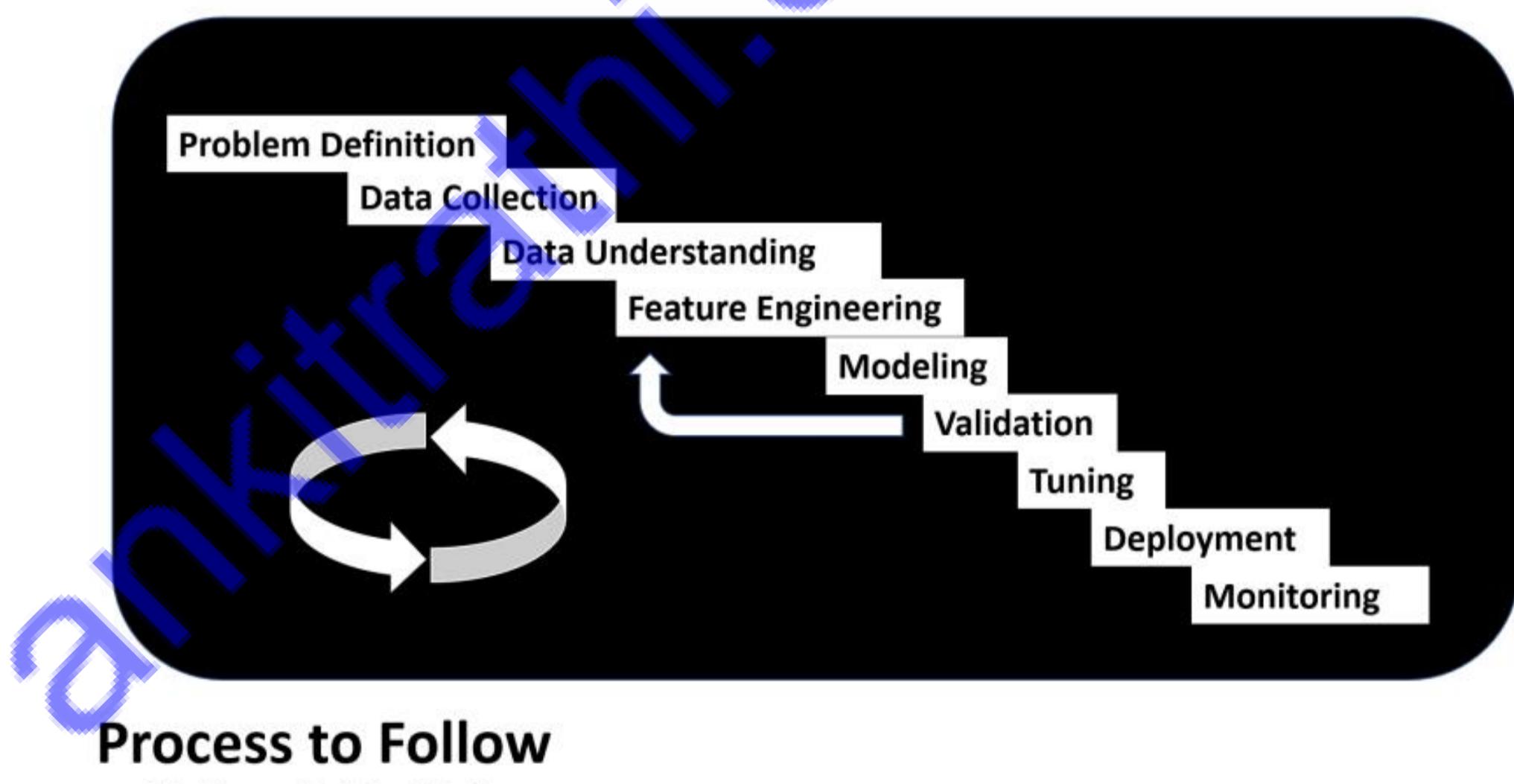
Once features are generated, knowing what relationships between variables are plausible help for basic sanity checks. Being able to glance at

the outcome of a model and determine if they make sense goes a long way for quality assurance of any analytical work.

*Finally, one of the biggest reasons a strong understanding of the data is important is because you have to interpret the results of analyses and modelling work.*

Knowing what results are important and which are trivial is important for the presentation and communication of results. It's also important to know what results are actionable.

### 3.2 Process to Follow



## Problem Definition

The first thing you have to do before you solve a problem is to define exactly what it is. You need to be able to translate data questions into something actionable.

You'll often get ambiguous inputs from the people who have problems. You'll have to develop the intuition to turn scarce inputs into actionable outputs—and to ask the questions that nobody else is asking.

## Data Collection

Once you've defined the problem, you'll need data to give you the insights needed to turn the problem around with a solution. This part of the process involves thinking through what data you'll need and finding ways to get that data, whether it's querying internal databases, or purchasing external data-sets.

## Data Understanding

The difficulty here isn't coming up with ideas to test, it's coming up with ideas that are likely to turn into insights. You'll have a fixed deadline for your data science project, so you'll have to prioritize your questions.

You'll have to look at some of the most interesting patterns that can help explain why sales are reduced for this group. You might notice that they don't tend to be very active on social media, with few of them having Twitter or Facebook accounts. You might also notice that most of them are older than your general audience. From that you can begin to trace patterns you can analyze more deeply.

# Feature Engineering

Feature engineering is the process of using domain knowledge of the data to create features that make machine learning algorithms work. If feature engineering is done correctly, it increases the predictive power of machine learning algorithms by creating features from raw data that help facilitate the machine learning process. Feature Engineering is in fact an art.

## Modeling

Depending on the type of question that you're trying to answer, there are many modelling algorithms available. You run the selected algorithm/s on the training data to build the models.

## Validation

Validation is a step used to evaluate the trained model on validation data. You use a series of steps competing for machine-learning algorithms along with the various associated tuning parameters that are geared toward answering the question of interest with the current data.

## Tuning

Tuning an algorithm or machine learning technique can be simply thought of as a process which one goes through in which they optimize the parameters that impact the model in order to enable the algorithm to perform the best.

# Deployment

After you have a set of models that perform well, you can operationalize them for other applications to consume. Depending on the business requirements, predictions are made either in real-time or on a batch basis. To deploy models, you expose them with an open API interface. The interface enables the model to be easily consumed from various applications.

U03.2: DS/AI Process: <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview>

## 3.3 Tools to Master



**Tools to Master**  
Working on Building Blocks

*The list mentioned here is not exhaustive, it depends more on what kind of problem you are solving and in what tech stack you are working.*

## SQL

Structured Query Language (SQL) is a standard computer language for relational database management and data manipulation. SQL is used to query, insert, update and modify data. Most relational databases support SQL.

As data collection has increased exponentially, so has the need for people skilled at using and interacting with data; to be able to think critically, and provide insights to make better decisions and optimize their businesses. The skills necessary to be a good data scientist include being able to retrieve and work with data and to do that you need to be well versed in SQL, the standard language for communicating with database systems.

U03.3.1: SQL: <https://www.tutorialspoint.com/sql/>

## R

R is a programming language and software environment for statistical analysis, graphics representation and reporting. In the world of data science, R is an increasingly popular language for a reason. It was built with statistical manipulation in mind, and there's an incredible ecosystem of packages for R that let you do amazing things — particularly in data visualization.

U03.3.2: R: <https://www.statmethods.net/r-tutorial/index.html>

## Python

Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. Python is no-doubt the best-suited language for a Data Scientist. It is a free, flexible and powerful open-source language. Python cuts development time in half with its simple and easy to read syntax. With Python, you can perform data manipulation, analysis, and visualization. Python provides powerful libraries for Machine learning applications and other scientific computations.

U03.3.3: Python: [http://www.tutorialspoint.com/python/python\\_data\\_science.htm](http://www.tutorialspoint.com/python/python_data_science.htm)

## Tensorflow

Currently, the most famous deep learning library in the world is Google's TensorFlow. Google product uses machine learning in all of its products to improve the search engine, translation, image captioning or recommendations.

TensorFlow is the best library of all because it is built to be accessible to everyone. Tensorflow library incorporates different API to built at scale deep learning architecture like CNN or RNN. TensorFlow is based on graph computation; it allows the developer to visualize the construction of the neural network with Tensorboard. This tool is helpful to debug the program. Finally, Tensorflow is built to be deployed at scale. It runs on CPU and GPU.

U03.3.4: Tensorflow: <https://www.guru99.com/tensorflow-tutorial.html>

## Keras

Keras is a high-level neural networks API, capable of running on top of Tensorflow, Theano, and CNTK. It enables fast experimentation through a high level, user-friendly, modular and extensible API.

Keras allows for easy and fast prototyping (through user-friendliness, modularity, and extensibility). It supports both convolutional networks and recurrent networks, as well as combinations of the two. It runs seamlessly on CPU and GPU.

U03.3.5: Keras: <https://www.guru99.com/keras-tutorial.html>

ankitrathi.com/youtube

## Chapter 4

---

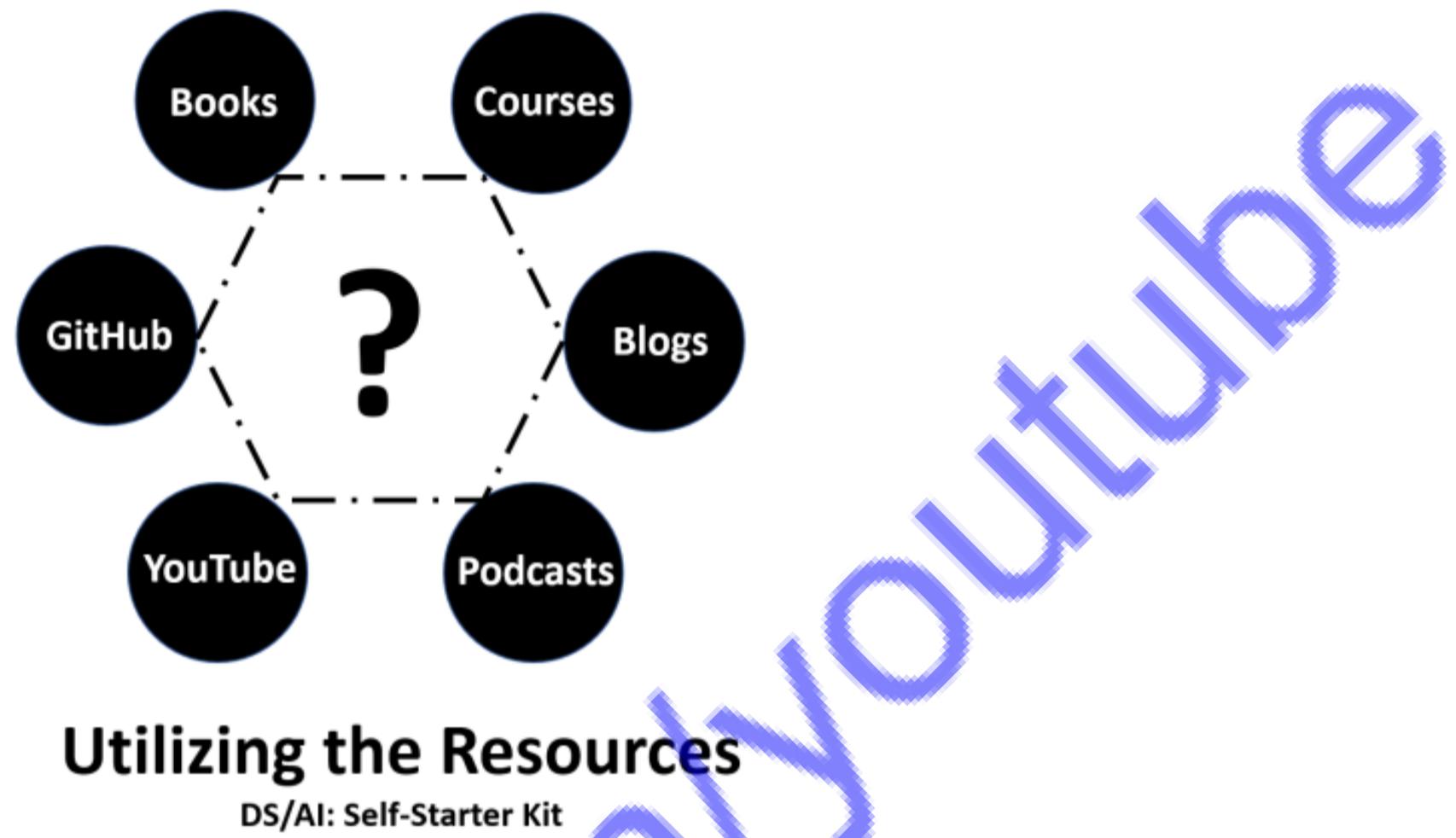
### Utilizing the Resources



After getting to know DS/AI landscape and its building blocks in previous chapter, you may ask which are the resources to refer, how do I know if a book or course is worth to spend time and/or money?

This is not an exhaustive list by any means, but it is good enough to keep as your reference. You can build your own list of references once you get more awareness of the field.

This chapter is my attempt to make your task easier. I am listing down major quality resources (mostly free) here and also going to provide you with my view of these resources, which will help you to make an informed decision.



*You need not go through each and every resource mentioned here, I would suggest you build the foundation first using a course or a book and keep other resources for your reference.*

## 4.1 Books to refer



### Books to Refer Utilizing the Resources

## Machine Learning with R

This is an excellent book for the R starter who wants to apply ML to any kind of project. All the main ML models are presented, as well as different performance metrics, bagging, pruning, tuning, ensembling etc. Easy to scan through, many tips with fully-solved textbook problems. Certainly, a very good starting point if you plan to compete on Kaggle. If you already master both R and ML, this books is obviously not for you.

U04.1.1: Machine Learning with R: <https://www.goodreads.com/en/book/show/26033041-machine-learning-with-r---second-edition>

## Python Machine Learning

This is a fantastic introductory book in machine learning with python. It provides enough background about the theory of each (covered)

technique followed by its python code. One nice thing about the book is that it starts implementing Neural Networks from scratch, providing the reader with the chance of truly understanding the key underlying techniques such as back-propagation. Even further, the book presents an efficient (and professional) way of coding in python, the key to data science.

U04.1.2: Python Machine Learning: <https://www.goodreads.com/book/show/25545994-python-machine-learning>

## ISLR

The book explains the concepts of Statistical Learning from the very beginning. The core ideas such as bias-variance trade-off are deeply discussed and revisited in many problems. The included R examples are particularly helpful for beginners to learn R. The book also provides a brief, but concise description of functions' parameters for many related R packages. Compared to *The Elements of Statistical Learning*, it is easy for the reader to understand. It does a wonderful job of breaking things down complex concepts. If one wishes to learn more about a particular topic, I'd recommend *The Element of Statistical Learning*. These two pair nicely together.

U04.1.3: ISLR: <https://www.goodreads.com/book/show/17397466-an-introduction-to-statistical-learning>

## Deep Learning

This is the book to read on deep learning. Written by luminaries in the field — if you've read any papers on deep learning, you must have heard about Goodfellow and Bengio before — and cutting through much of the BS surrounding the topic: like ‘big data’ before it, ‘deep learning’ is not something new and is not deserving of a special name. Networks with more hidden layers to detect higher-order features, networks of different

types chained together in order to play to their strengths, graphs of networks to represent a probabilistic model.

This is a theoretical book, but it can be read in tandem with Hands-On Machine Learning with Scikit-Learn and TensorFlow, almost chapter-for-chapter. The Scikit-Learn and Tensorflow example code, while only moderately interesting on its own, helps to clarify the purpose of many of the topics in the Goodfellow book.

U04.1.4: Deep Learning: <https://www.goodreads.com/en/book/show/24072897-deep-learning>

## Hands-On Machine Learning with Scikit-Learn and TensorFlow

This book provides a great introduction to machine learning for both developer and non-developers. Authors suggest to just go through even if you don't understand math details. Highlights of this book are:

- Extraction of field expert knowledge is very important, you should know which model will serve better for the given solution. Luckily, a lot of models are available already from other scientists.
- Training data is the most important part, the more you have it the better. So if you can you should accumulate as much data as you can, preferably categorized, you may not still know how you will apply the accumulated data in the future but you will need it.
- Labelling training data is very important too, to train neural network you need to have at least thousands of labelled data samples, the more the better.
- Machine learning algorithms and neural networks are pretty common for years but the latest breakthrough is possible because of new optimization, new autoencoders ( that may help to artificially

generate training data) allowing to do training faster and with fewer data.

- Machine learning is still pretty time and resources consuming process. To train a machine learning model you need to know how to tweak parameters and how to use different training approaches fitting the particular model.

The book demonstrates (including the code) different approaches using Scikit-Learn python package and also the TensorFlow.

U04.1.5: Hands-On Machine Learning with Scikit-Learn and TensorFlow:

<https://www.goodreads.com/book/show/32899495-hands-on-machine-learning-with-scikit-learn-and-tensorflow>

## Data Science for Business

This is probably the most practical book to read if you are looking for an overview of data science. Either you know when terms like k-means and ROC curves are to be used or you have some context when you start digging deeper into how some of these algorithms are implemented. You will find it at the right level because there is just enough math to explain the fundamental concepts and make them stick in your head.

This isn't a book on implementing these concepts or a bunch of algorithms. This gives the book the advantage of being something you can refer to an intelligent manager or interested developer, and they can both get a lot out of it. And if they are interested in the next level of learning there are plenty of pointers. You will also find the chapter on presenting results through ROC curves, lift curves, etc. pretty interesting. It would be cool if this book had some more hands-on, but you can go to Kaggle and browse around the current and past competitions to apply what you learn here.

U04.1.6: Data Science for Business: <https://www.goodreads.com/book/show/17912916-data-science-for-business>

## 4.2 Courses to Attend

The screenshot displays two course pages from different platforms:

- Coursera:** Shows the "Machine Learning" course by Stanford University. It has a rating of 4.9 stars based on 103,013 ratings and 25,588 reviews. Enrollment starts on July 08. A large blue watermark "Analytics.com" runs diagonally across the page.
- fast.ai:** Shows the "Machine Learning" course offered by fast.ai in collaboration with Stanford. It has a rating of 4.9 stars based on 103,013 ratings and 25,588 reviews. Enrollment starts on July 08. A "Financial aid available" link is present.
- Kaggle:** Shows the "Deep Learning Specialization" course. It has a rating of 4.9 stars based on 219,522 already enrolled. Enrollment starts on June 30. A "Financial aid available" link is present.
- fast.ai in the news:** A sidebar lists several news articles about fast.ai, including:
  - The Economist: [New schemes teach the classes to code](#)
  - MIT Tech Review: [The startup diversity of AI tools just “bothered”](#)
  - The New York Times: [Finally, a Machine That Can Teach Your Software](#)
  - The Verge: [An AI speed test shows how coders can still beat tech giants like Google and Intel](#)

### Courses to Attend

Utilizing the Resources

## Machine Learning

The screenshot shows the "Machine Learning" course page on Coursera:

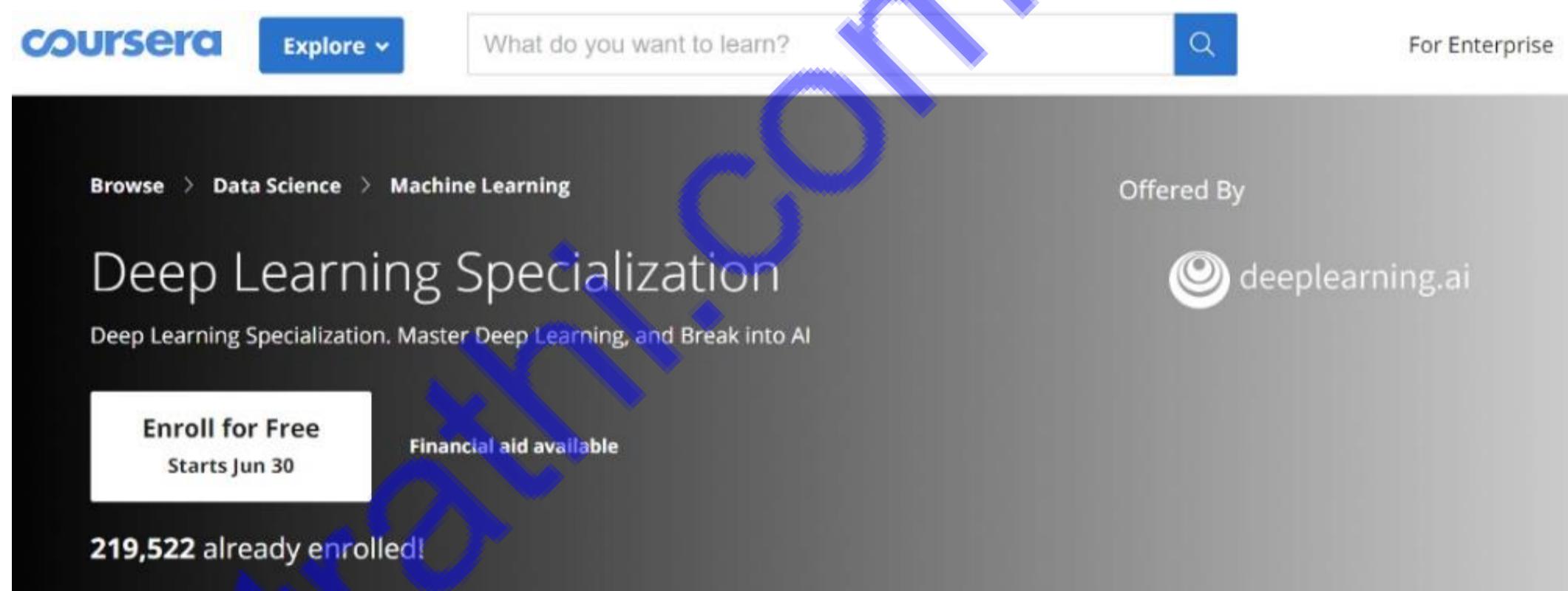
- Coursera:** The course is offered by Stanford University.
- Rating:** 4.9 stars based on 103,013 ratings and 25,588 reviews.
- Enrollment:** Starts Jul 08. Enrollment count: 2,375,041 already enrolled!
- Financial aid available:** Link provided.

Machine Learning is one of the first programming MOOCs. Coursera put online by Coursera founder and Stanford Professor Andrew Ng. This course assumes that you have basic programming skills and you have some understanding of Linear Algebra. Knowledge of Statistics & Probability is not required though.

Andrew Ng does a good job explaining dense material and slides. The course gives you a lot of structure and direction for each homework, so it is generally pretty clear what you are supposed to do and how you are supposed to do it.

U04.2.1: Machine Learning: <https://www.coursera.org/learn/machine-learning>

## Deep Learning

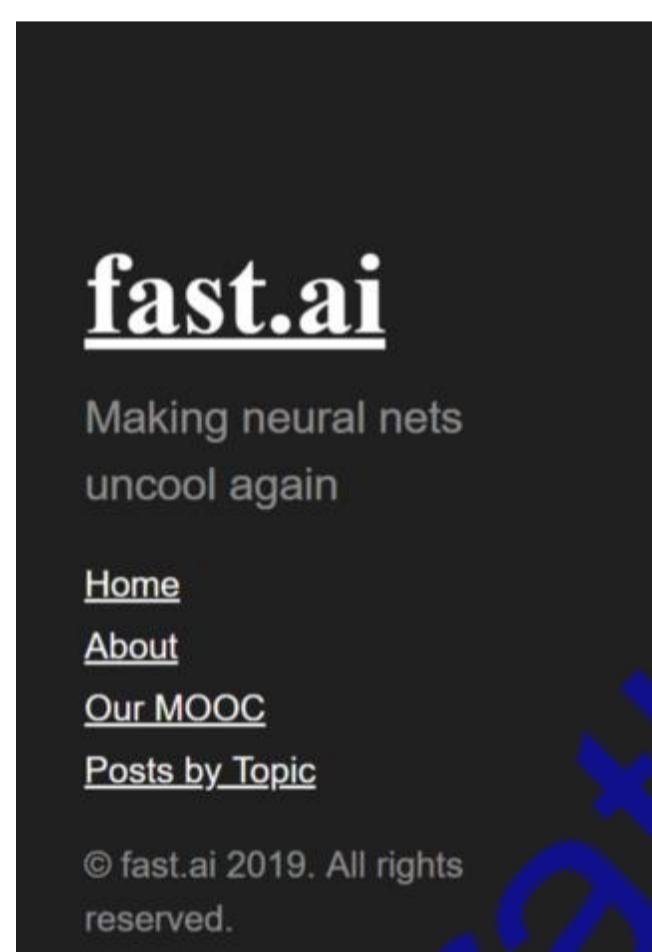


When you are rather new to the topic, you can learn a lot of doing the deeplearning.ai specialization. First and foremost, you learn the basic concepts of NN. How does a forward pass in simple sequential models look like, what's a backpropagation, and so on? I experienced this set of courses as a very time-effective way to learn the basics and worth more than all the tutorials, blog posts and talks, which I went through beforehand.

Doing this specialization is probably more than the first step into DL. I would say, each course is a single step in the right direction, so you end up with five steps in total. I think it builds a fundamental understanding of the field. But going further, you have to practice a lot and eventually it might be useful also to read more about the methodological background of DL variants. But doing the course work gets you started in a structured manner — which is worth a lot, especially in a field with so much buzz around it.

U04.2.2: Deep Learning: <https://www.deeplearning.ai/deep-learning-specialization/>

## Fast AI



**Our online courses (all are free and have no ads):**

- [Practical Deep Learning for Coders](#)
- [Cutting Edge Deep Learning for Coders](#)
- [Introduction to Machine Learning for Coders](#)
- [Computational Linear Algebra](#)

**Our software: [fastai v1 for PyTorch](#)**

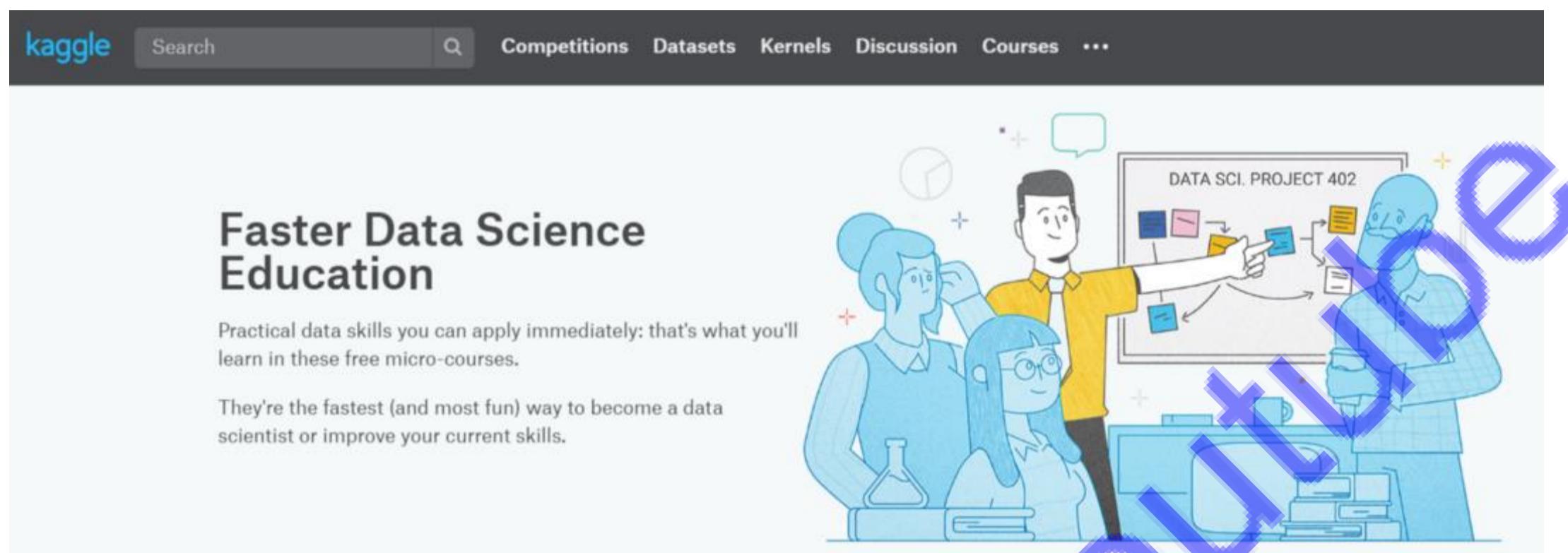
**fast.ai in the news:**

- The Economist: [New schemes teach the masses to build AI](#)
- MIT Tech Review: [The startup diversifying AI workforce beyond just "techies"](#)
- The New York Times: [Finally, a Machine That Can Finish Your Sentence](#)
- The Verge: [An AI speed test shows clever coders can still beat tech giants like Google and Intel](#)

If your goal is to be able to learn about deep learning and apply what you've learned, the fast.ai course is a better bet. If you have the time, interleaving the deeplearning.ai and fast.ai courses is ideal — you get the practical experience, applicability, and audience interaction of fast.ai, along with the organised material and theoretical explanations of deeplearning.ai.

U04.2.3: Fast AI: <https://www.fast.ai/>

# Kaggle Learn



Practical data skills you can apply immediately: that's what you'll learn in these free micro-courses. They're the fastest (and most fun) way to become a data scientist or improve your current skills.

U04.2.4: Kaggle Learn: <https://www.kaggle.com/learn/overview>

## 4.3 Blogs to follow

A screenshot of a blog homepage with a dark blue header. The header includes the 'KDnuggets' logo, social media links (Twitter, Facebook, LinkedIn), and a 'Contact' button. Below the header, there's a banner for 'Analytics Vidhya' with the tagline 'Discover, understand, and catalog your data'. The main content area has a dark blue background with white text. It features three article thumbnails: 'Machine Learning, Data Mining, Big Data', 'Generating Beatles' Lyrics with Machine Learning', and 'Portable Computer Vision: Tensorflow 2.0 on a Raspberry Pi'.

### Blogs to Follow

Utilizing the Resources

# KD Nuggets

The screenshot shows the KDnuggets homepage. At the top, there's a yellow header bar with the KDnuggets logo, social media links (Twitter, Facebook, LinkedIn), and a 'Contact' link. Below the header is a navigation bar with links to 'SOFTWARE | News/Blog | Top stories | Opinions | Tutorials | JOBS | Companies | Courses | Datasets | E...' and a search bar. A large banner in the center says 'Discover, understand, and catalog your data' with a subtext 'Smart Data Discovery - AI-driven Catalog'. To the right of the banner is a logo for 'Io-Tahoe'. Below the banner, a blue banner reads 'Discover, understand, and catalog your data with Io-Tahoe. Smart Data Discovery - AI-...'. On the left side, there's a sidebar with categories like 'Machine Learning, Data Science, Data Mining, Big Data, Analytics, AI', 'Software (Suites, Text, Visualization)', 'Jobs - Industry | Academic Meetings, Conferences', 'Companies (Consulting, Products)', 'Courses in Big Data, Data Science', 'Datasets (APIs/Markets, Gov)', and 'Data Mining Course | Gregory Piatetsky'. On the right, there's a 'Most Recent' section with a list of articles:

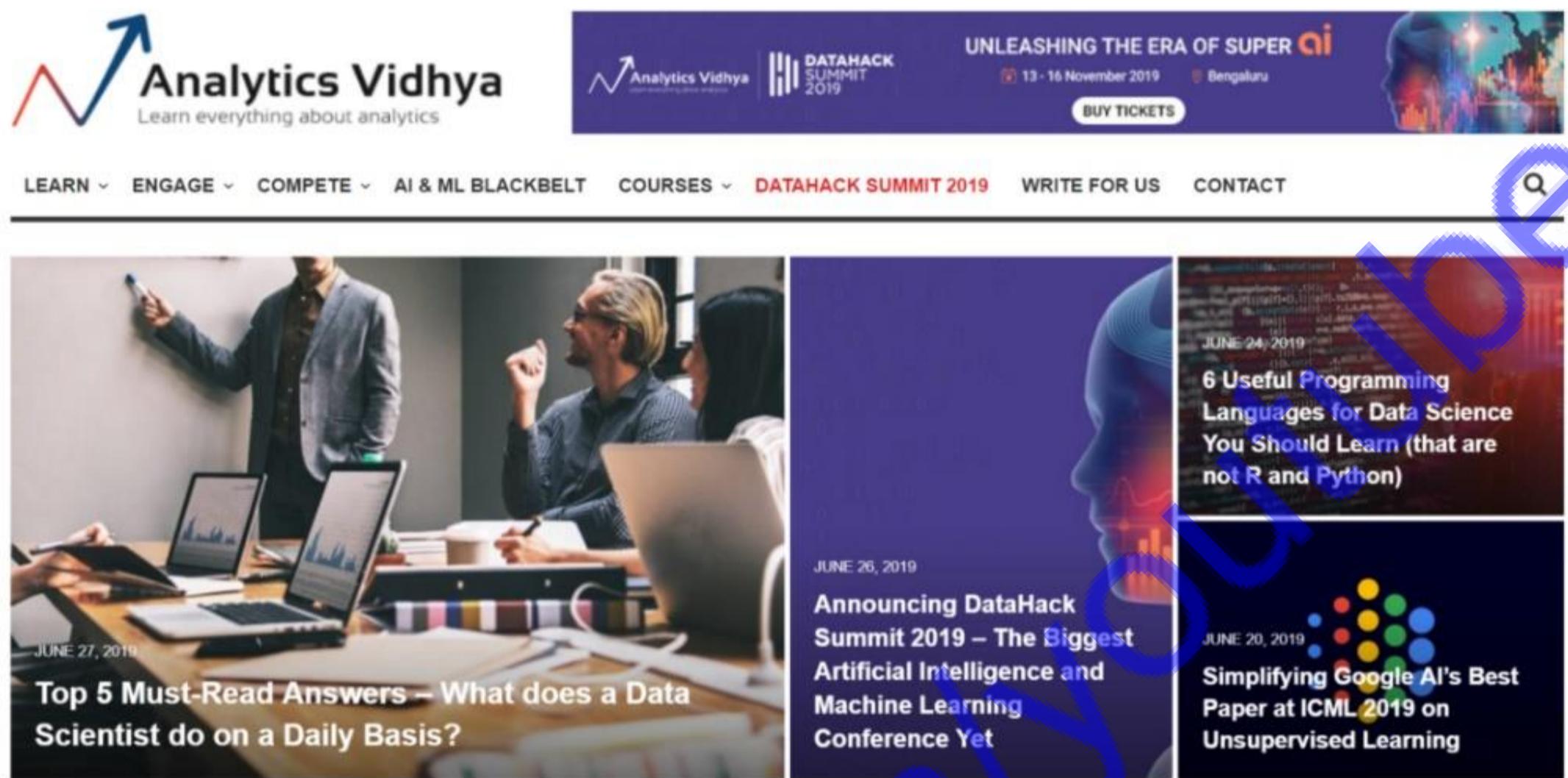
- Make your Data Talk!
- An Overview of Human Pose Estimation with Deep Learning
- RS21: Senior Data Scientist [Albuquerque, NM]
- How To Get Funding For AI Startups
- PySyft and the Emergence of Private Deep Learning
- An Overview of Outlier Detection Methods from PyOD R...

At the bottom right, there's a green bar with the text 'MS IN DATA SCIENCE at the' and a small logo.

KDnuggets is a leading site on AI, Analytics, Big Data, Data Mining, Data Science, and Machine Learning and is edited by Gregory Piatetsky-Shapiro and Matthew Mayo. KDnuggets was founded in February of 1997. Before that, Gregory maintained an earlier version of this site, called Knowledge Discovery Mine, at GTE Labs (1994 to 1997).

U04.3.1: KD Nuggets: <https://www.kdnuggets.com/>

# Analytics Vidhya



Analytics Vidhya provides a community-based knowledge portal for Analytics and Data Science professionals. The aim of the platform is to become a complete portal serving all knowledge and career needs of Data Science Professionals.

U04.3.2: Analytics Vidhya Blog: <https://www.analyticsvidhya.com/blog/>

# Towards Data Science



Generating Beatles' Lyrics with Machine Learning



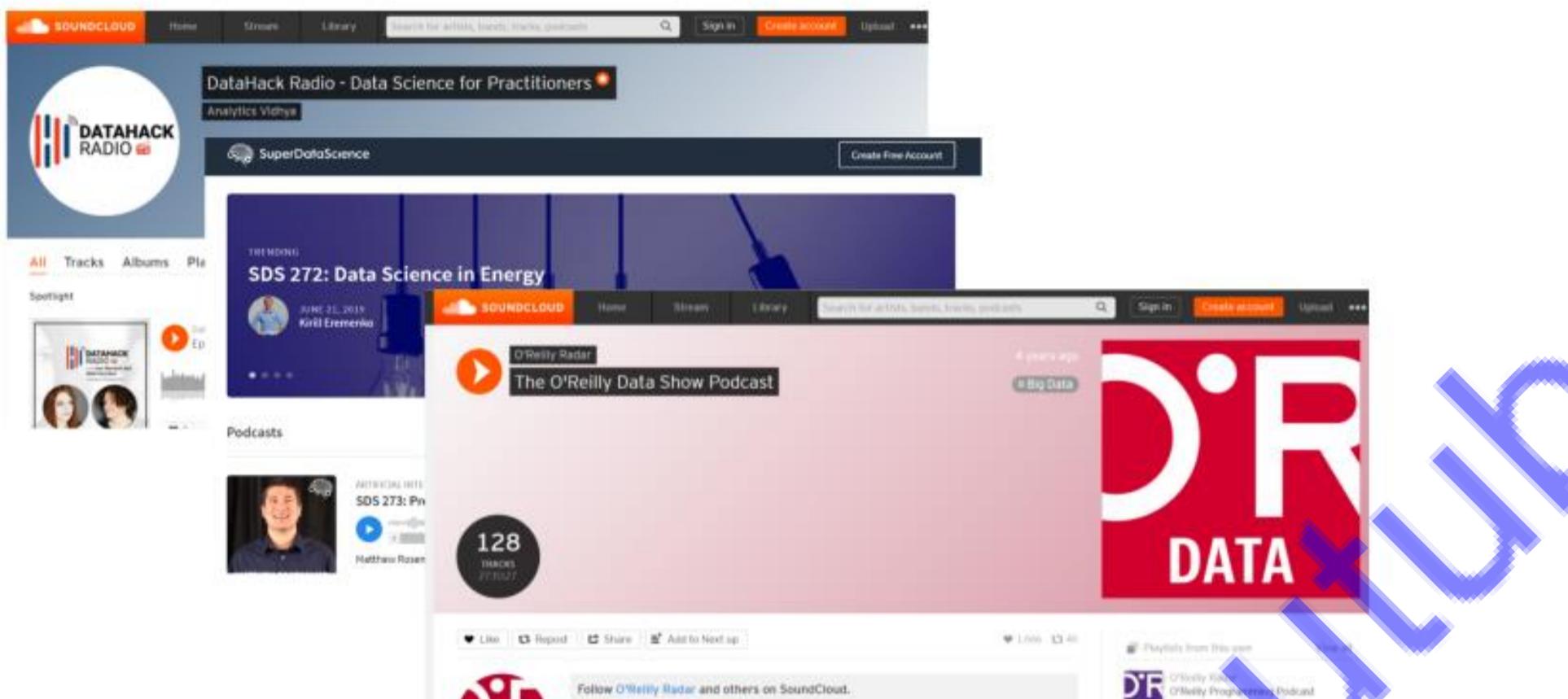
Portable Computer Vision: Tensorflow 2.0 on a Raspberry Pi

TDS joined Medium's vibrant community in October 2016. In the beginning, their goal was simply to gather good posts and distribute them to a broader audience. Just a few months later, they were pleased to see that they had a very fast-growing audience and many new contributors.

Today they are working with more than 10 Editorial Associates to prepare the most exciting content for our audience. They provide customized feedback to our contributors using Medium's private notes. This allows them to promote their latest articles across social media without the added complexity that they might encounter using another platform.

U04.3.3: Towards Data Science: <https://towardsdatascience.com/>

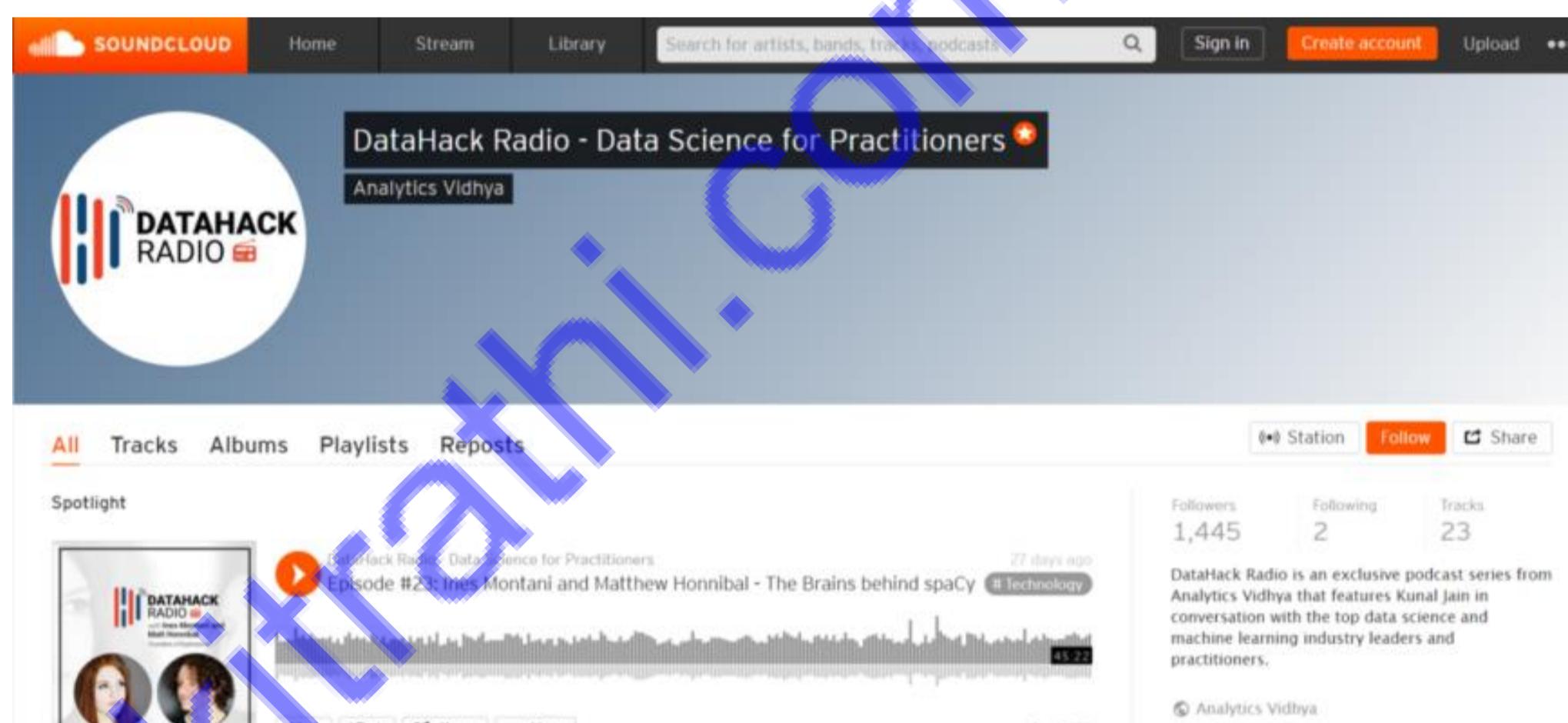
## 4.4 Podcasts to listen



### Podcasts to Listen

Utilizing the Resources

### Data Hack



This is Analytics Vidhya's exclusive podcast series which will feature top leaders and practitioners in the data science and machine learning industry.

So in every episode of DataHack Radio, they bring you discussions with one such thought leader in the industry. They have discussions about

their journey, their learnings and plenty of other data science-related things.

U04.4.1: Data Hack Radio: <https://datahack.analyticsvidhya.com/>

## Super Data Science

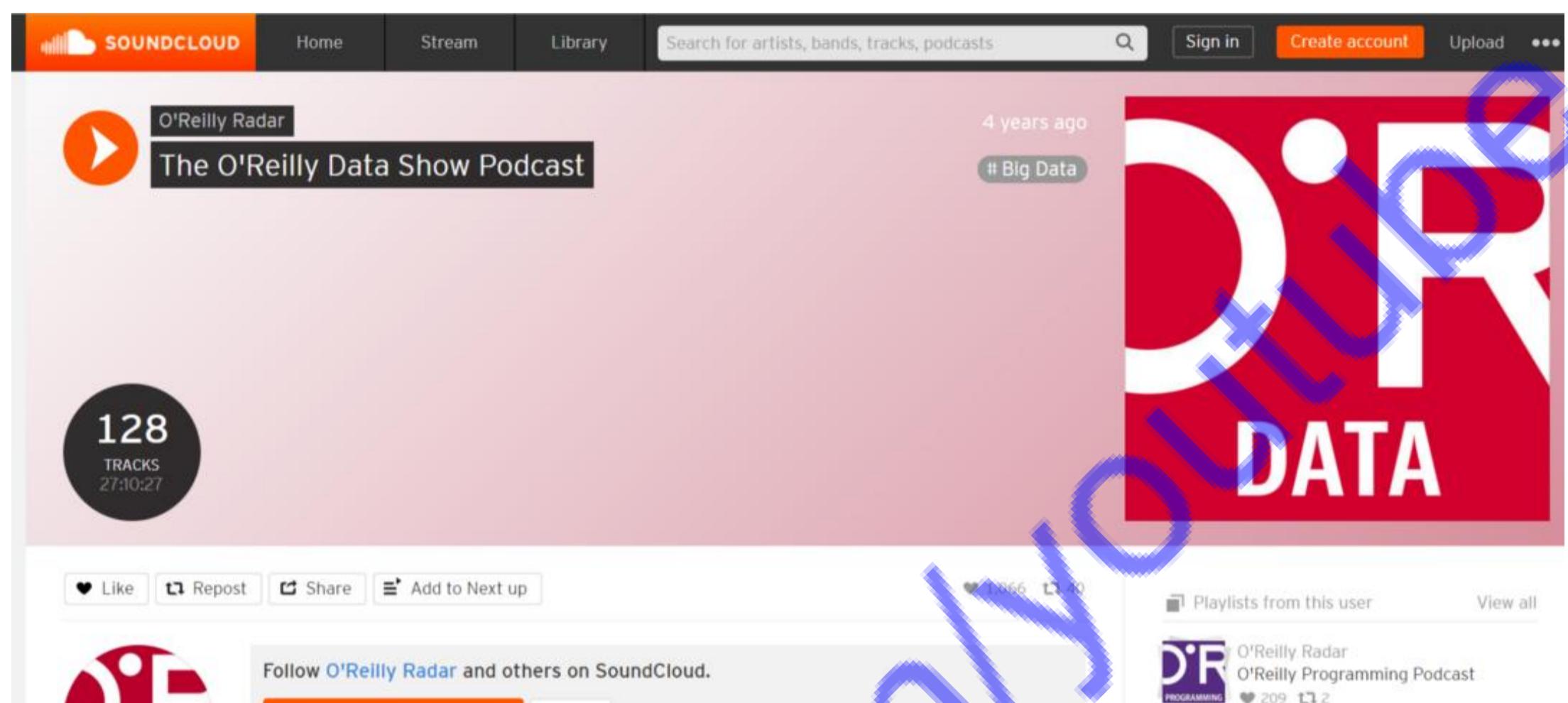
The screenshot shows the homepage of the Super Data Science website. At the top, there's a dark header with the logo 'SuperDataScience' and a 'Create Free Account' button. Below the header is a large, dark purple rectangular area featuring the text 'TRENDING' and 'SDS 272: Data Science in Energy'. It includes a small circular profile picture of Kirill Eremenko, the date 'JUNE 21, 2019', and his name 'Kirill Eremenko'. Below this main section, there's a navigation bar with 'Podcasts' on the left, a search bar in the center containing 'Search for a Podcast...', and dropdown menus for 'All Guests', 'All Categories', and 'Recent'. The main content area below the navigation bar displays another podcast episode: 'ARTIFICIAL INTELLIGENCE, DATA SCIENCE • 56 MINS' titled 'SDS 273: Predict, Prevent, Detect: Cyber Security'. It features a photo of Matthew Rosenquist, the date 'June 26, 2019', and the text 'Invited Guest'. To the right of this episode, there's a 'Recommendations for you' sidebar with a thumbnail for 'SDS 271: Making the Public Graphically Literate' by Alberto Cairo, which is 65 mins long.

*Kirill Eremenko* is a Data Science coach and lifestyle entrepreneur. The goal of the *Super Data Science* podcast is to bring you the most inspiring Data Scientists and Analysts from around the World to help you build your successful career in Data Science.

Data is growing exponentially and so are salaries of those who work in analytics. This podcast can help you learn how to skyrocket your analytics career. Big Data, visualization, predictive modelling, forecasting, analysis, business processes, statistics, R, Python, SQL programming, tableau, machine learning, Hadoop, databases, data science MBAs, and all the analytics tools and skills that will help you better understand how to crush it in Data Science.

U04.4.2: Super Data Science: <https://www.superdatascience.com/podcast>

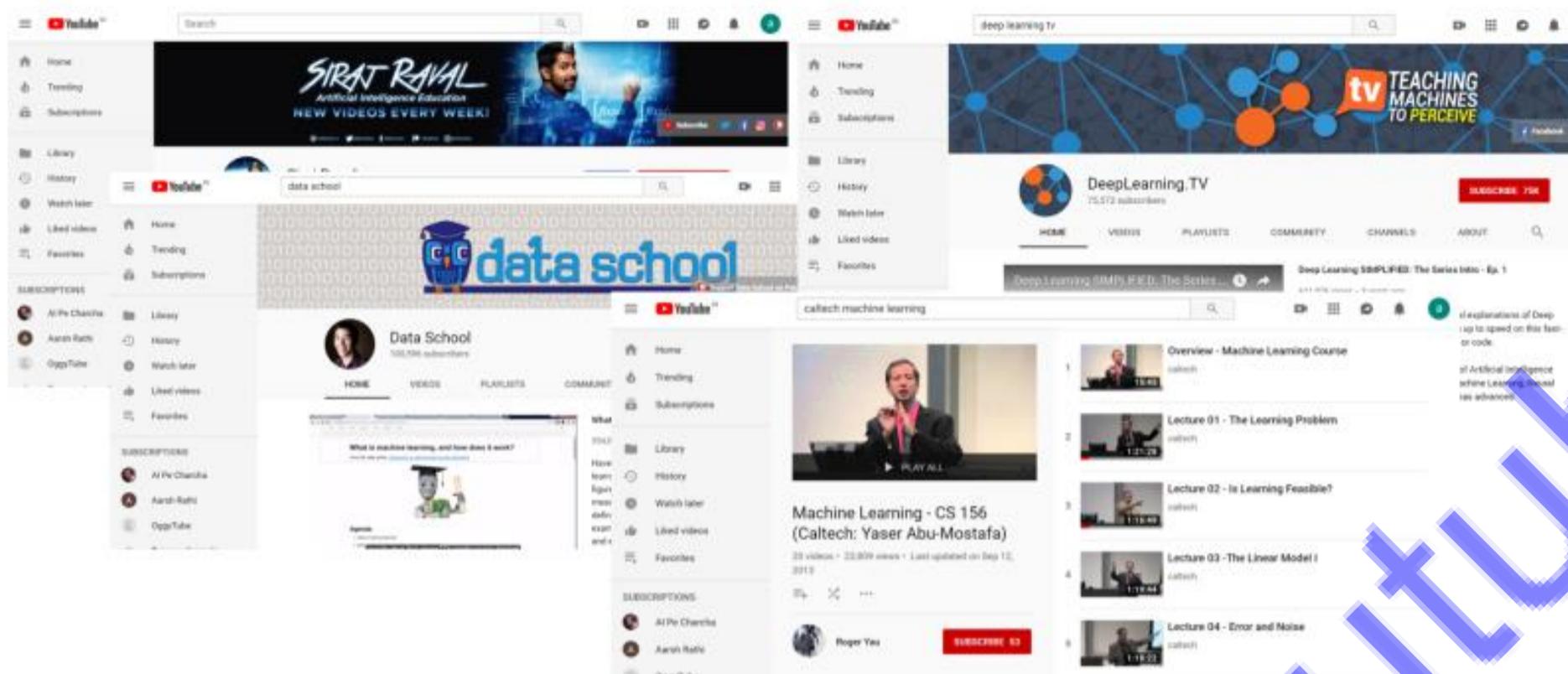
## The O'Reilly Data Show Podcast



Known as the father of all other data shows, “the O'Reilly Data Show” features Ben Lorica, O'Reilly Media's chief data scientist. Lorica conducts interviews with other experts about big data and data science current affairs. While it does get technical and may not be the best place for a beginner to start, it provides interesting insights into the future of the data science industry.

U04.4.3: The O'Reilly Data Show Podcast: <https://www.oreilly.com/topics/oreilly-data-show-podcast>

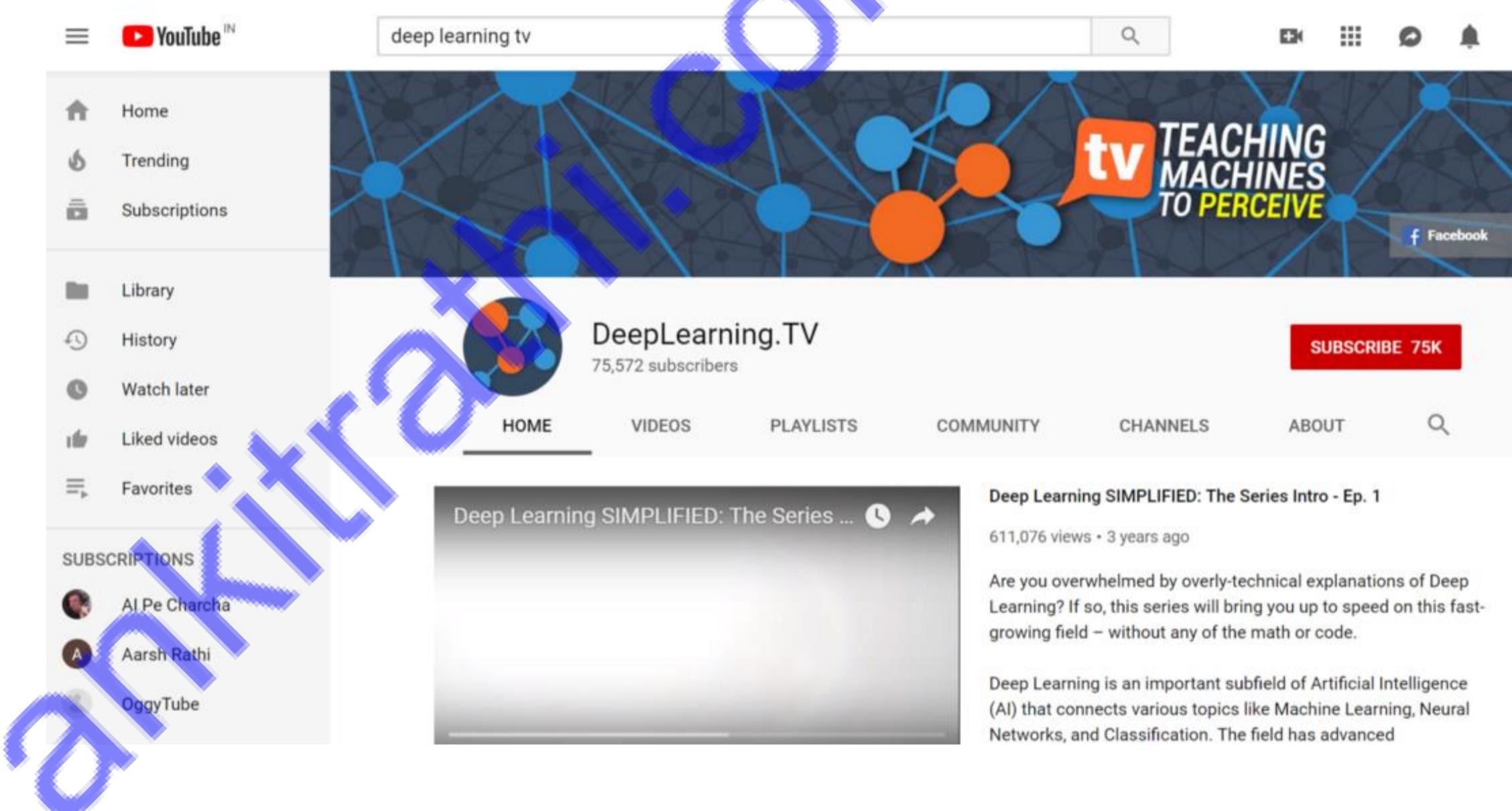
## 4.5 YouTube Channels



### YouTube Channels

Utilizing the Resources

### DeepLearning.TV



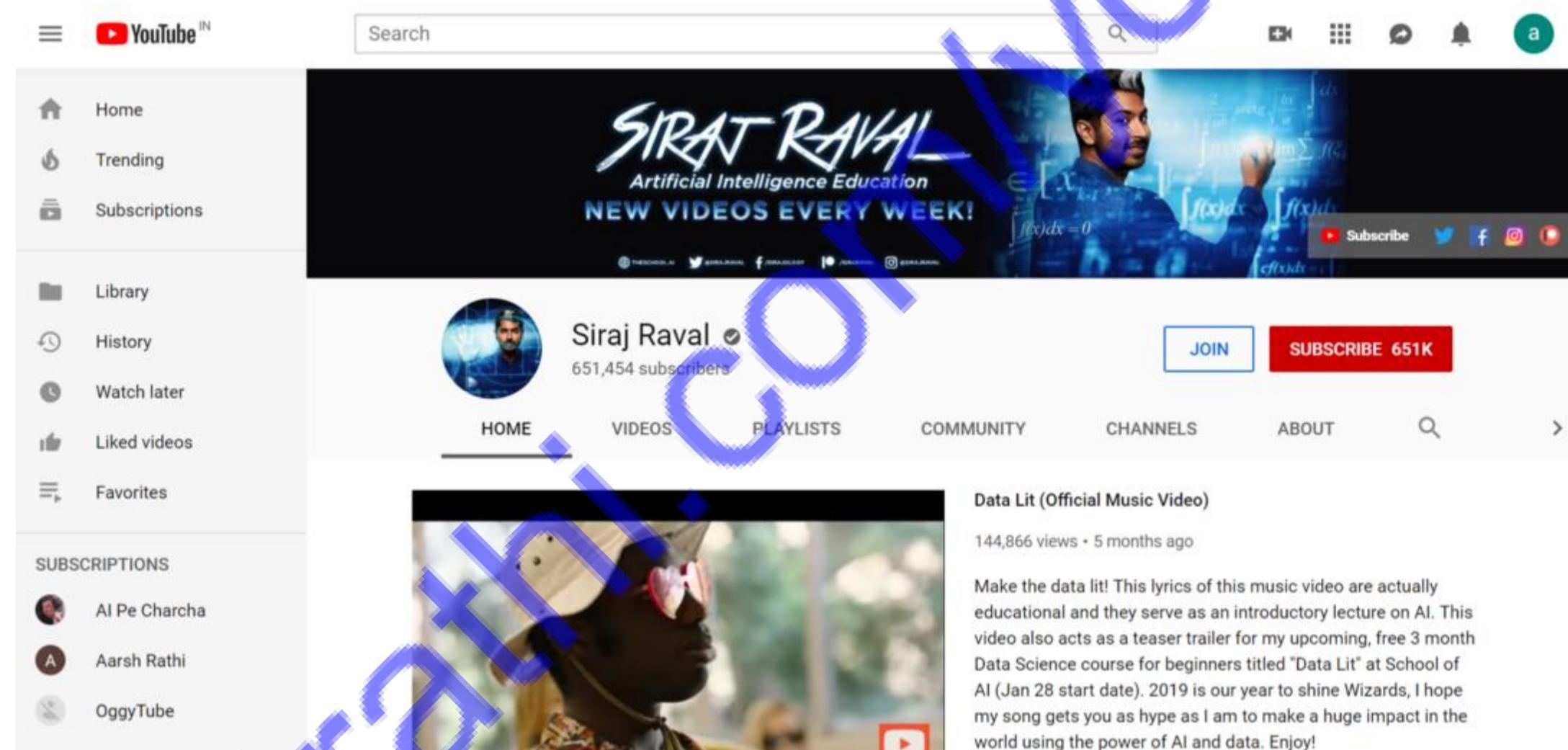
DeepLearning.TV is all about Deep Learning, the field of study that teaches machines to perceive the world. Starting with a series that simplifies Deep Learning, the channel features topics such as How To's, reviews of software libraries and applications, and interviews with key

individuals in the field. Through a series of concept videos showcasing the intuition behind every Deep Learning method, they show you that Deep Learning is actually simpler than you think. Their goal is to improve your understanding of the topic so that you can better utilize Deep Learning in your own projects. They provide a window into the cutting edge of Deep Learning and bring you up to speed on what's currently happening in the field.

U04.5.1: DeepLearning.TV:

<https://www.youtube.com/playlist?list=PLJjh1vISEYgvGod9wWiydumYl8hOXixNu>

## Sirajology



Your host here is Siraj. He is on a warpath to inspire and educate developers to build Artificial Intelligence. Games, music, chatbots, art, he teaches you how to make it all yourself. This is the fastest-growing AI community in the world. Their mission: Solve AI. Use it to benefit humanity.

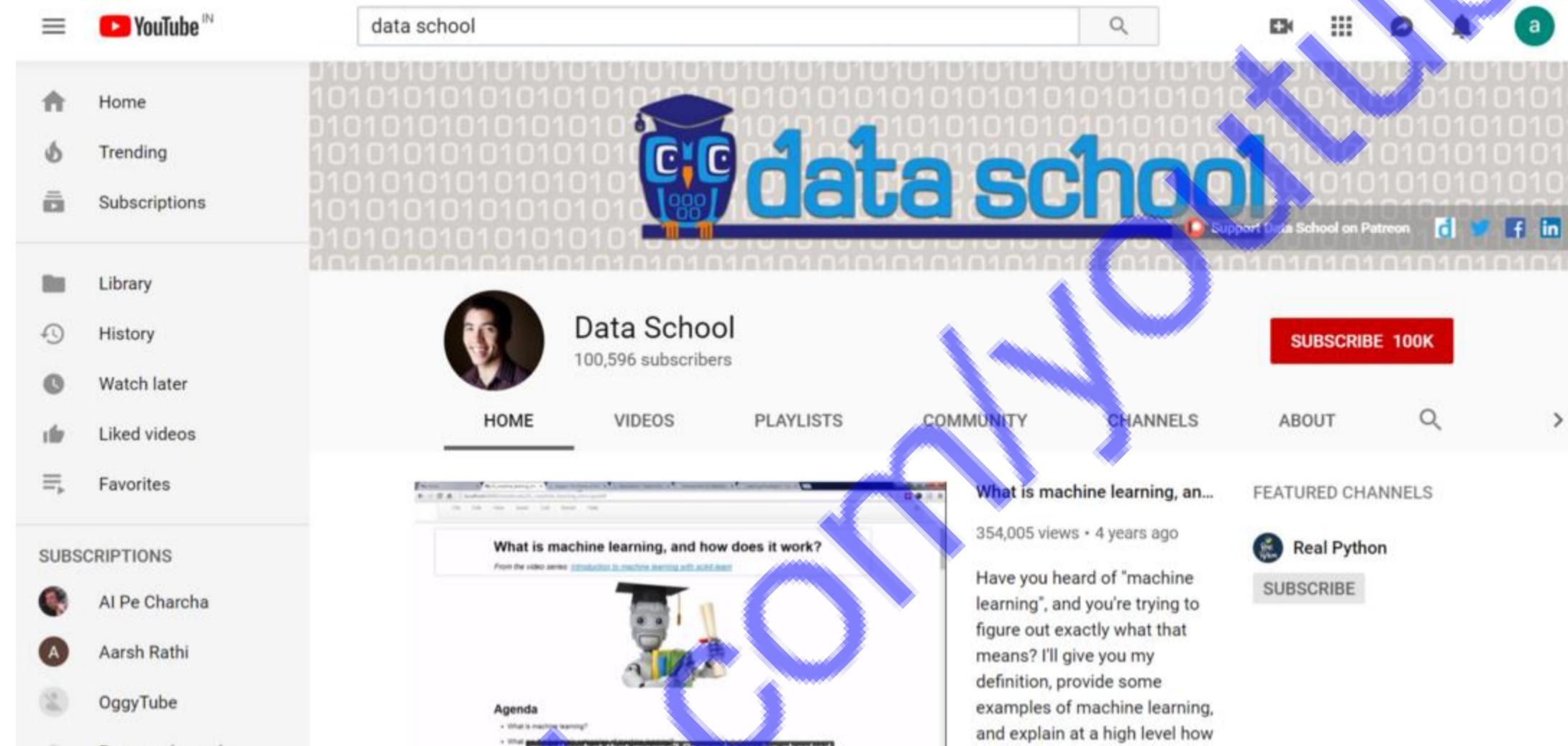
He is an AI Researcher, his latest paper is here —

<https://drive.google.com/file/d/0BwUv84INDk72Q1gzaXgwR2U3U2NWVIZSOFk4amZIRmV1QXIO/view>

He is also a Data Scientist, AI Educator, Rapper, Author, and Director of the School of AI ([www.theschool.ai](http://www.theschool.ai))

U04.5.2: Sirajology: <https://www.youtube.com/channel/UCWN3xxRkmTPmbKwht9FuE5A>

## Data School

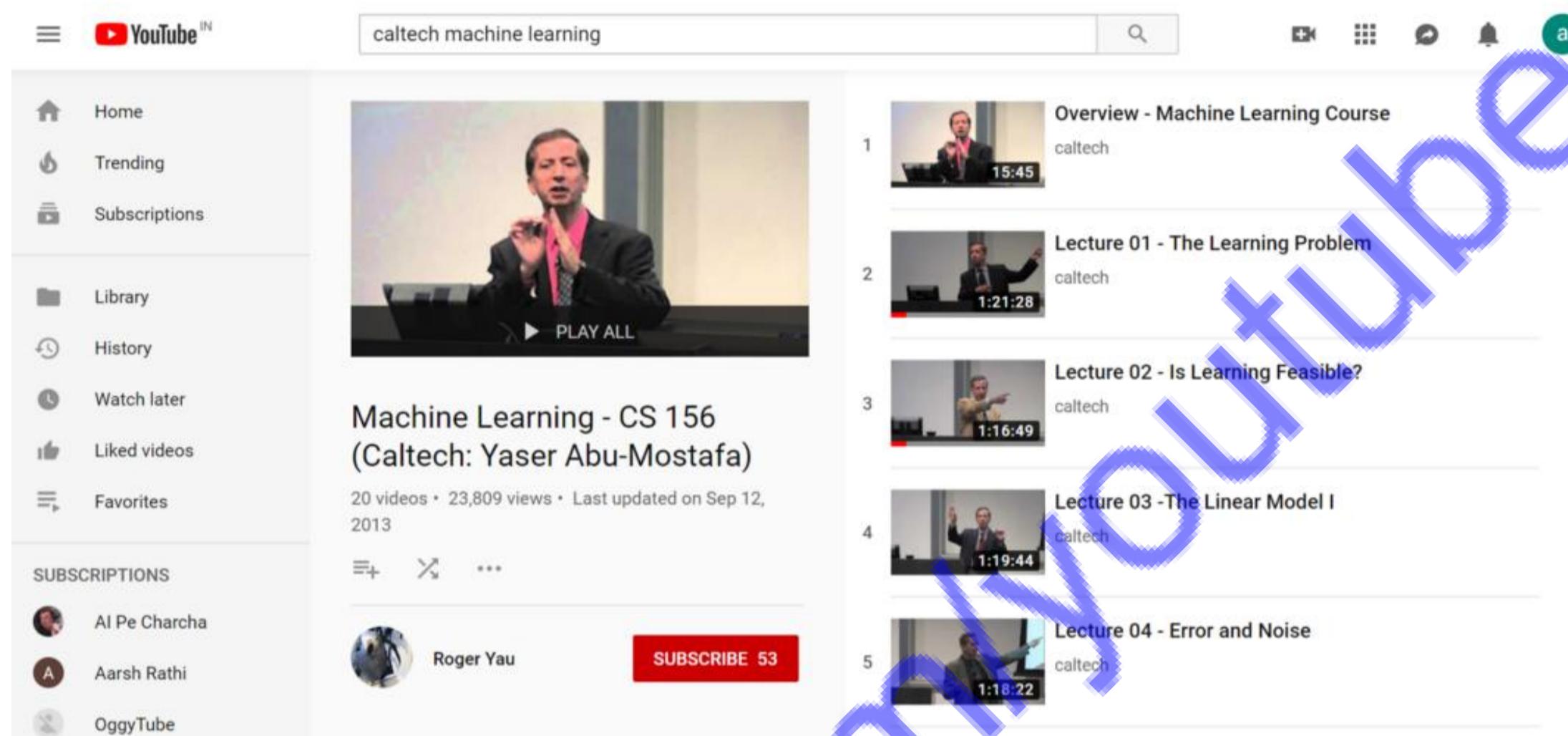


Are you trying to learn data science so that you can get your first data science job? You're probably confused about what you're "supposed" to learn, and then you have the hardest time actually finding lessons you can understand! Data School focuses you on the topics you need to master first, and offers in-depth tutorials that you can understand regardless of your educational background.

Your host here is Kevin Markham, and he is the founder of Data School. He has taught data science using the Python programming language to hundreds of students in the classroom, and hundreds of thousands of students (like you) online. Finding the right teacher was so important to his data science education, and so he sincerely hopes that he can be the right data science teacher for you.

U04.5.3: Data School: <https://www.youtube.com/dataschool>

## Caltech Machine Learning



This is an introductory course by Caltech Professor Yaser Abu-Mostafa on machine learning that covers the basic theory, algorithms, and applications. Machine learning (ML) enables computational systems to adaptively improve their performance with experience accumulated from the observed data. ML techniques are widely applied in engineering, science, finance, and commerce to build systems for which we do not have a full mathematical specification (and that covers a lot of systems). The course balances theory and practice and covers the mathematical as well as the heuristic aspects.

U04.5.4: CalTech Machine Learning: <https://www.youtube.com/playlist?list=PLD63A284B7615313A>

## 4.6 GitHub Repos

The screenshot shows the homepage of the [Awesome Data Science](#) GitHub repository. It features a sidebar with a table of contents containing links to various resources like Motivation, Infographic, What is Data Science?, Colleges, MOOC's, Data Sets, Bloggers, Newsletters, Podcasts, Books, Facebook Accounts, Twitter Accounts, YouTube Videos & Channels, Telegram Channels, and Toolboxes - Environment. The main content area displays the [cheatsheets-ai](#) repository, which provides essential cheat sheets for deep learning and machine learning researchers. Below it is a section titled [HackerMath for Machine Learning](#), which includes a video thumbnail and a brief description of the workshop.

### GitHub Repos

Utilizing the Resources

## Awesome Data Science

### Awesome Data Science

An open source Data Science repository to learn and apply towards solving real world problems.

#### Table of contents

- [Motivation](#)
- [Infographic](#)
- [What is Data Science?](#)
- [Colleges](#)
- [MOOC's](#)
- [Data Sets](#)
- [Bloggers](#)
- [Newsletters](#)
- [Podcasts](#)
- [Books](#)
- [Facebook Accounts](#)
- [Twitter Accounts](#)
- [YouTube Videos & Channels](#)
- [Telegram Channels](#)
- [Toolboxes - Environment](#)

This Repo answer the questions, “What is Data Science and what should you study to learn Data Science?” An awesome Data Science repository to learn and apply for real-world problems.

As the aggregator says, “Our favourite data scientist is Clare Corthell. She is an expert in data-related systems and a hacker and has been working on a company as a data scientist. Clare’s blog. This website helps you to understand the exact way to study as a professional data scientist.”

“Secondly, Our favourite programming language is Python nowadays for Data Science. Python’s — Pandas library has full functionality for collecting and analyzing data. We use Anaconda to play with data and to create applications.”

U04.6.1: Awesome Data Science: <https://github.com/bulutyazilim/awesome-datasience>

## Essential Cheat Sheets for Machine Learning and Deep Learning Engineers

### cheatsheets-ai

Essential Cheat Sheets for deep learning and machine learning researchers

- [Tensorflow](#)
- [Keras](#)
- [Neural Networks Zoo](#)
- [Numpy](#)
- [Scipy](#)
- [Pandas-1](#)
- [Pandas-2](#)
- [Pandas-3](#)
- [Scikit-learn](#)
- [Matplotlib](#)
- [ggplot2-1](#)
- [ggplot2-2](#)
- [PySpark](#)
- [PySpark-RDD](#)
- [PySpark-SQL](#)
- [R Studio\(dplyr & tidyr\)-1](#)
- [R Studio\(dplyr & tidyr\)-2](#)

Machine learning is complex. For newbies, starting to learn machine learning can be painful if they don’t have the right resources to learn from. Most of the machine learning libraries are difficult to understand

and the learning curve can be a bit frustrating. Kailash Ahirwar has created a repository on Github (cheatsheets-ai) containing cheatsheets for different machine learning frameworks, gathered from different sources. Have a look at the Github repository, also, contribute cheat sheets if you have any.

U04.6.2: Cheatsheets-AI: <https://github.com/kailashahirwar/cheatsheets-ai>

## HackerMath for Machine Learning

### HackerMath for Machine Learning

"Study hard what interests you the most in the most undisciplined, irreverent and original manner possible." — Richard Feynman

Math literacy, including proficiency in Linear Algebra and Statistics, is a must for anyone pursuing a career in data science. The goal of this workshop is to introduce some key concepts from these domains that get used repeatedly in data science applications. Our approach is what we call the "Hacker's way". Instead of going back to formulae and proofs, we teach the concepts by writing code. And in practical applications. Concepts don't remain sticky if the usage is never taught.

The focus will be on depth rather than breadth. Three areas are chosen - Hypothesis Testing, Supervised Learning and Unsupervised Learning. They will be covered to sufficient depth - 50% of the time will be on the concepts and 50% of the time will be spent coding them.

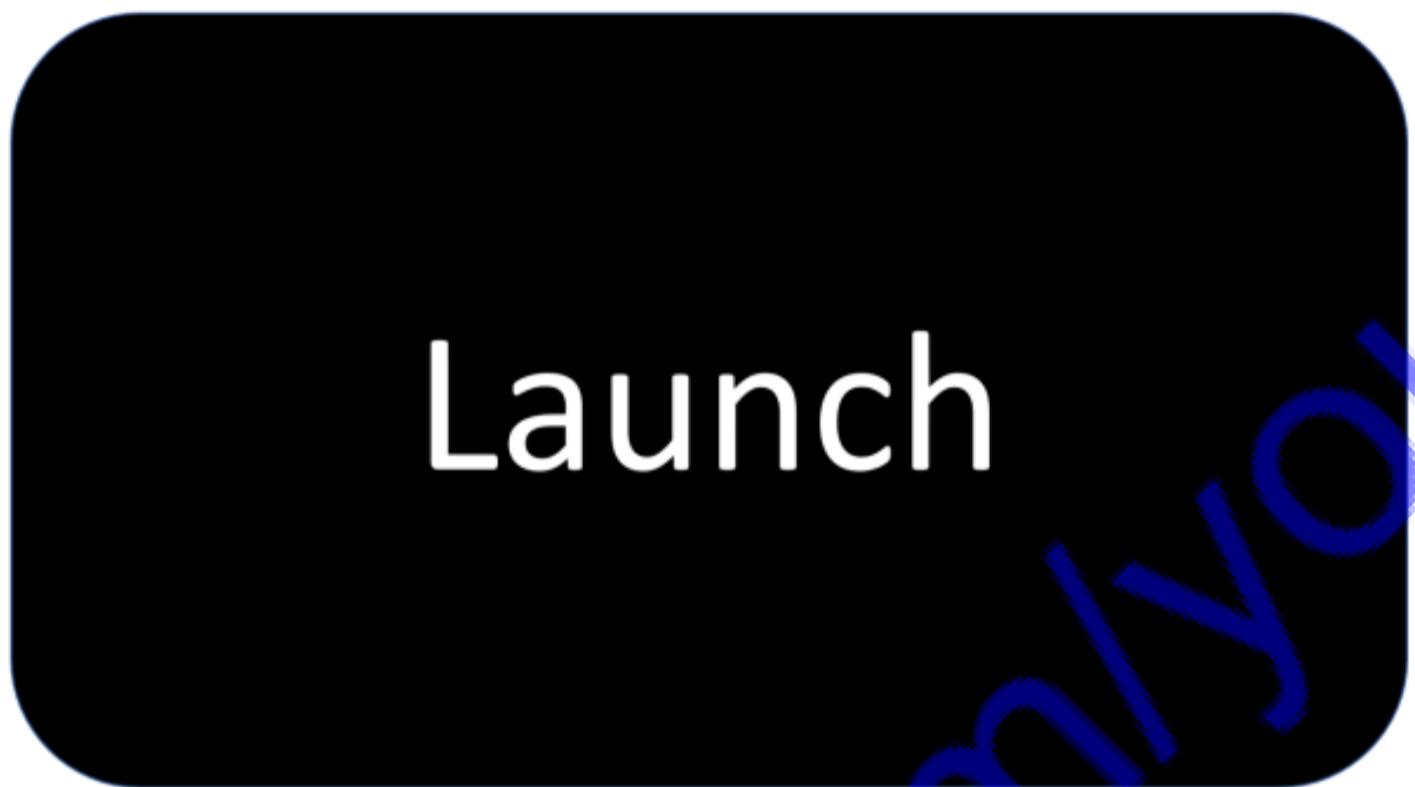


Math literacy, including proficiency in Linear Algebra and Statistics, is a must for anyone pursuing a career in data science. The goal of this workshop is to introduce some key concepts from these domains that get used repeatedly in data science applications.

As outlined by Amit Kapoor, “Our approach is what we call the ‘Hacker’s way’. Instead of going back to formulae and proofs, we teach the concepts by writing code. And in practical applications. Concepts don’t remain sticky if the usage is never taught.”

The focus here is on depth rather than breadth. Three areas are chosen — Hypothesis Testing, Supervised Learning and Unsupervised Learning. They are covered to sufficient depth — 50% of the time on the concepts and 50% of the time spent coding them.

U04.6.3: HackerMath for Machine Learning: <https://github.com/amitkaps/hackermath>



**DS/AI: Self-Starter Kit**  
Build Your Own Roadmap

ankitrathi.com/youtube

[ankitratthi.com/youtube](http://ankitratthi.com/youtube)

## Chapter 5

---

### Building your Portfolio

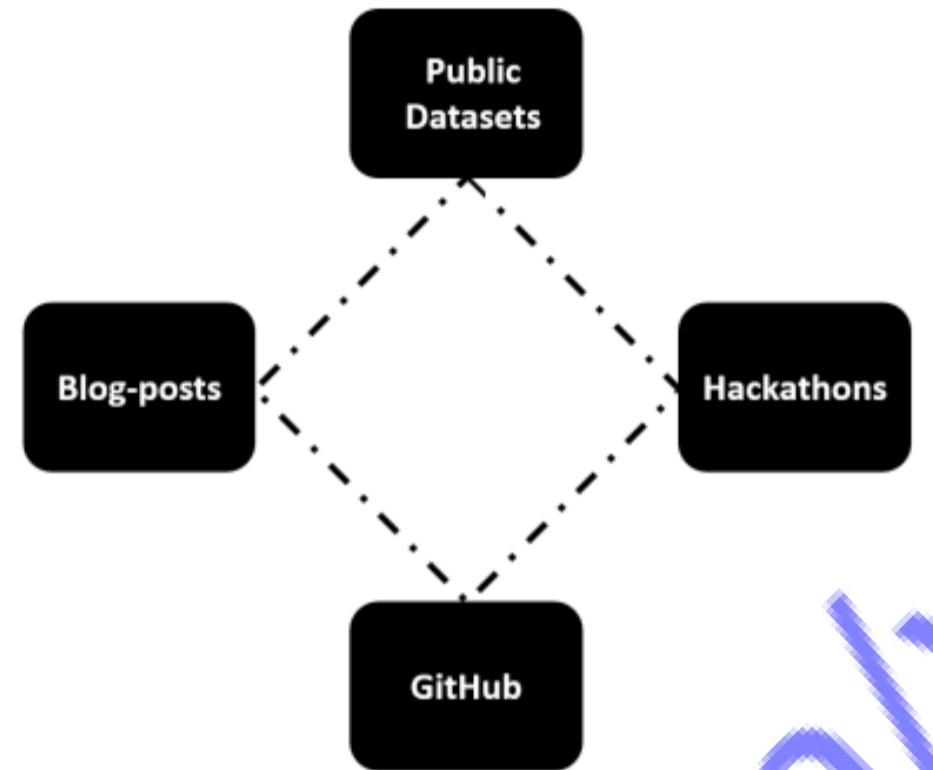


This chapter talks about how you can build your DS/AI portfolio. Lets first understand, why a portfolio is important in DS/AI field?

*Besides the benefit of learning by making a portfolio, a portfolio is important as it can help get you employment.*

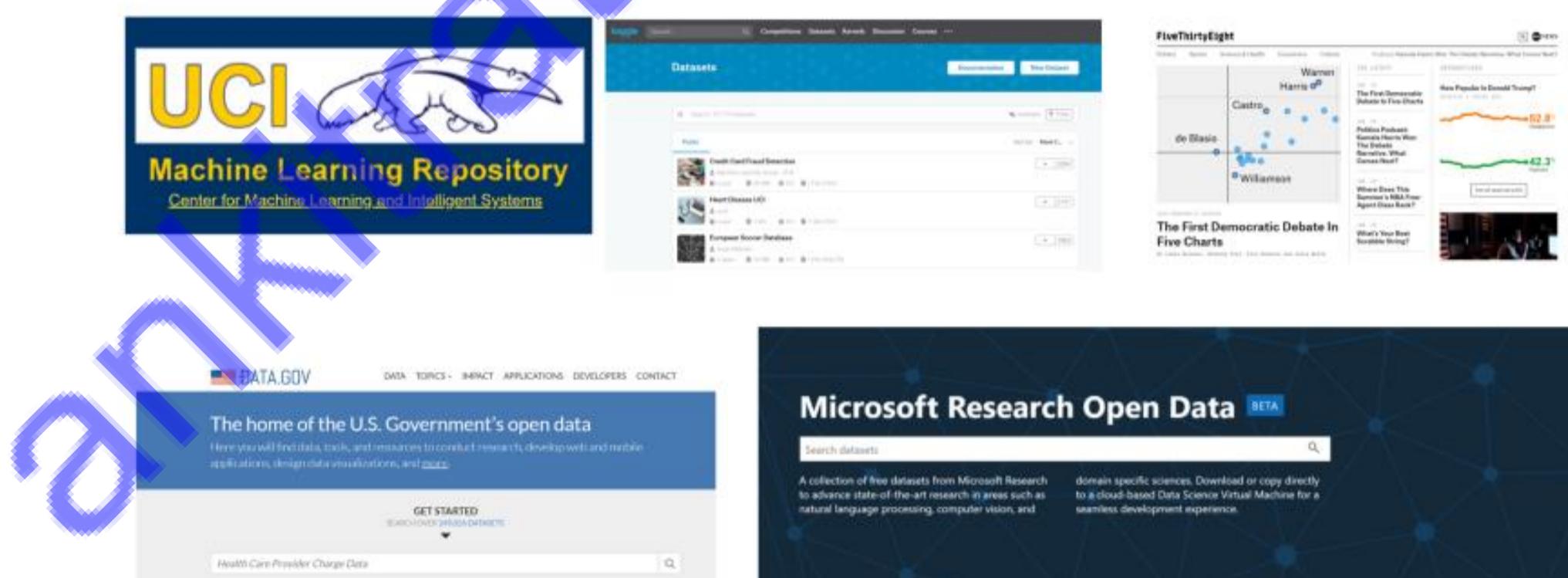
For the purpose of this article, let's define a portfolio as public evidence of your DS/AI skills.

People often forget that software engineers and data scientists also google their issues. If these same people have their problems solved by reading your public work, they might think better of you and reach out to you.



## Building your Portfolio DS/AI: Self-Starter Kit

### 5.1 Work on Public data-sets



## Work on Public data-sets

Building your Portfolio

*You can gain more DS/AI skills by working on prediction problems rather than getting stuck in endless learning loop.*

But you will not get a project to work on from day one of your learning. Still, there are platforms where you can apply and learn DS/AI.

UCI ML



The UCI Machine Learning Repository is a collection of data-sets that are used by the machine learning community for the analysis of machine learning algorithms. The archive was created as an FTP archive in 1987 by David Aha and fellow graduate students at UC Irvine. Since that time, it has been widely used by students, educators, and researchers all over the world as a primary source of machine learning data-sets. As an indication of the impact of the archive, it has been cited over 1000 times, making it one of the top 100 most cited “papers” in all of computer science. The current version of the web site was designed in 2007 by Arthur Asuncion

and David Newman, and this project is in collaboration with Rexa.info at the University of Massachusetts Amherst. Funding support from the National Science Foundation is gratefully acknowledged.

U05.1.1: UCI ML: <https://archive.ics.uci.edu/ml/index.php>

## Kaggle Datasets

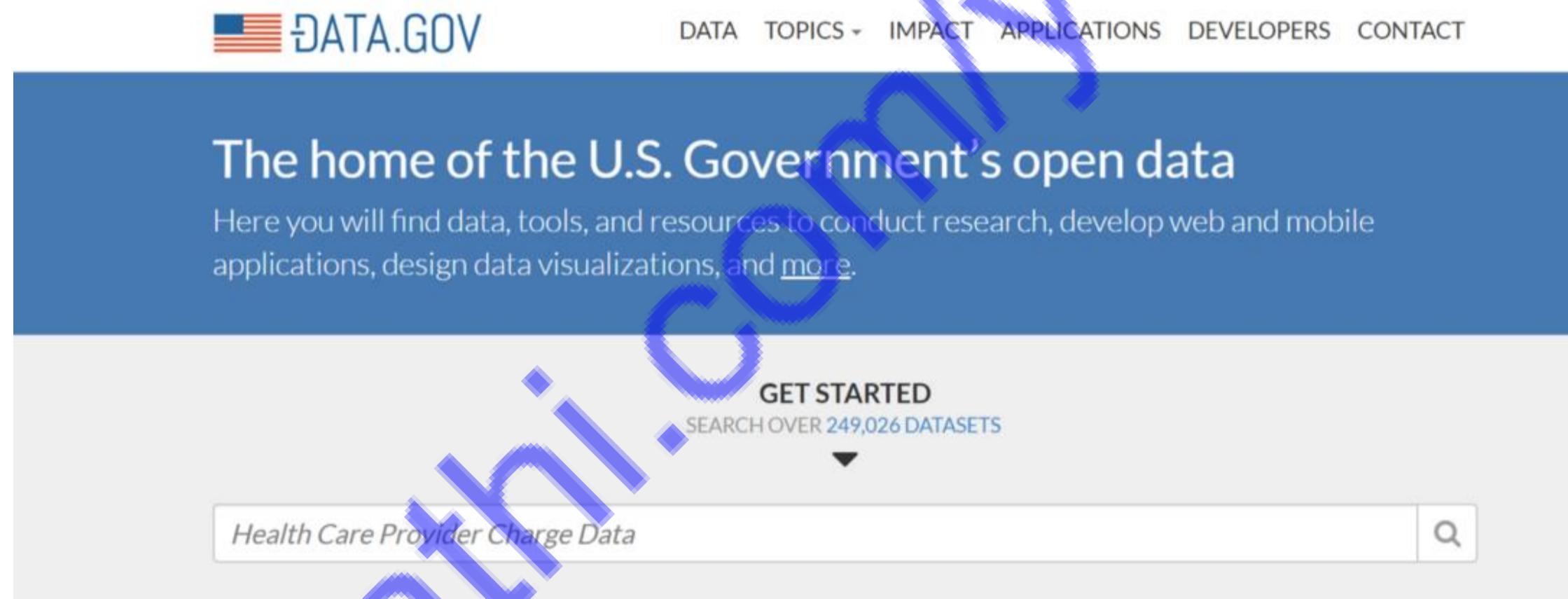
The screenshot shows the Kaggle Datasets homepage. At the top, there's a navigation bar with links for Search, Competitions, Datasets, Kernels, Discussion, Courses, and more. Below the navigation is a blue header bar with the word "Datasets". On the right side of this bar are buttons for "Documentation" and "New Dataset". The main content area has a search bar at the top with the placeholder "Search 18,710 datasets". Below the search bar, there's a "Public" section containing three dataset cards. The first card is for "Credit Card Fraud Detection" by the Machine Learning Group - ULB, uploaded a year ago, with 66 MB, 8.5 rows, and 1 File (CSV), and 3284 votes. The second card is for "Heart Disease UCI" by ronit, uploaded a year ago, with 3 KB, 8.5 rows, and 1 File (CSV), and 2141 votes. The third card is for "European Soccer Database" by Hugo Mathien, uploaded 3 years ago, with 34 MB, 8.5 rows, and 1 File (SQLITE), and 1863 votes. To the right of the cards, there's a "Sort by: Most V..." dropdown and a "Filter" button.

Kaggle is where many data scientists spend their nights and weekends. It's a crowd-sourced platform to attract, nurture, train and challenge data scientists from all around the world to solve data science, machine learning and predictive analytics problems. It has over half a million active members from 190+ countries and it receives close to 150K submissions per month. Started from Melbourne, Australia Kaggle moved to Silicon Valley in 2011, ultimately been acquired by the Google in March of 2017. Kaggle is the number one stop for data science enthusiasts all around the world who compete for prizes and boost their Kaggle rankings. There are only a handful of Kaggle Grandmasters in the world to this date.

Do you know that most data scientists are only theorists and rarely get a chance to practice before being employed in the real-world? Kaggle solves this problem by giving data science enthusiasts a platform to interact and compete in solving real-life problems. The experience you get on Kaggle is invaluable in preparing you to understand what goes into finding feasible solutions for big data.

U05.1.2: Kaggle Datasets: <https://www.kaggle.com/datasets>

## Data.Gov



This is the home of the U.S. Government's open data. Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more. Data.gov is managed and hosted by the U.S. General Services Administration, Technology Transformation Service. Data.gov is powered by two open source applications: CKAN and WordPress, and it is developed publicly on GitHub. Learn how you can contribute to Data.gov and these larger open source projects here.

U05.1.3: Data.Gov: <https://www.data.gov/>

# Amazon Data-sets

The screenshot shows the Registry of Open Data on AWS website. At the top, there's a dark header with the AWS logo. Below it, the main content area has a dark sidebar on the left containing sections like 'About', 'Search datasets', and 'Add to this registry'. The main content area on the right is titled 'Sentinel-2' and includes a summary, a 'Usage examples' section with a list of links, and a 'Landsat 8' section. A large blue diagonal watermark reading 'COMING SOON' is overlaid across the entire page.

**Registry of Open Data on AWS**

**About**  
This registry exists to help people discover and share datasets that are available via AWS resources. [Learn more about sharing data on AWS.](#)  
[See all usage examples for datasets listed in this registry.](#)

**Search datasets (currently 110 matching datasets)**

**Add to this registry**  
If you want to add a dataset or example of how to use a dataset to this registry, please follow the instructions on the [Registry of Open Data on AWS GitHub repository](#).  
Unless specifically stated in the applicable dataset documentation, datasets available through the Registry of Open Data on AWS are not provided and maintained by AWS. Datasets are provided and maintained by a variety of third parties under a variety of licenses. Please check dataset licenses and related documentation to determine if a dataset may be used for your application.

**Sentinel-2**  
[earth observation](#) [satellite imagery](#) [geospatial](#) [natural resource](#) [sustainability](#) [disaster response](#)  
The [Sentinel-2 mission](#) is a land monitoring constellation of two satellites that provide high resolution optical imagery and provide continuity for the current SPOT and Landsat missions. The mission provides a global coverage of the Earth's land surface every 5 days, making the data of great use in on-going studies. L1C data are available from June 2015 globally. L2A data are available from April 2017 over wider Europe region and globally since December 2018.  
[Details →](#)  
**Usage examples**

- [Satellite Search by Remote Pixel by Remote Pixel](#)
- [Sentinel-2 Cloudless Atlas by EOX](#)
- [Sterling Geo Using Sentinel-2 on Amazon Web Services to Create NDVI by Sterling Geo](#)
- [Using Vector tiles and AWS Lambda, we can build a really simple API to get Landsat and Sentinel images by Remote Pixel](#)
- [Spectator - tracking Sentinel 2, accessing the data and quick preview by Spectator](#)

[See 16 usage examples](#)

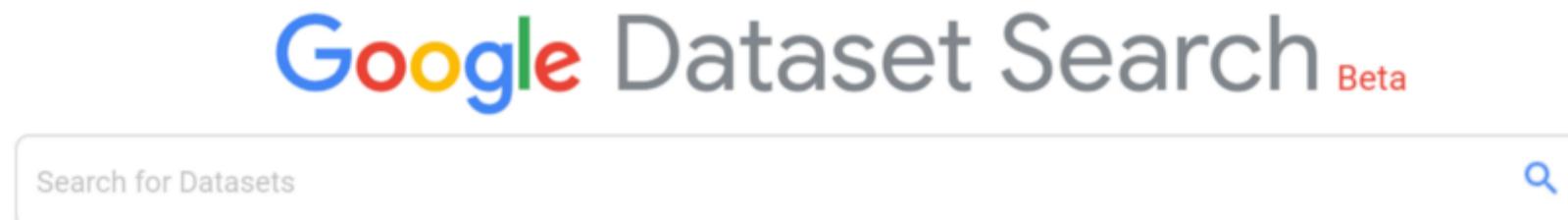
**Landsat 8**

This source contains many datasets in different fields such as Public Transport, Ecological Resources, Satellite Images, etc. It also has a search box to help you find the dataset you are looking for and it also has dataset description and Usage examples for all datasets which are very informative and easy to use!

The datasets are stored in Amazon Web Services (AWS) resources such as Amazon S3 — A highly scalable object storage service in the Cloud. If you are using AWS for machine learning experimentation and development, that will be handy as the transfer of the datasets will be very quick because it is local to the AWS network.

U05.1.4: Amazon Datasets: <https://registry.opendata.aws/>

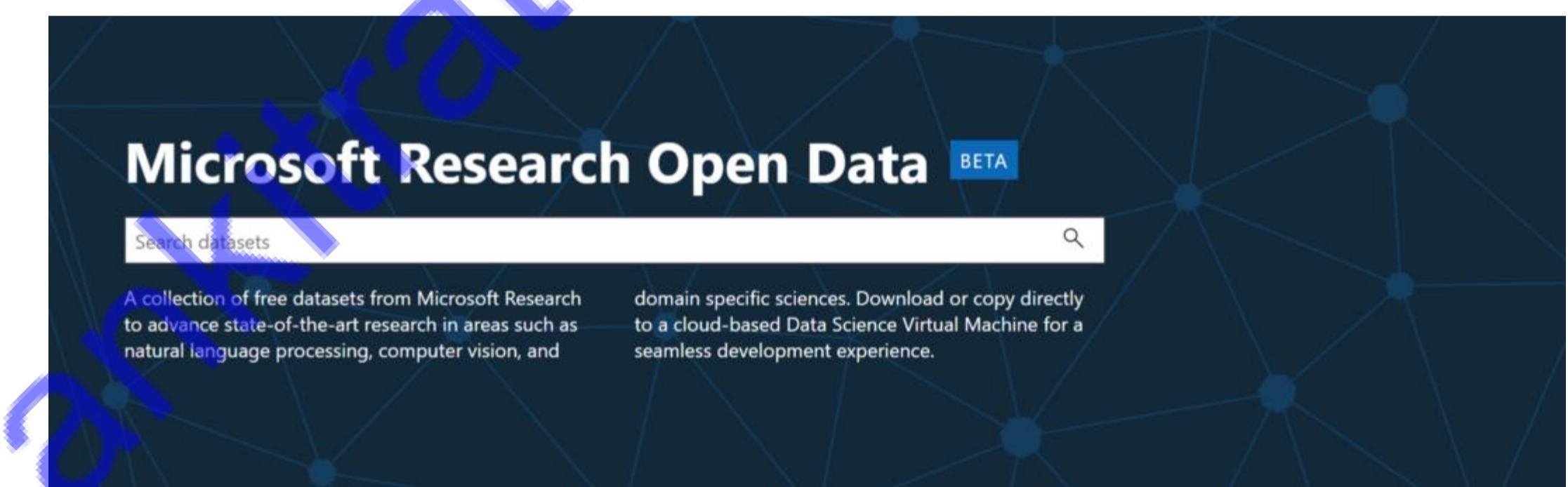
## Google's Datasets Search Engine



In late 2018, Google did what they do best and launched another great service. It is a toolbox that can search for datasets by name. Their aim is to unify tens of thousands of different repositories for datasets and make that data discoverable.

U05.1.5: Google Dataset Search: <https://toolbox.google.com/datasetsearch>

## Microsoft Data-sets

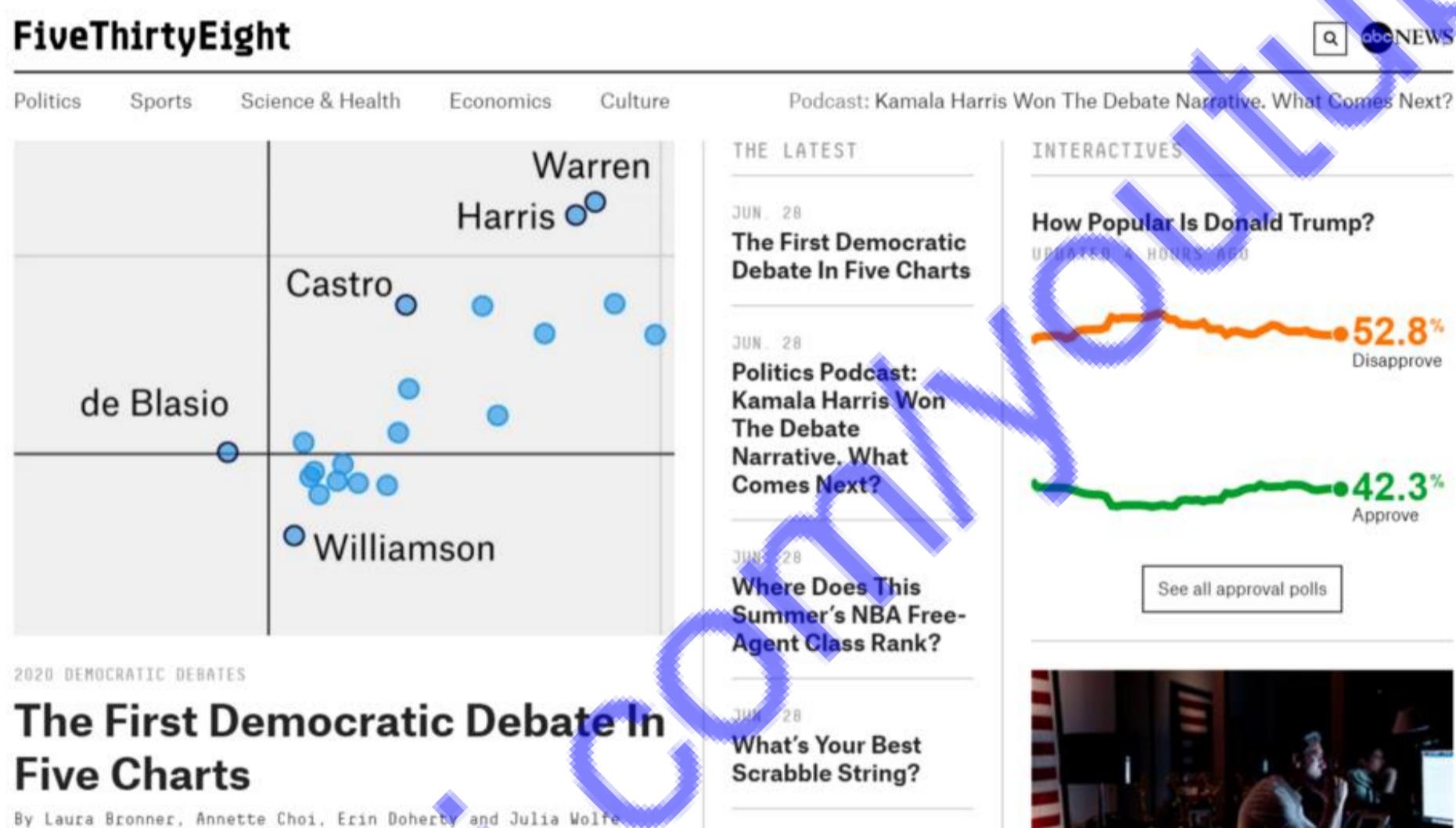


In July 2018, Microsoft along with the external research community announced the launch of “Microsoft Research Open Data”. It contains a data repository in the cloud dedicated to facilitating collaboration across

the global research community. It offers a bunch of curated datasets that were used in published research studies.

U05.1.6: Microsoft Datasets: <https://msopendata.com/>

## FiveThirtyEight

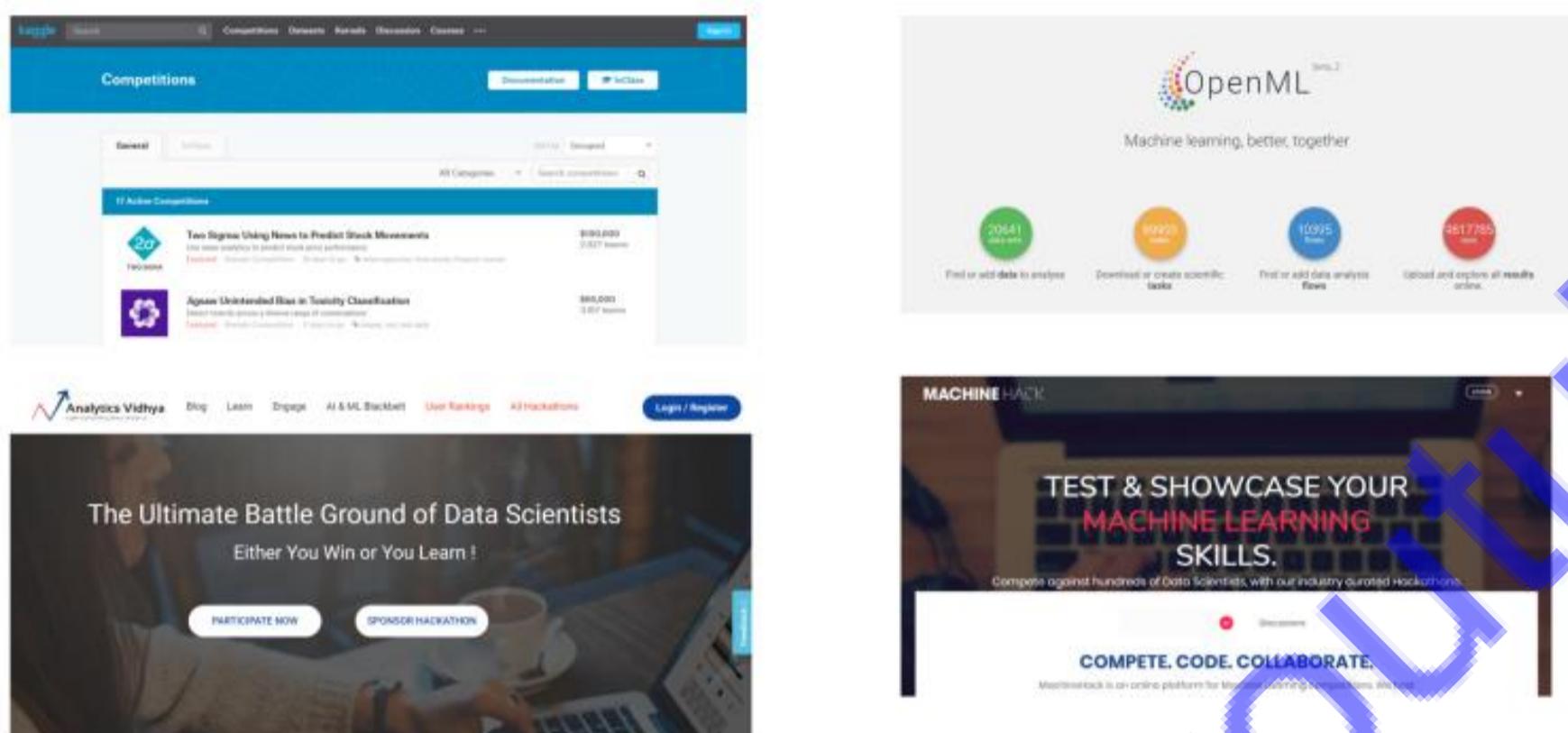


FiveThirtyEight, sometimes rendered as 538, is a website that focuses on opinion poll analysis, politics, economics and sports blogging. The website, which takes its name from the number of electors in the United States electoral college, was founded on March 7, 2008, as a polling aggregation website with a blog created by analyst Nate Silver.

You can find the data and code behind some of the popular articles and graphics here. You can use it to check others' work and to create stories and visualizations of your own.

U05.1.7: FiveThirtyEight: <https://fivethirtyeight.com/>

## 5.2 Participate in Hackathons

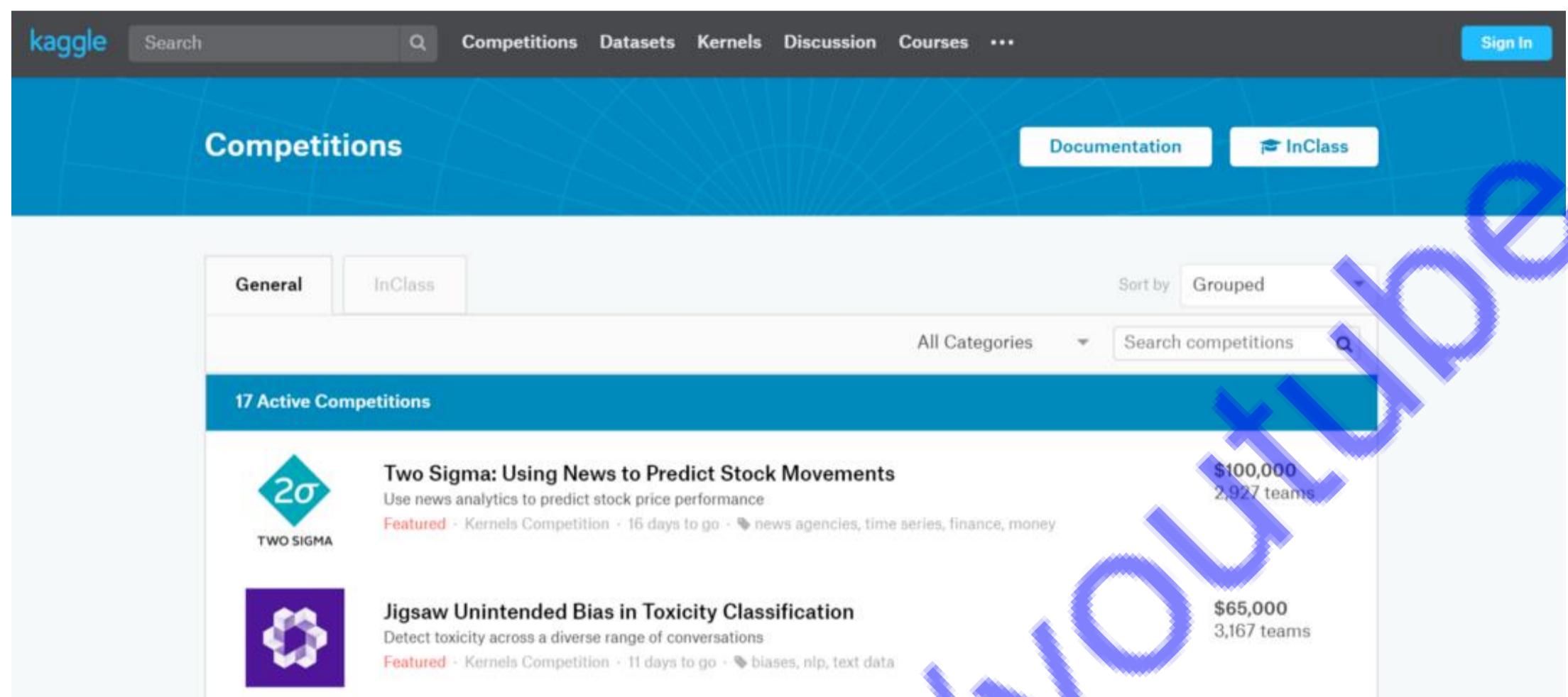


### Participate in Hackathons

Building your Portfolio

Participating in DS/AI competitions is one of the most frequent paths taken by data scientists, while it doesn't dish you all the challenges, it can help you to build your exploratory, modelling & cross-validation skills. You can also learn from fellow competitors about their approaches once the competition is over.

# Kaggle Competitions

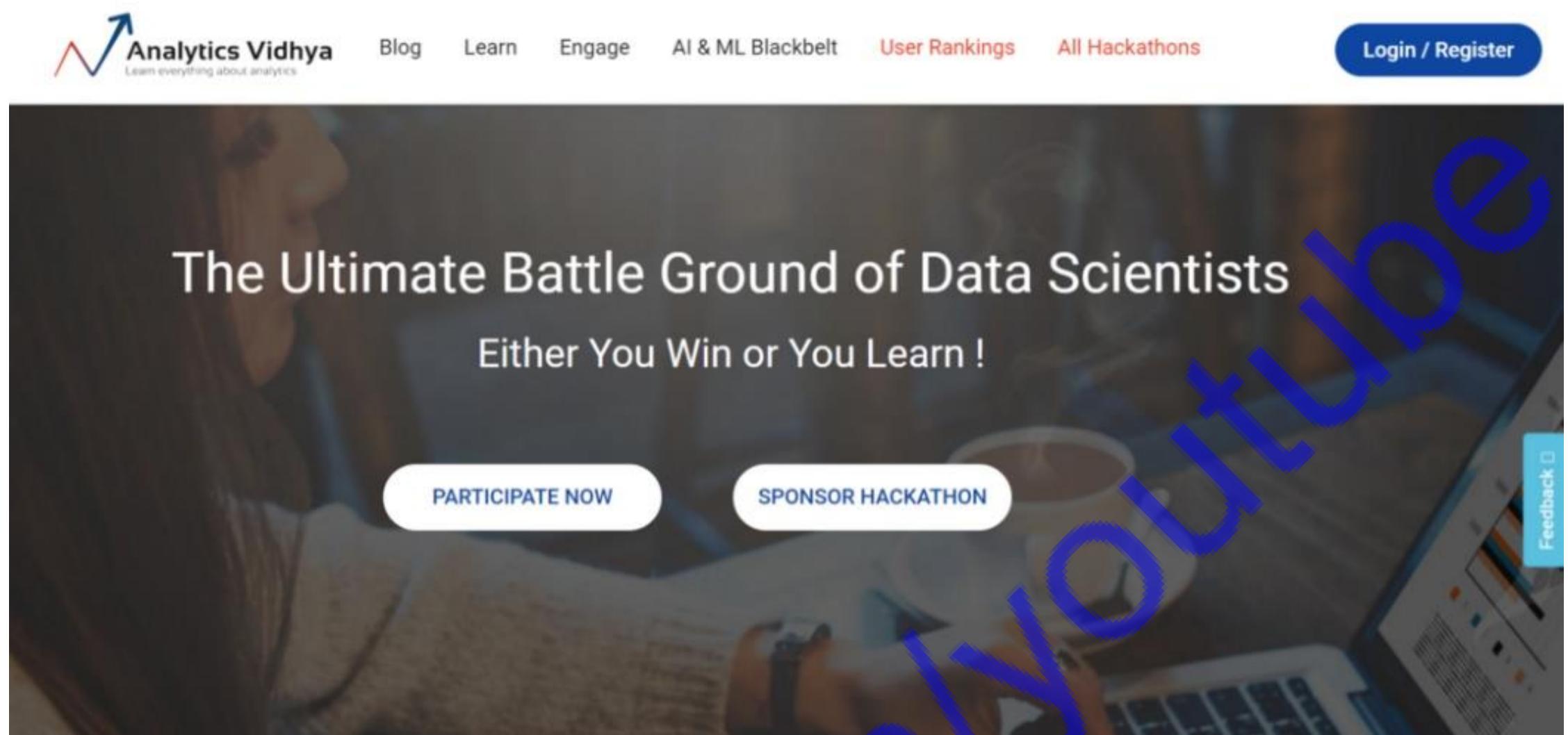


Kaggle runs a variety of different kinds of competitions, each featuring problems from different domains and have different difficulties. Before you start, navigate to the Competitions listing. It lists all of the currently active competitions.

If you click on a specific Competition in the listing, you will go to the Competition's homepage.

U05.2.1: Kaggle Competitions: <https://www.kaggle.com/competitions>

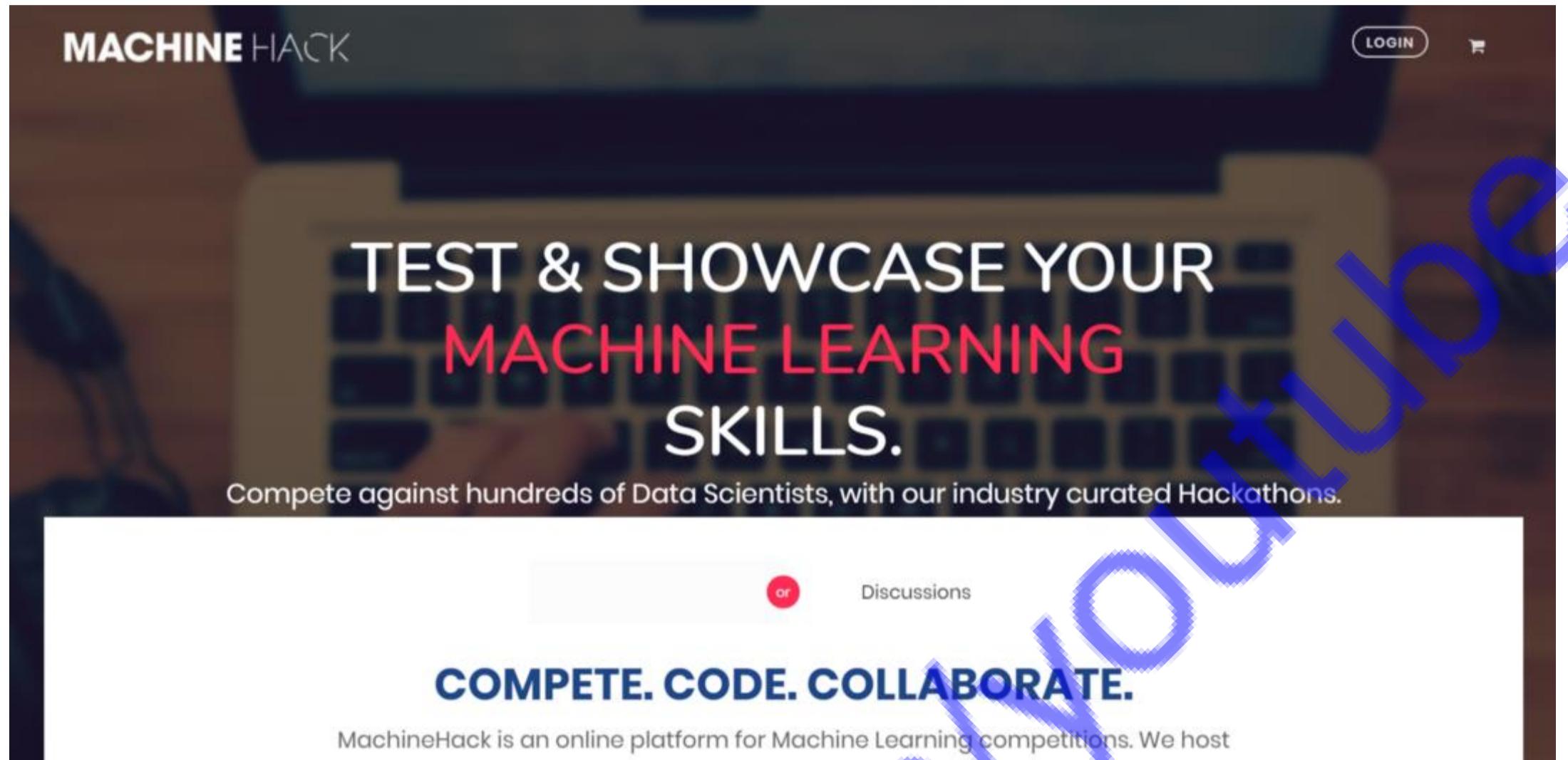
## DataHack by AnalyticsVidhya



AnalyticsVidhya Data Hack is also a platform where you can compete with the best in the world on real-life data science problems. You can learn by working on real-world problems. You can also upskill yourself and get hired in the listed companies. You can showcase your expertise and get hired in top firms. If you happen to be at the top of competitions, you can also win lucrative prizes.

U05.2.2: Data Hack: <https://datahack.analyticsvidhya.com/>

## Machine Hack

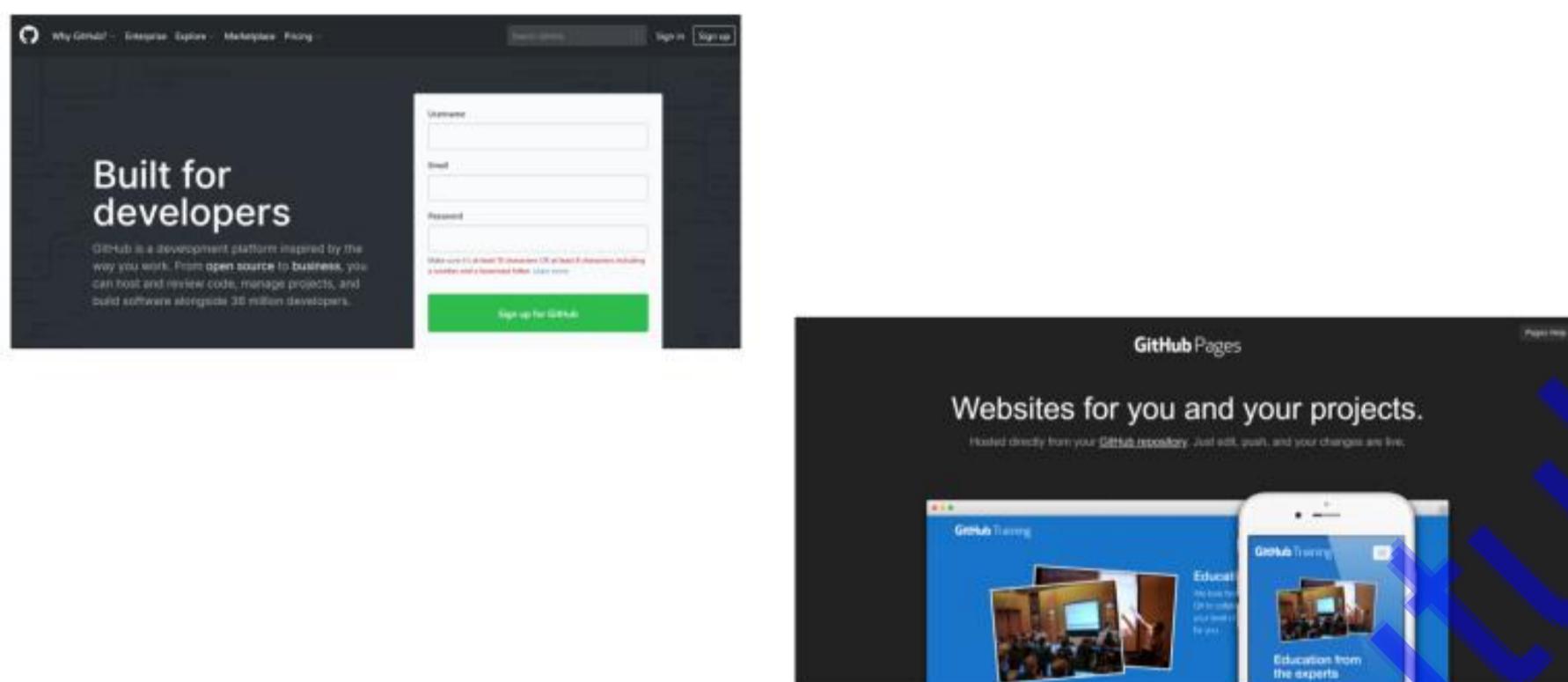


MachineHack is an online platform for Machine Learning competitions. They host the toughest business problems that can now find solutions using Machine Learning & Data Science techniques. Companies can hire better data scientists, the can discover & evaluate talented data scientists.

Just like Kaggle & DataHack, you can enrol in competitions here and help host solve their business problem. In return, you get near real-world project experience, you can learn from fellow competitors once the competition is over.

U05.2.3: Machine Hack: <https://www.machinehack.com/>

## 5.3 Publish on Git-hub

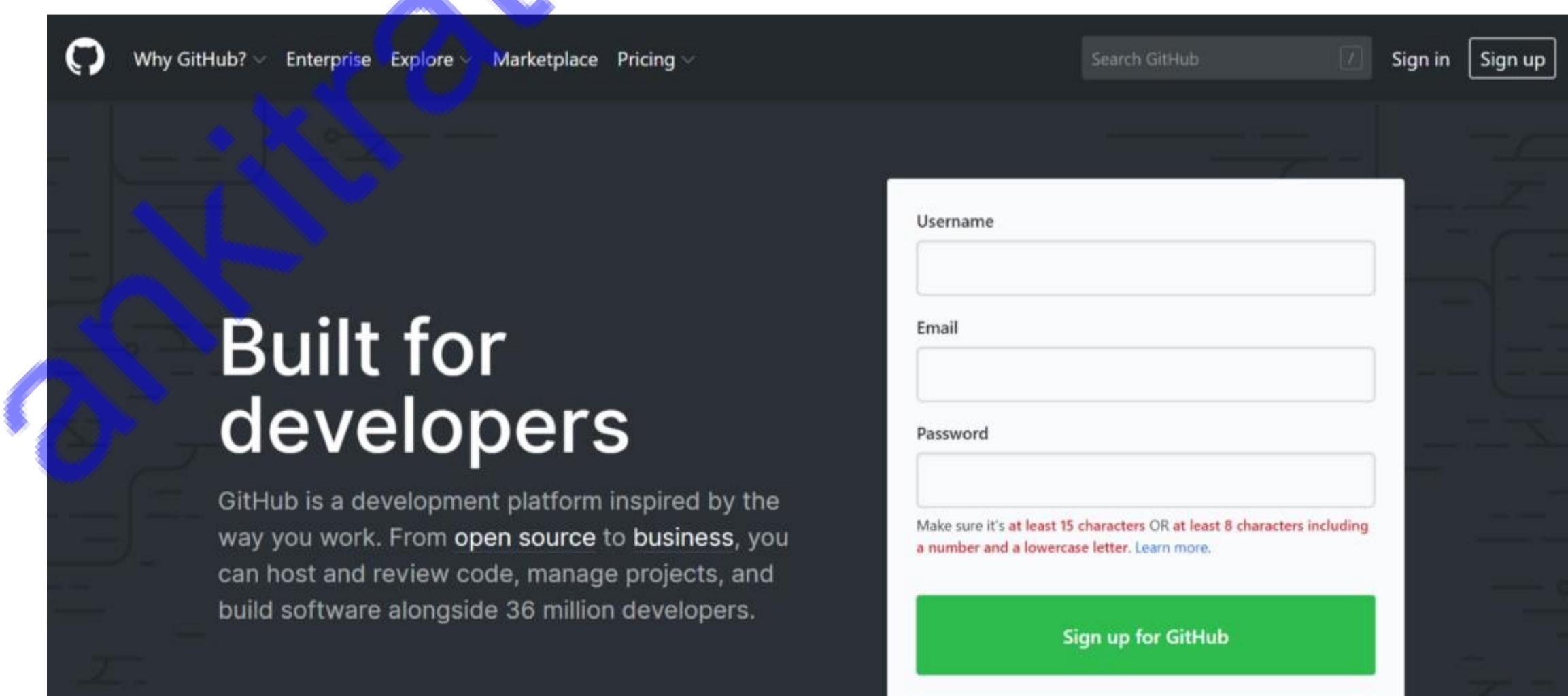


### Publish on Github

Building your Portfolio

GitHub is a powerful platform for software development, but at its heart, it's about empowering people like you by helping you learn from other developers, build the software that matters to you, and propel yourself to the next stage of your life as a software developer.

### Understand GitHub

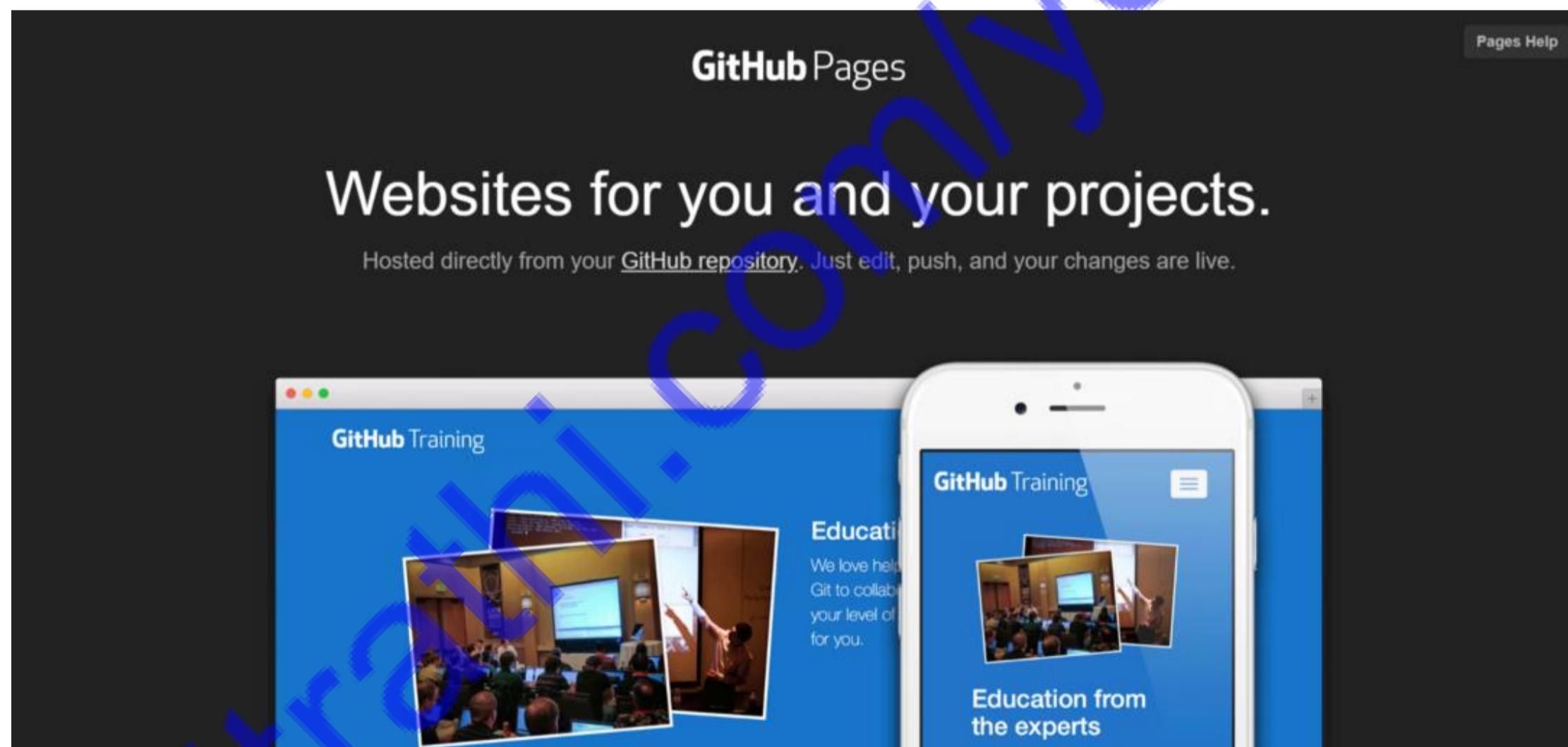


GitHub is a code hosting platform for version control and collaboration. It lets you and others work together on projects from anywhere.

In order to work on GitHub, you need to learn essentials like repositories, branches, commits, and Pull Requests. You'll create your own Hello World repository and learn GitHub's Pull Request workflow, a popular way to create and review code.

U05.3.1: GitHub: <https://github.com/>

## GitHub Pages



GitHub Pages are public webpages hosted and easily published through GitHub. The quickest way to get up and running is by using the Jekyll Theme Chooser to load a pre-made theme. You can then modify your GitHub Pages' content and style remotely via the web or locally on your computer.

U05.3.2: GitHub Pages: <https://pages.github.com/>

## 5.4 Write a Blog



Writing blogs is an effective way to showcase your expertise and skills. You can write what you have learnt recently, any interesting problem you have solved or worked on any project.

Writing an engaging blog-post is an art in itself, here are few tips to write and promote your blog-posts.

## Take notes for ideas

Start by writing down ideas as they occur to you. Make it a habit and keep doing it consistently by installing a note-taking app (like Keep, EverNote etc) on your mobile device.

Ideas occur to us all the time. You need a way to capture them when they do so that you can turn them into a great blog-post in the future.

## Build a simple outline

It is an essential step to develop an easy-to-follow outline before you sit down to write a blog-post.

Once you've picked a topic to write about, from the list of ideas that you've written down, create an outline. The outline contains a heading, introduction, major points you want to write about and conclusion.

To get the juices flowing, you should actually write the introduction and the conclusion first, then add a list of things that you'll cover in the body.

## Start with a story

Entertainment is the biggest factor in engaging your audience. If you're just about to start a blog, keep this at in your mind.

Stories engage people in and help clear the doubts. You are able to develop a scene which people can relate to.

Become a memorable writer by integrating stories into your blog-posts. It doesn't have to be your own story, you can tell interesting stories about others.

## Solve common problems

Consistent writing is one of the easiest ways to become a better writer. The question is, what should you write about? As a beginner, write blog-posts that answer questions.

Look for the problems that are common in your field, what most of the people are struggling with. Research about that topic, try to explore the problem and its possible solution.

## Learn & Share

When I write a blog-post, I read a lot about the subject. On the web and in real life, there are too many questions with too few answers.

Many a time, you will end up learning yourself in an attempt to write the post on a certain topic.

## Read other great writers

The truth is that if you don't read great writers, you don't really know how to do it and that successful blog that you dream of will evade you.

I've learned that I get a better education from studying authors' best work than I do from waiting for a piece of advice from them.

## Mentor Others

As part of being a successful and well-rounded data scientist, giving back can be a rewarding and beneficial aspect. Becoming a mentor, or mentoring those who want to follow in your steps of being a data scientist can sharpen your expertise and credentials.

## Build a Personal Brand

Building a brand is about giving yourself more opportunities to help and connect with people in your industry. And one of the best ways to build a brand is through blogging.

*A blog is a hub for your advice. It also has the added benefit of helping you rank on search engines.*

I hope that reading this inspires at least a few of you who want to become a data scientist and want to get better day by day by following the above-mentioned approach.

## Chapter 6

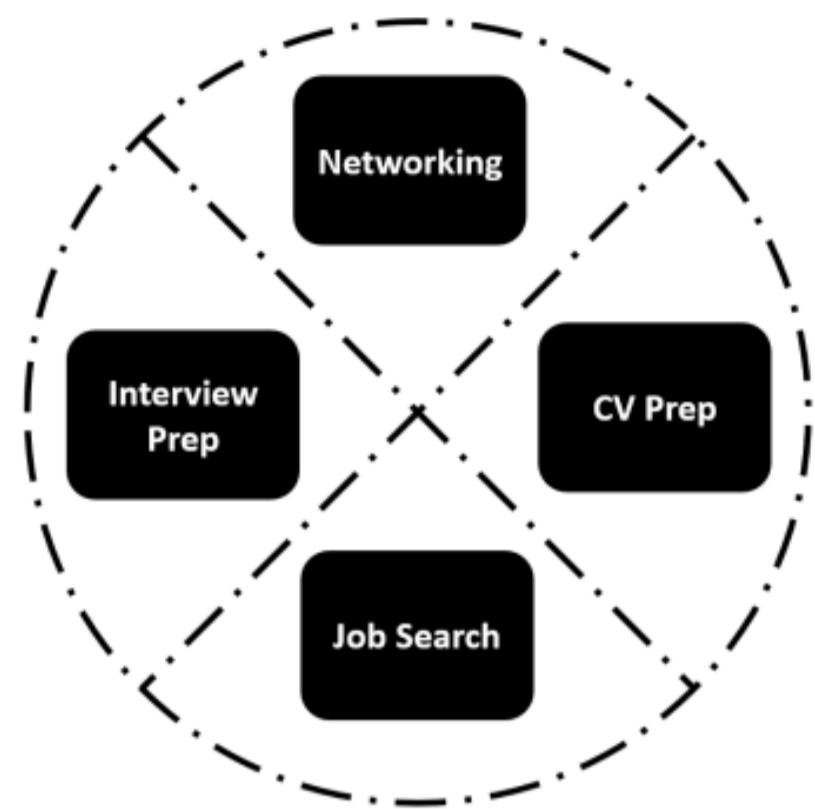
---

### Networking & Landing the Job



So you have done all your homework, you traverse through the DS/AI landscape, you built required concepts and learnt relevant tools and you have started working on your portfolio as well.

Now its time to network with people, work on your CV/resume, start looking for a job and prepare for the interview.



## Networking & Landing the Job

DS/AI: Self-Starter Kit

### 6.1 Network on LinkedIn

When it comes to networking on LinkedIn, there are two primary functions — just like in real-world networking: Building your network and nurturing relationships. I'll cover both in this chapter.

## Enrich Profile

Before using LinkedIn for networking, you need to update your profile and complete all sections to make it authentic, relevant, and compelling.

I would suggest you to avoid reaching out to new contacts and accepting connection requests until your profile was in tip-top shape. You need to assume that people will check out your profile when you are connecting. And with many connections, that's your first impression.

*Deny it all you want but first impression matters a lot.*

Make sure your photo, headline, and summary all tell a compelling story of who you are and what you have to offer. Show that you're worth connecting with! You can also request a recommendation from someone you've worked with so new connections have an impression of how you have helped others.

## Interact

Being on LinkedIn for namesake is not enough, you need to interact with people to expand the network and get the visibility. Provide status updates on a regular basis. It keeps you visible to the people in your brand community.

Like and comment on LinkedIn posts that you think are valuable, and share the posts with your connections and other groups you belong to. Share content you find at other sites — like Fast Company or Forbes or Huffington Post — that you think would be valuable.

*When sharing, remember to add content saying why you think it is valuable and expressing your point of view.*

## **Publish**

If you don't have a blog, you can publish your articles on LinkedIn. As for the blog, you need to keep your article engaging, either it should entertain or educate or inspire. If your article does all of the three, nothing is better than that.

Write articles to solve problems. Divide your article into logical sections, like heading, introduction, main points and conclusion.

*Keep images, embed content if required, to tell the story in your words.*

More post likes should also get you LinkedIn shares, post views, and comments according to correlation data. You can encourage people to like your post with a call to action.

## Reach Out

When you're reaching out, remember to customize the request instead of using the default "I'd like to add you to my network." You can even customize requests when using the LinkedIn mobile app.

## Accept Requests

Know your criteria for accepting requests (and remember what you lose if you are a closed networker). LinkedIn alerts you when you have requested. Get in the habit of accepting them soon after receiving them.

## Guide/Help Others

LinkedIn does the heavy lifting when it comes to staying on top of people in your network. They provide notifications when someone you know has a birthday, work anniversary, or new job.

You have the option of "liking" the notification or sending a message. I suggest choosing "send a message" and writing a personal note. Determine a time of day you will check-in and get in the habit of doing it daily so you don't miss any of your connections' important dates.

Now you have the keys for unlocking the power of networking on LinkedIn.

*In all of these interactions, remember that successful personal branding is the result of being authentic and being consistent.*

## 6.2 Prepare the CV

**Ankit Rathi**  
Lead Data Architect

Delhi/NCR, India  
+91 9891650969  
rathi.ankit@gmail.com  
[linkedin.com/in/ankitrathi](https://www.linkedin.com/in/ankitrathi)  
[medium.com/@rathi.ankit](https://medium.com/@rathi.ankit)

**Summary**

- Seasoned Data Science Architect, Kaggle expert & Data blogger/author with 13 years in data engineering/architecture and 5 years in data science/machine learning.
- Designed & developed many data intensive technology solutions using various tools in data architecture, data science, big data & cloud.
- Translated complex business problems into technology & analytics solutions for data-driven decision making.
- Demonstrated knowledge of the business success drivers, industry trends, regulatory issues and competitive marketplace.
- Continuously in sync with latest developments in the analytics field methodologies and technologies.
- Participating in Kaggle (Data Science) competitions since 2014, achieved Kaggle Expert level in 2017.
- Experienced in leading technical teams and handling technical & business stakeholders.

### Keep it Short

You don't have to list everything you've ever done. Try making a master resume which has all of your work history and then pulls what's most relevant for each job.

### Don't put Objective

Objective serves no purpose in your resume since it doesn't help you distinguish yourself while taking up spaces that can be better utilized to showcase more useful information about you such as projects, experiences, etc.

## Photo on CV

Based on in which region you are applying, you need to decide whether you can put your picture. While this is common in some countries in Europe and South America, it is not appropriate in the US and Asia.

## Always Proofread

Having typos or grammar errors in your resume can be the quickest way to have your application eliminated. Use spell-check and have a friend to check it over.

## Mention Accomplishments

Quantify your accomplishments where possible, use verbs to start the bullet points.

## Add Cover Letter

If there is a place to submit a cover letter, do so. Just like your resume, you can have a master cover letter that you pull paragraphs from. Tailor at least the first and closing paragraph to the company and make sure you get their name correct.

## Put Results & Links

Include the results and links of your projects. For instance, if you are doing a competition on Kaggle, it will be important to mention your ranking & you can put your solution on GitHub.

## Mention Portfolio

Mention your LinkedIn, GitHub, Kaggle profiles links on your resume. It can reveal more useful information about yourself and it is also good to showcase the relevant work you have done.

## Tailor for Job Description

It is crucial that you tailor your description of experience towards the job's requirement because it can make your skills and experience seem more relevant and better fit for the position.

## Highlight Coursework

You should list coursework that is relevant to the position you are applying to. It is a quick way to show hiring managers your background, which can have some influences in their decision on inviting you for interviews.

## Don't put Common Projects

Common projects and homework are highly guided problems that will help you to stand out. You should include projects that demonstrate your ability to solve real-world problems and your interests.

## Don't Rate your Skills

Numerical rating your skills are not meaningful since they are not standardized or calibrated.

## 6.3 Search for the Job



### Have a Growth Mindset

Growth Mindset is believing that their most basic abilities can be developed through dedication and hard work — brains and talent are just the starting point. Don't use "talent" to describe others as an excuse for your laziness. What you need is to learn the right way and practice many times until you are good at it.

### Take Notes

Take note of all the interview questions you have been asked, especially those questions you failed to answer. You can fail again, but do not fail with the same question. Most of the time, you learn more about the topic when you reflect on the notes taken.

## Browse widely

Jobs in DS/AI go by many names besides data scientist. These include product analyst, data analyst, research scientist, quantitative analyst, and machine learning engineer. Try searching for all of these terms to find positions and then use the description to evaluate the fit.

## Self-reflect

Rather than applying to every type of data science job you find, think about your strengths and where you want to specialize. There are data scientists who have strong statistics skill and the ability to work with messy data and communicate results. While other data scientists have very strong coding skills, maybe have a background in software engineering, and focus on putting machine learning models into production.

## Don't demand perfection

Your first job in the field probably would not be your dream one. You may need to start out by moving to a position where you can leverage your other skills. That doesn't mean you shouldn't have certain requirements and preferences, but it does mean you'll want to have some flexibility. The most important criteria for your first job may be that it has a supportive environment, with lots of other analysts, where you can learn a lot.

## Don't undersell yourself

Job descriptions are generally wish-lists with some flexibility. If you meet 80% of the requirements but are otherwise a good fit, you should still apply. With that said, be wary of job descriptions that describe a unicorn

where they need every skill on earth. It usually means they don't know what they're looking for and they expect a data scientist to come and solve all their problems without any support.

## Look on LinkedIn

Check if you know anyone at the company you're interested in. If you don't know anyone, see if there's anyone in your alumni networks. You can also check for second connections and see if the person who bridges you can introduce you. Many jobs get hundreds if not thousands of applications for each position and having someone refer you or give you feedback on what the team is looking for is enormously helpful.

## Check out Meetups/Conferences

Sometimes hiring managers will come to meetups or conferences to recruit. You may also meet someone in the company or sub-industry you're interested in. If you ask if their company has an opening or if they can refer you, though, you'll probably be directed to the company's career page. This is why it's important to build your network before you need it, means starting off with a strong ask is not a great way to build a mutually fulfilling relationship.

## 6.4 Crack the interview



### Prepare & Practice

Most of the questions (at least 60–70%) are based on your background and skills required for the job. Which means you can prepare a list of questions that can be asked to you and keep your responses comprehensive yet short.

You can also ask the recruiter about the number of rounds and what is expected at each stage so get as much information as you can and keep yourself prepared.

## The Resume is a Fair Game

Whatever you have mentioned in your resume is a fair game. So keep everything on your fingertips. If you have mentioned a skill or a course, be prepared to be grilled about that.

## Research about the Company

You may have done some research when writing your cover letter, but once you get the interview, dig a little deeper. In addition to the tips in the thread below, find out about your interviewers' professional accomplishments. I was very impressed when a candidate I was interviewing asked some technical questions about a presentation I had given.

## Have your Questions Ready

Each interviewer should leave time for questions at the end. If they don't, that's a bad sign. Interviewing is a two-way process: you're evaluating them as much as they're evaluating you. If you don't know what to ask, you can prepare them based on what is important to you like the quality of life, culture, and management practices. You can ask each interviewer different questions to maximize how many you can get answered, but you could also try to ask multiple people the same questions to see if and how their answers differ.

## Never Mention the Numbers

While it is a cultural thing, in some countries it is OK to mention your current salary while in other law prohibits the recruiter to even ask for it. Whatever is the case, never mention your expected salary.

If you name a number, you risk them giving you a lower offer than they would have otherwise. Their offer should not depend on your current salary or expectations; it should be your worth in the market and similar to the salary of your peers there.

## Handle Rejection Gracefully

You will almost inevitably get rejected in some job interviews, maybe dozens of them. DS/AI is a competitive field and this is a very normal process that everyone goes through. If they reject you or you don't hear back from them, you can express your disappointment politely and thank them for their consideration.

You can ask for feedback, but know that many hiring managers would not be able to give you any because they want to avoid the possibility of being sued. While it is okay to take a little time to wallow, just don't lash out in public or to the hiring manager. It won't help the situation, but it will hurt your professional reputation.



Maintain

**DS/AI: Self-Starter Kit**  
Build Your Own Roadmap

[ankitratthi.com/youtube](http://ankitratthi.com/youtube)

## Chapter 7

---

### Making Career Future-Proof



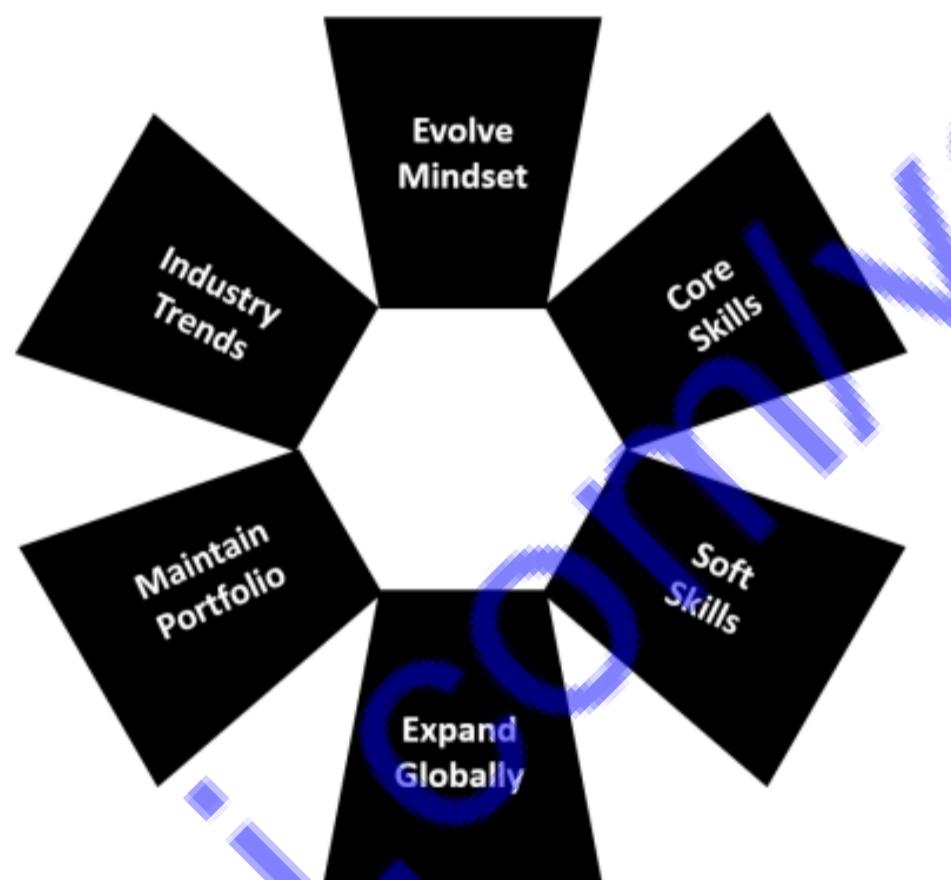
This chapter talks about how you can make your DS/AI career future proof. Let's first understand, why this is required? So what exactly does it mean to future proof your career?

*Future proofing your career is simply taking the extra steps to prepare yourself for constant*

*technology disruption, one that's going to rely heavily on adaptability.*

So rather than waiting for someone or technology to replace your labour, you'll take a proactive approach to put yourself in a position where potential employers can't afford to work with you.

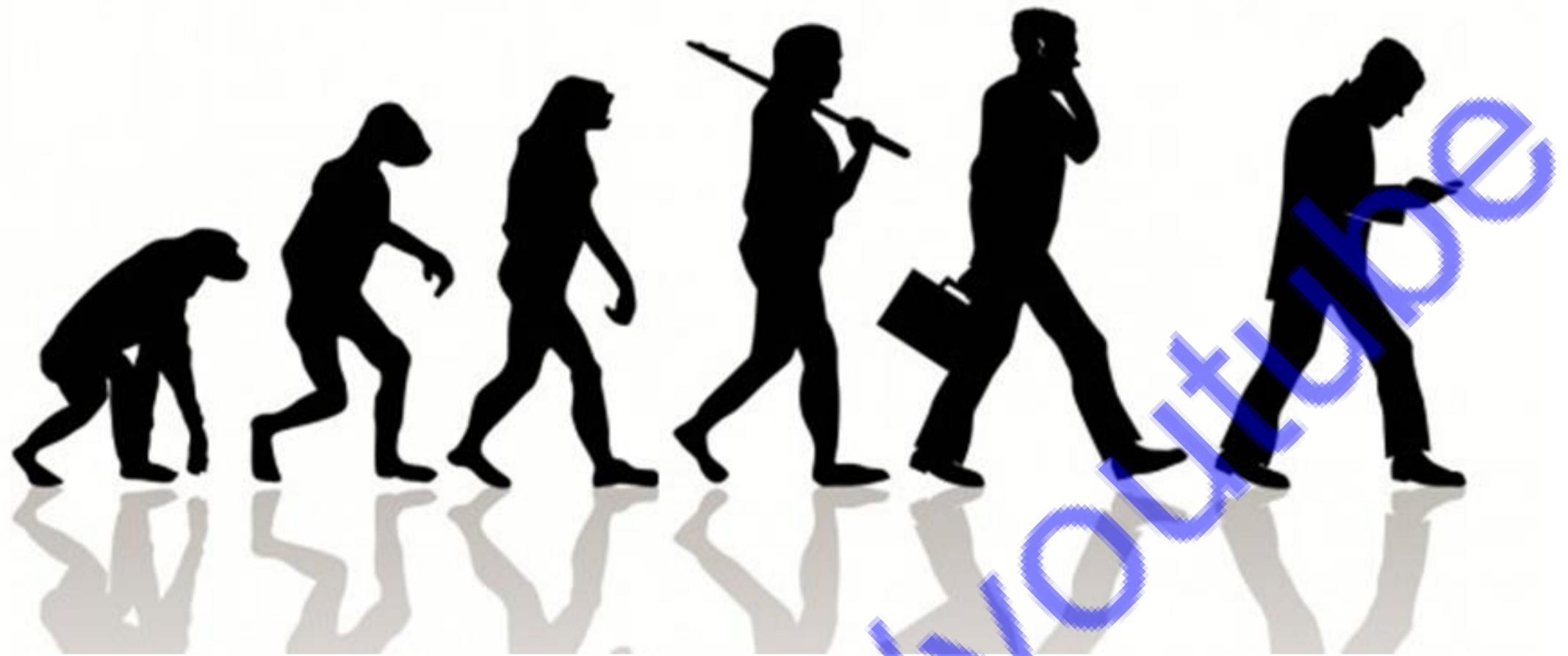
Follow these six steps and you will secure your place in the workforce alongside disruption in technology instead of getting edged out:



### Making Career Future-Proof

DS/AI: Self-Starter Kit

## 7.1 Build ‘Evolve’ Mindset



Technology is only going to keep evolving, and it's always going to get better. While you and I may not know exactly how we do know this change is inevitable. So as technology in the work environment evolves, so should the workforce.

*People who are more adaptable and resilient will be the ones who will make the cut. They will also be the employees who are not threatened by technology disruption.*

But how to become more resilient and adaptable?

First, you can prepare for the future like you are doing today and be ready to change course at a short notice. And two, when your environment begins changing, have an open mind about what this transition may bring — and be ready to take it head-on — instead of resisting it and sticking to your old habits.

One of the best ways to build the confidence necessary for this new technology-driven world is to level-up your digital skill-set.

## 7.2 Hone Core Skills



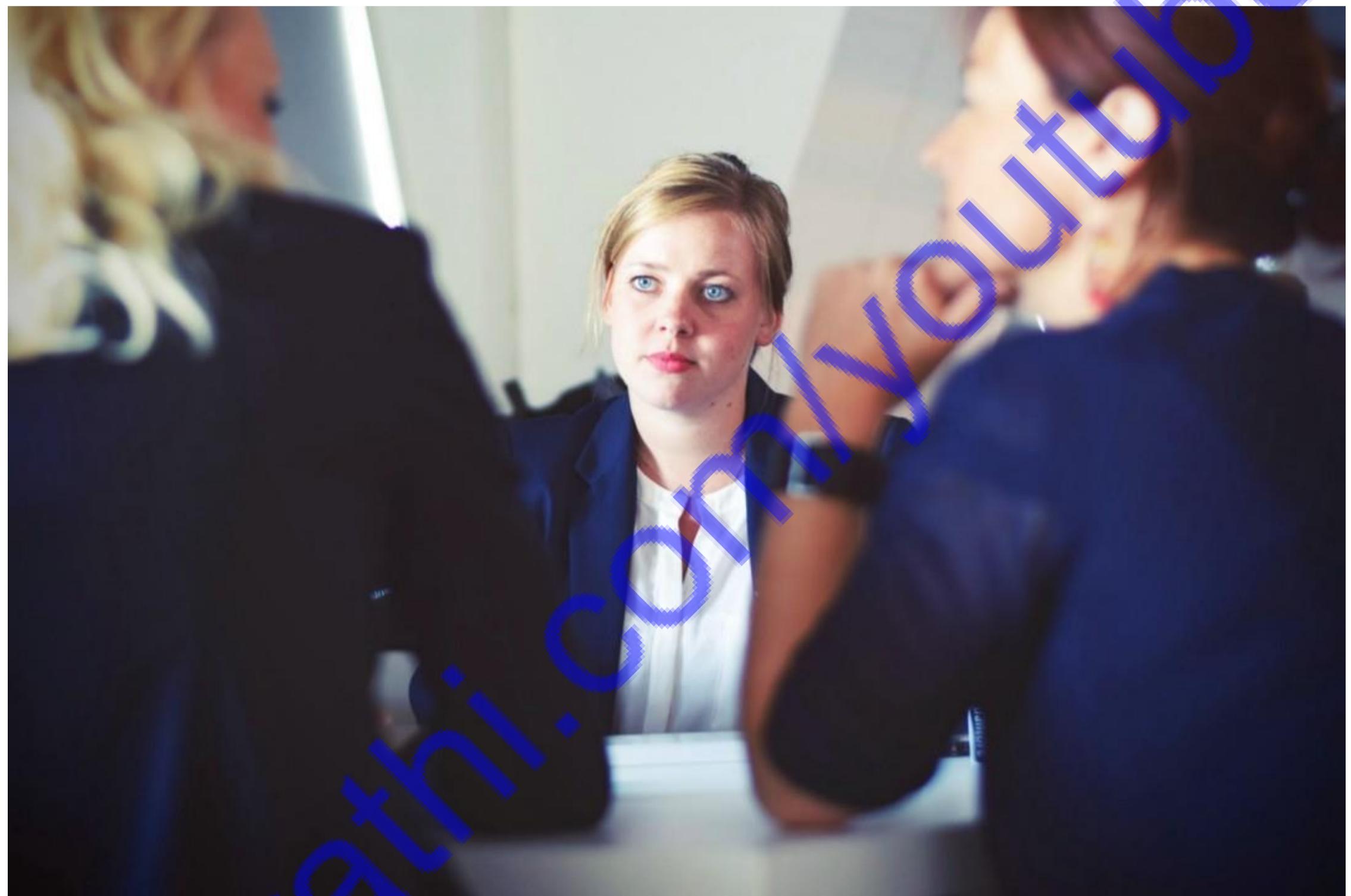
The technology disruption will also create a need for higher skill levels. You may have already witnessed the fact that having a college degree doesn't make you stand out anymore. In fact, it is going to get redundant in the near future.

*Organizations are emphasizing more on the core skills to do the job, the trick is to never stop learning and keep honing your core skills.*

You should continue to acquire new relevant skills as well—especially ones which will be in demand.

The best way professionals can do this is by enrolling in online courses or you can also learn on your own. Like degrees, don't focus on collecting training certificates but try to gain hands-on learning as much as you can.

### 7.3 Develop Soft Skills



You may be able to train a robot to automate the technical skills of your job, but soft skills such as leadership, communication, collaboration, and time management are still tasks only humans do well.

*Since technology is not at the point where robots have the same emotional intelligence as humans, these soft skills are and will continue to be in high-demand.*

The key is to consciously try to improve on your soft-skills. Fortunately, you can also hone your soft skills with online classes so you can be more proactive in this department too.

## 7.4 Maintain Digital Portfolio



In one of the lessons earlier, we learnt how to build our portfolio. Most people wait until they are ready to find a new job to update the portfolio. But this is a huge mistake since you may forget what you've been up to, or worse, forget to mention a major milestone or achievement.

*A better approach is to always update your projects and accomplishments as you work through them.*

This guarantees that you never forget to highlight something and you'll always have a list of your achievements on hand.

It is also a great idea to keep your performance reviews here, both the good ones and the bad. You can always refer back to these anytime you need a pick-me-up or if you want to narrow down your specific areas needing improvement.

## 7.5 Expand Network Globally



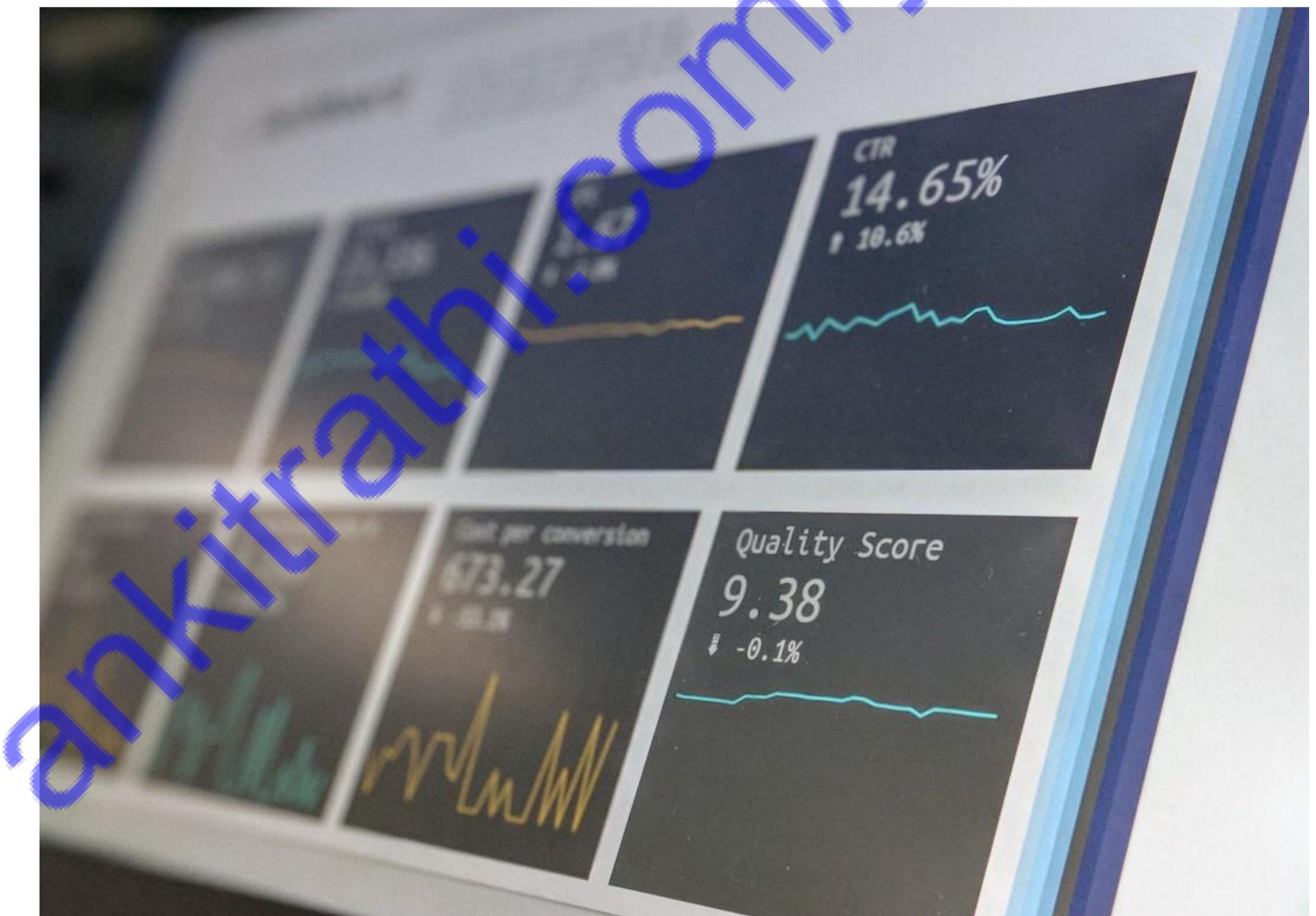
Couple the evolution of technology with the rise in telecommuting, remote work, and networking sites like LinkedIn, and you will quickly find connections outside of your local network.

To stay in touch with these global members of your team, you will want to become a pro at virtual project management tools like Trello and messaging platforms like Slack.

*Apart from your team members, become more familiar with the geographic regions that pertain to your job by reaching out to professionals in those countries too.*

If you get an opportunity to relocate or take on an international project, go for it and give you a leg up on your competition, it will go a long way to future-proofing your career too.

## 7.6 Monitor Industry Trends



Understanding the future of your industry is a giant factor in how well you can truly future proof your career. Essentially, there's no point specializing in a language/tools that may be completely redundant in a few years.

The tasks most at risk are those requiring low-level repetitive tasks each day. So if that description matches your current position, it is time to add more skills to the fore and prepare for the redundancy in future.

*But remember, an increased use of technology in the workplace doesn't always mean your job is at risk.*

Rather, it could mean you will just need to know how to use upcoming technology/framework as a way to potentially do your job better or more efficiently.

Stay abreast with industry news to see if it impacts only specific to your company or your entire industry as a whole. If you see an industry-wide trend, that's a good sign you should learn those new skills and consider switching career paths.

Either way, follow these steps and you'll ensure your career — both now and in the future — is set up for success, even as technology disruption and automation move in.

[ankitrathi.com/youtube](http://ankitrathi.com/youtube)

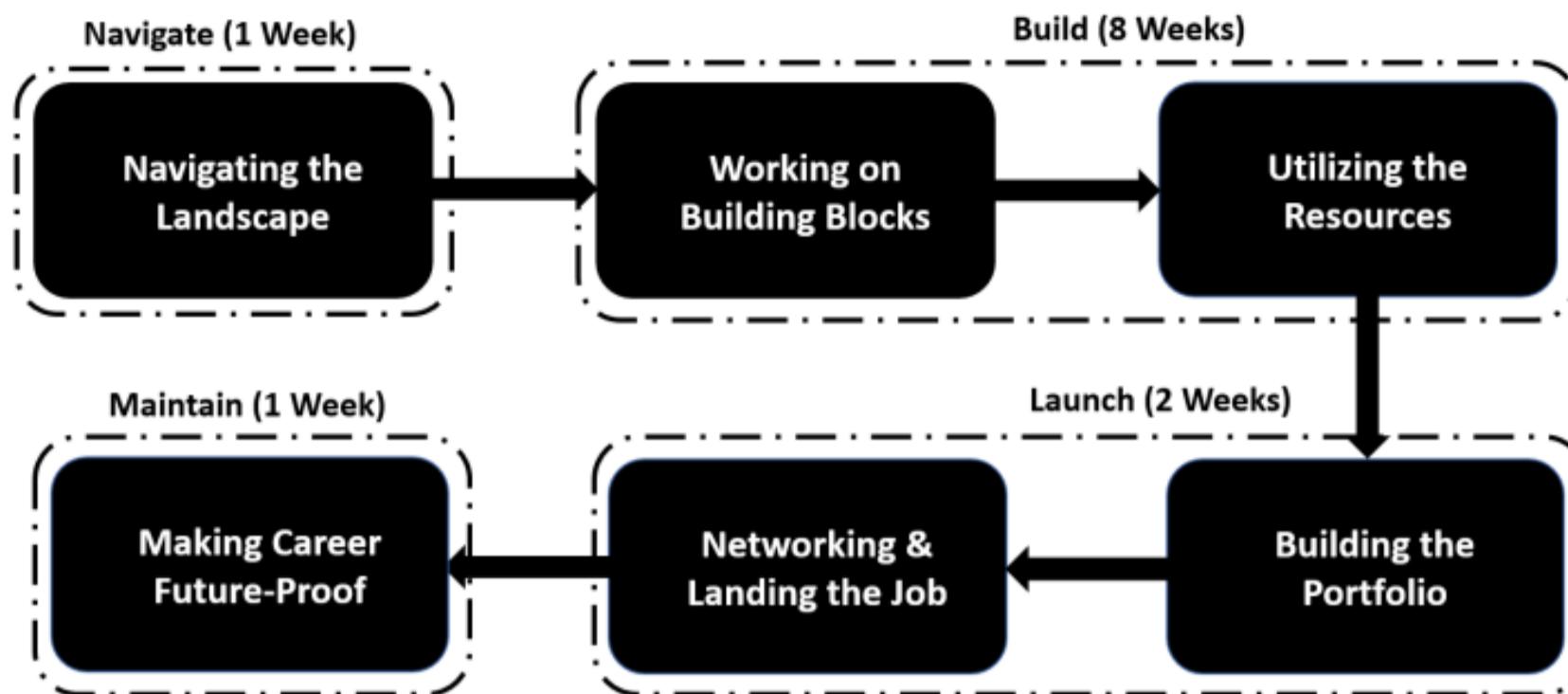
## Chapter 8

---

### Putting It All Together



Now that you have covered each and everything from the content perspective. Let's reflect on what you have learnt and after that, I will give you few tips to overcome some learning obstacles.



## Putting it All Together

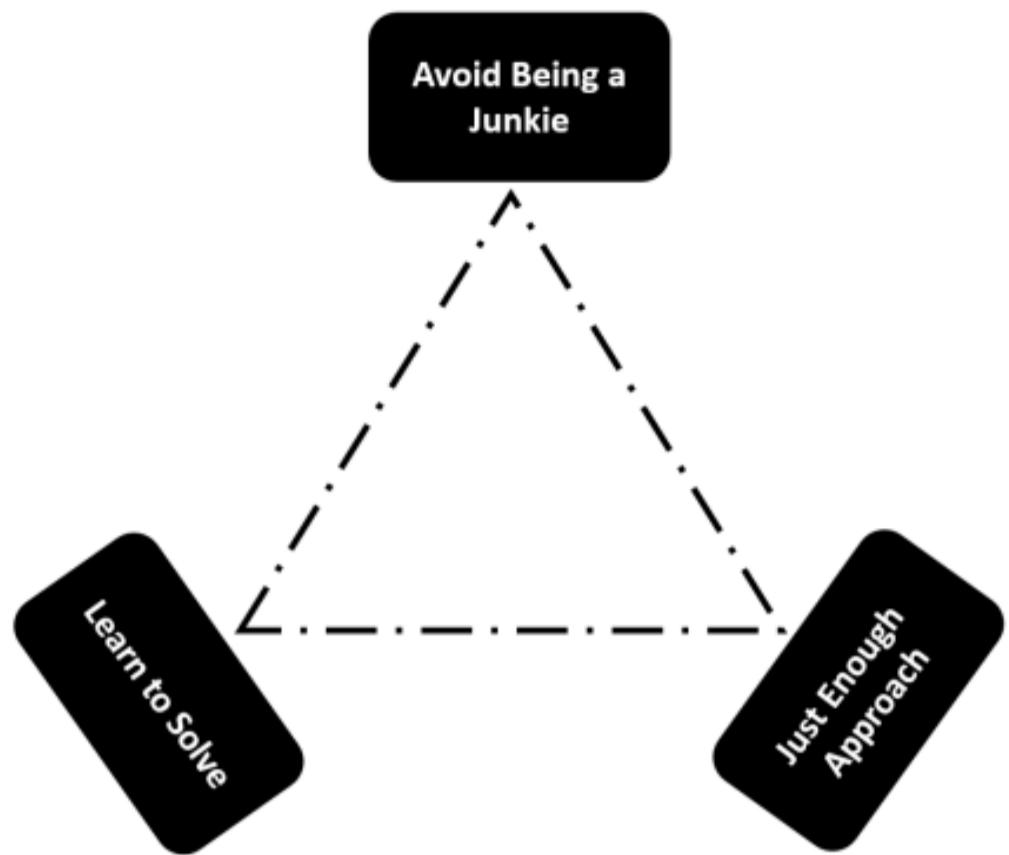
DS/AI: Self-Starter Kit

You can notice that all these activities can be completed in 12 weeks.

*As I mentioned earlier that you can not be an expert in DS/AI in such a short time but I can ensure that you have all the knowledge, concepts, processes and tools and techniques available to you to tackle any challenge in DS/AI.*

After going through the 'Navigate' step, you need to focus as much as you can on building your skills in 'Build' step, you can start some of the activities of 'Launch' step in parallel as it would help you to grasp and apply what you learnt in 'Build' step.

On your path to learn and build your skill-set in DS/AI, you will find many obstacles. Some of them I am covering here and what you need to do to overcome those obstacles.



## Putting it All Together

DS/AI: Self-Starter Kit

### 8.1 Avoid Being a Junkie



Online learning is like Chocolate. You can't take just one online course or subscribe to just one newsletter/blog. Once you start, you want to have it all!

But the trouble with Chocolate is that, even if you eat the whole damn packet, you don't have enough. The only thing that stops you is that the packet is empty, your stomach is full, and you feel kind of sick. But your appetite doesn't really go away.

*Similarly, having an unlimited access to all the online courses, MOOCs, webinars, workshops, challenges, e-books, blogs, TED talks, podcasts, and all the other free (or affordable) resources can get you in trouble.*

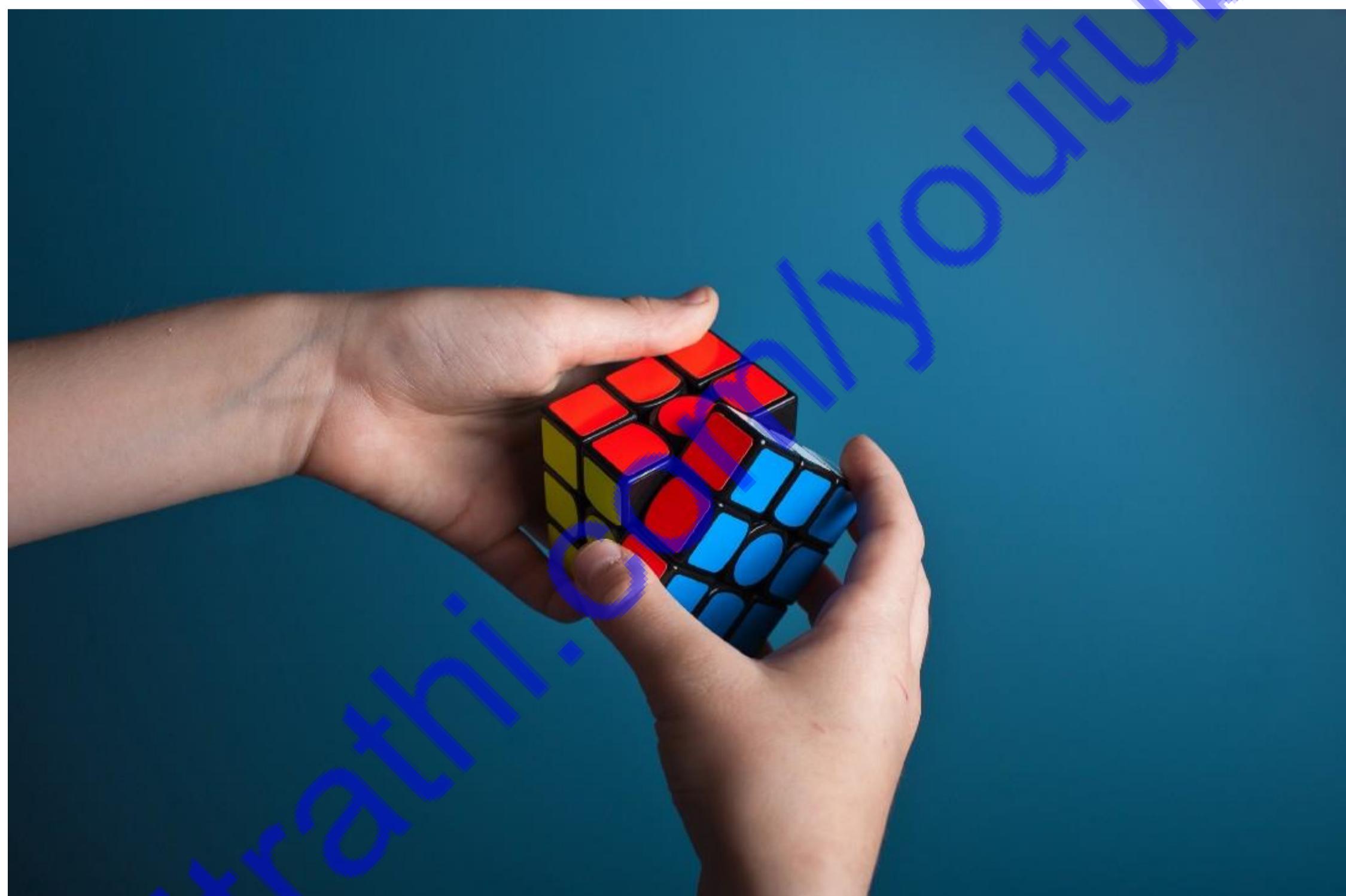
The key to avoiding the trouble is to prioritize, plan, schedule and log what you do. And analyse at regular interval how your progress is and what is working while what needs to be improvised.

When you subscribe to a webinar or an online course, put it in your calendar. Set up reminders; make a commitment to attend it as if it was compulsory. In case of live events, if you can't catch them live, schedule the time when you're going to watch the replay. When you watch a webinar or an online course, don't multitask. Close the door of your room, turn off your phone, and don't open other tabs to check Twitter. After the webinar/course video ends, take several minutes to debrief. Find something you can put into practice right away and put it on your to-do list.

At the end of the month, evaluate. What have you learned? What worked? What didn't?

Make a list of your favourite bloggers, podcasts, and other resources. Read them, listen to them or watch them as a part of your daily schedule. If you set aside time to do this, you will be able to concentrate on learning when you learn and to create when you create.

## 8.2 Learn to Solve



*Learning is 20% information and 80% action.*

If you aren't taking action, you aren't learning; you're just wasting time. The key to making learning more effective is Active Learning.

According to a study into learning-centred approaches to education, students learn more when they participate in the process of learning. Active learning is discussion, practice, review, or application. Problem-solving, exploring new concepts in groups. Working out the problem on a piece of paper.

*Active learning is any learning activity in which you participate or interact with the learning process, as opposed to passively taking in the information.*

When given the opportunity to actively engage with the information you are learning, you perform better. It nurtures the brain, giving it an extended opportunity to connect new and old information, correct previous misconceptions, and reconsider existing thoughts or opinions.

Active learning encourages your brain to activate cognitive and sensory networks, which helps process and store new information. One more similar research at Cornell University found that learner attention starts to wane every 10–20 minutes during lectures — which means instructors are continuously fighting to keep attention. Incorporating regular, varied active learning moments is a great solution to recapture an audience.

Similarly, when you are learning DS/AI theory, try to apply as soon as what you have learnt. This will keep you engaged and motivated for a longer period of time.

## 8.3 Just Enough Approach



Just Enough Approach is learning just enough to perform a specific task. As we saw in the above sections that it does not help to be a course junkie. You need to learn in order to solve a problem. So don't get into an endless learning loop and just learn enough to solve the problem at hand.

Unless you apply what you have learnt, there is no benefit of learning. And the depth in the theory of DS/AI is kind of endless, without any problem to solve, you may go as much deep as you want and still get no idea how to apply what you have learnt.

*I always suggest my students to apply Just Enough Approach while learning DS/AI. This is the path of optimal learning, you learn the concept and solve the problem as well.*

As we learnt in this chapter, there are few obstacles in the path of learning and application of DS/AI and you can overcome them by just being aware that they exist.

## 8.4 Conclusion

We have looked at the overall landscape, what are DS/AI terminologies and what are the roles in DS/AI projects. Then we worked on the building blocks by listing out what are the concepts, processes, and tools you need to learn. After that, we listed the resources we need to refer to during the learning stage.

Post that we learned how to build an impressive portfolio. How we can start building a network and prepare for the interview in order to search and land the job. Then we looked at the ways to make our career future-proof.

*Now you have all the knowledge, concepts, processes & tools with you to tackle any challenge you face in DS/AI field.*

I believe all the content provided in this book is helpful to you. If you liked (or not) the content, I would ask you to provide your feedback to me.