

CAREER IN DATA SCIENCE / ARTIFICIAL INTELLIGENCE

Yogesh Haribhau Kulkarni



Outline

① BACKGROUND

② INTRODUCTION

③ ROLES

④ TOOLS

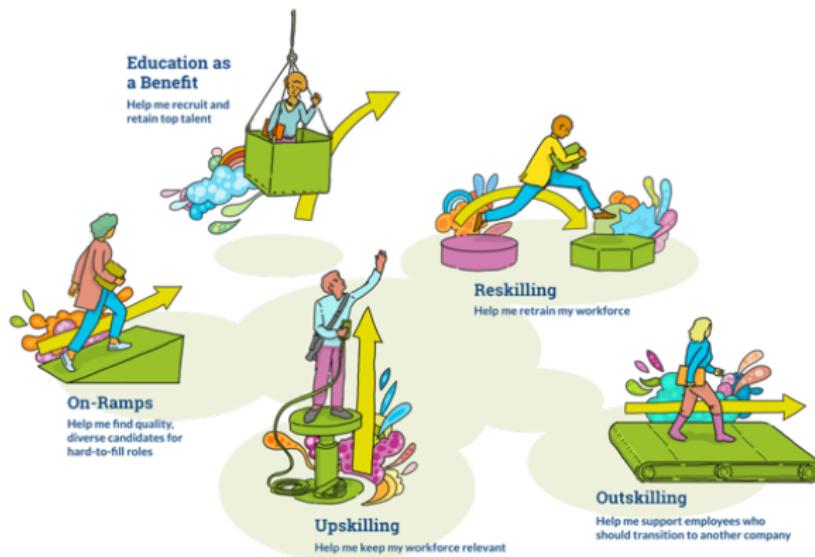
⑤ PREPARATION

⑥ REFERENCES

Career in AI: Background

Current State

- ▶ 44% of US workforce < \$18K/yr (< poverty line), works 80-100 hrs/week
- ▶ Automation CAGR 7% (as per BCG), to reach \$114B by 2025

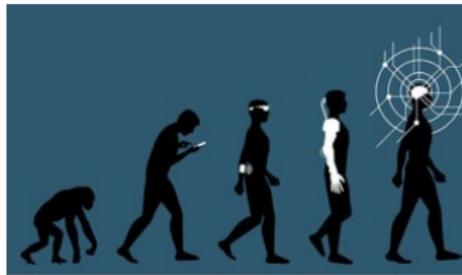


(Ref: As Pressure To Upskill Grows, 5 Models Emerge – Forbes.com)

Examples

Less mechanical, automatable

- ▶ Bill and account collectors
- ▶ Data entry operators
- ▶ Computer network support
- ▶ Secretaries and administrative assistants
- ▶ Insurance sales agents
- ▶ Office clerks



More Cognitive, Creative, Human

- ▶ Software developers
- ▶ Customer service representatives
- ▶ General and operations managers
- ▶ Human resources managers
- ▶ Personal finance advisors
- ▶ Psychologists
- ▶ Artists
- ▶ Sportsman
- ▶ Researchers
- ▶ Nurses, care

(Source: The Simplistic Debate Over Artificial Intelligence – Preston
Estep)

Changes

Technological

- ▶ AI deluge
- ▶ Digitization → Data + APIs
- ▶ Remote *

Social

- ▶ Over interaction + isolation
- ▶ Obsolete roles, emergence of new
- ▶ Lifelong re-skilling



(Source: Rise of the Chatbots: How AI Changed Customer Service – Salesforce.com)

Financial

- ▶ Widening gap
- ▶ Flatter world
- ▶ Gigs over jobs

Data Science is critical in bringing intelligent automation

What are Data Sciences?

ie What is Artificial Intelligence? Machine Learning? Deep Learning?

Data Science

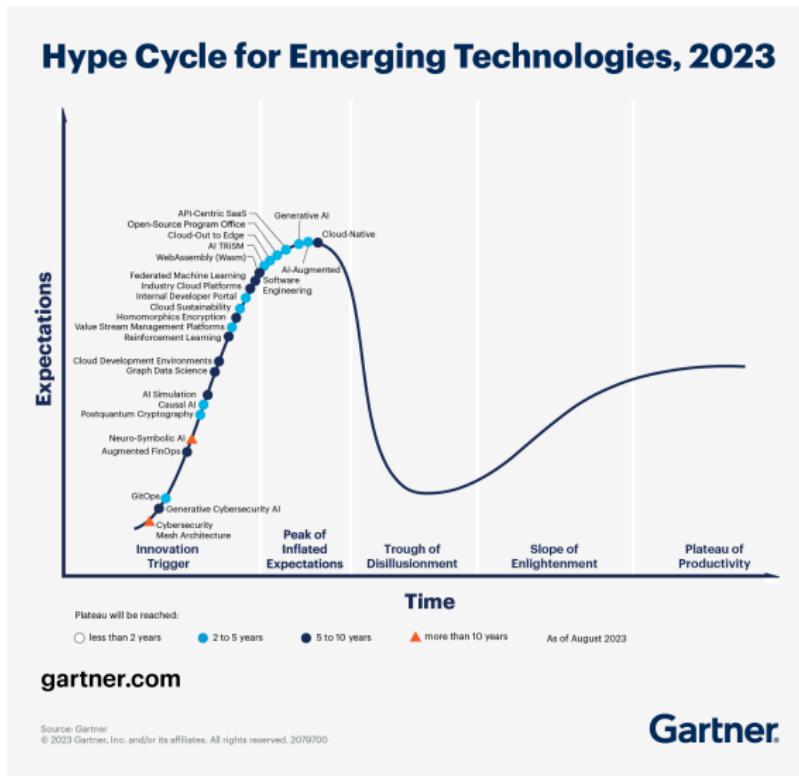
- ▶ Science of Data (obviously)
- ▶ Use of Data for Applications
- ▶ Some parts of AI uses Data to find patterns and insights which are helpful in multiple applications
- ▶ Machine and Deep Learning that part of AI that leverages data.

So, more on AI-ML here ...

Why now?

- ▶ Flood of data (Internet, IoT)
- ▶ Increasing computational power
- ▶ Easy/free availability of algorithms
- ▶ Increasing support from industries

Gartner Hype Cycle Emerging Technologies 2023



Introduction to Artificial Intelligence

The Problem

Every company is claiming to be working in AI-ML

- ▶ Is it really so?
- ▶ What exactly is AI (ML)?
- ▶ What is not AI?

Or is it just a plain BIG hype?

What is the Core Idea?

What's the core idea?

- ▶ behind problem solving?
- ▶ behind writing software algorithms?
- ▶ solving research problems?

Desire

- ▶ To find a “function”
- ▶ To find a relation
- ▶ To find a transformation
- ▶ To build a model
- ▶ From given inputs to desired outputs.

That's it.

Functions

- ▶ Some functions are straight forward
- ▶ *"In summer, ice-cream sale goes up"*
- ▶ Cause and effect
- ▶ Relation (function, Mathematical model) is found out
- ▶ Here, simple rule based programming suffices

Functions

- ▶ But some functions are complex
- ▶ *"More you put efforts, your business flourishes."*
- ▶ Cause and effect again, but the relation is far to complex
- ▶ Too many variables
- ▶ Here, simple rule based programming not humanly possible.
- ▶ Lots of research needed to come up with equations.

Functions

- ▶ $E = mc^2$
- ▶ What's this? a function?
- ▶ Input variable(s)?
- ▶ Output variable(s)?
- ▶ Parameters?
- ▶ How's the relation? linear?

Controversial Example

- ▶ Even astrology is a model, based on the past cases.
- ▶ Could claim imperical evidence.
- ▶ Given this planetary position, it predicts.
- ▶ Represented by “Horoscope”
- ▶ Got weights for each planets (real or fictitious)
- ▶ Reliable??

Functions

- ▶ But most real-life functions are not deterministic
- ▶ Some are probabilistic, some non-linear.
- ▶ *“Detecting if the tumor is benign or malignant”*
- ▶ *“At any state in the game of chess, what's the next move?”*

Chess: next move?

- ▶ Needs extreme expertise
- ▶ Needs “intelligence”
- ▶ How do you get that?
 - ▶ Built by lots of training.
 - ▶ By studying lots of past games.
- ▶ This is how Humans build intelligence

Intelligence

- ▶ Can machine (software/program) also do the same?
- ▶ Can it play chess?
- ▶ Can it build intelligence?
- ▶ By looking at past experiences (data),
- ▶ Training Data: games played, moves used, etc.

Yes, it can!! Thats Artificial Intelligence.

What is AI?

What is Artificial Intelligence (AI)?

My definition:

“If machines (or computer programs) start doing some/all of these “intelligent” tasks, then that’s Artificial Intelligence”

Intelligence: the differentiation

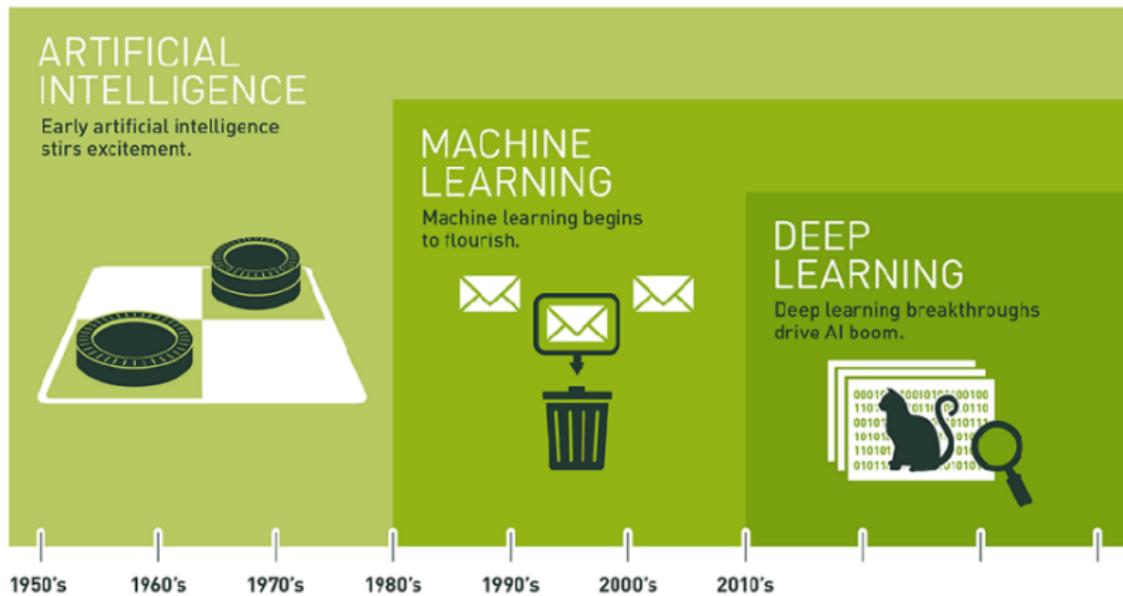
- ▶ Ability to think various domains
- ▶ Ability produce something new
- ▶ Ability to detect the unseen
- ▶ Ability to enhance knowledge (rules, patterns)

All these, AI has started doing. The AI era has arrived!!

What is Artificial Intelligence (AI)?

As Bernard Marr comments in Forbes, there is a need to distinguish between “the ability to replicate or imitate human thought” that has driven much AI to more recent models which “use human reasoning as a model but not an end goal”.

Relationship between AI, ML, DL



(Ref: <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>)

Everyday usage

Artificial intelligence seems to have become ubiquitous.

- ▶ Replying to our emails on Gmail
- ▶ Learning how to drive our cars,
- ▶ Sorting our holiday photos.
- ▶ etc.

Too good to be true, isn't it, sort of Magical !!

But then ...

- ▶ When its too good, you start suspecting
- ▶ Is it for real!!
- ▶ How can such thing happen?
- ▶ How far will it go?

The next thing you know, people are worrying about exactly how and when AI is going to doom humanity.

Is AI a threat?

Is AI a threat?

If you believe in what Elon Musk says, then YES.



Elon Musk recently commented on Twitter that artificial intelligence (AI) is more dangerous than North Korea

(Ref: What is Artificial Intelligence — Artificial Intelligence Tutorial For Beginners — Edureka)

Is AI a threat?

If you believe in these movies, then YES.



The Terminator



I, Robot



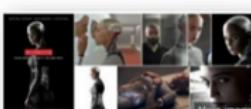
The Matrix



Tron: Legacy



War Games



Ex Machina

Well, AI based War robots are not impossible anymore.

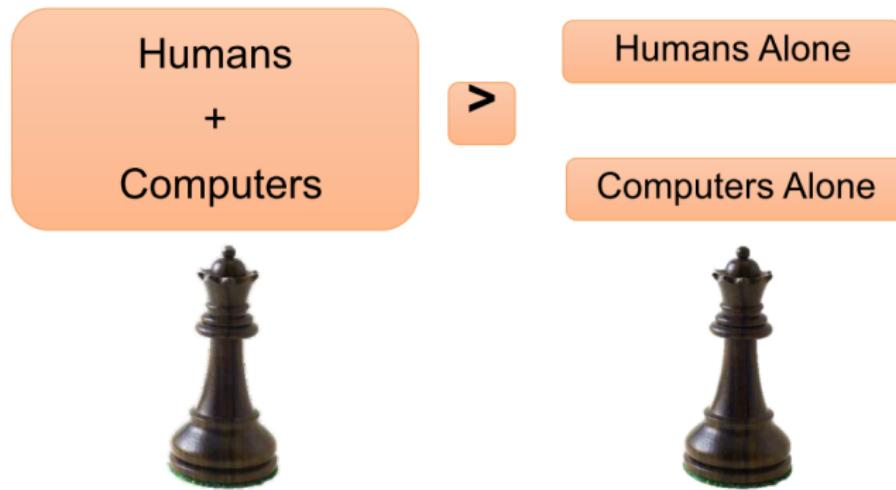
(Ref: What is Artificial Intelligence — Artificial Intelligence Tutorial For Beginners — Edureka)

Fear: Are we being replaced?

- ▶ Yes. in tasks that are repetitive
- ▶ But not which require complex thinking and creativity

Mostly

Technology Enhancing (Not Replacing) Humans



(Ref: "Artificial Intelligence Overview" - Harry Surden)

Limits on Artificial Intelligence

- ▶ Many things still beyond the realm of AI
- ▶ No thinking computers
- ▶ No Abstract Reasoning
- ▶ Often AI systems Have Accuracy Limits
- ▶ Many things difficult to capture in data
- ▶ Sometimes Hard to interpret Systems

Finding Your Persona . . .

Choosing Your Persona for Transition to Data Science

The USER Persona

- ▶ Leverage domain expertise.
- ▶ Utilize low/no code platforms like Weka or Knime.
- ▶ Build machine learning workflows easily.
- ▶ Ideal for managers, marketing professionals.
- ▶ Make data-driven decisions without extensive coding.

The DEVELOPER Persona

- ▶ Technical enthusiast with a programming flair.
- ▶ Expertise in scikit-learn, TensorFlow, PyTorch.
- ▶ Develop robust data science applications.
- ▶ Dive deep into machine learning techniques.
- ▶ Transform ideas into impact-ful solutions.

The RESEARCHER Persona

- ▶ Passionate about mathematics and innovation.
- ▶ Invent new techniques, contribute to research.
- ▶ Caters to deep R&D professionals, PhD holders.
- ▶ Shape the data science landscape with created libraries.
- ▶ Be at the forefront of groundbreaking discoveries.

Common to All

- ▶ Solid foundation in machine and deep learning is crucial.
- ▶ Regardless of persona, these principles are essential.
- ▶ Embrace a continuous learning mindset.
- ▶ Stay updated with industry trends.
- ▶ Be open to expanding knowledge horizons.

Choosing a Persona

- ▶ Select a persona based on skills and interests.
- ▶ Allow projects and interests to guide your journey.
- ▶ Explore new areas of expertise.
- ▶ Expand your skill set accordingly.
- ▶ The right persona empowers excellence and lasting impact.

So, Roles/Persona to chose from ...

- ▶ Researcher: Advanced Mathematics, Programming, domain, Adv ML/DL, etc. You invent new things and write NeurIPS paper.
- ▶ Implementation:
 - ▶ Data Analyst: Bridge the gap between the data scientists and the business analysts, organizing and analyzing data to answer the questions the organization poses.
 - ▶ Data Engineer: Focus on developing, deploying, managing, and optimizing the organization's data infrastructure and data pipelines.
 - ▶ Data Scientist: Use/mine data, clean, build applications, reports, etc
- ▶ User level: Product Manager, Sales, Manager, etc

Career in AI: Tools

Scikit-Learn



- ▶ Popular machine learning library
- ▶ Provides simple and efficient tools for data mining and data analysis

TensorFlow



TensorFlow

- ▶ TensorFlow is an open-source library for machine learning and artificial intelligence
- ▶ Developed by the Google Brain team
- ▶ Allows easy deployment of computation to CPUs, GPUs, and TPUs
- ▶ Provides a Python API as well as APIs for other languages

Key Features

- ▶ Eager Execution for interactive coding
- ▶ Keras for building and training models
- ▶ TensorFlow Hub for reusable models
- ▶ TensorFlow Lite for mobile and embedded devices
- ▶ TensorFlow Extended (TFX) for model pipelines

Code Example: MNIST Digits Classification

```
import tensorflow as tf

mnist = tf.keras.datasets.mnist
(x_train, y_train), (x_test, y_test) = mnist.load_data()

model = tf.keras.models.Sequential([
    tf.keras.layers.Flatten(input_shape=(28, 28)),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(10, activation='softmax')
])

model.compile(optimizer='adam', loss='sparse_categorical_crossentropy',
model.fit(x_train, y_train, epochs=5)
model.evaluate(x_test, y_test)
```

Future Developments

- ▶ Improved support for multi-cloud and hybrid environments
- ▶ Continued focus on performance, scalability, and efficiency
- ▶ Integration with emerging hardware like AI accelerators
- ▶ Expanded model libraries and pre-trained models
- ▶ Advanced features for research and experimentation

Pytorch



- ▶ Open-source machine learning library
- ▶ Developed by Facebook's AI Research lab (FAIR)

ChatGPT - A Tipping Point for Generative AI

- ▶ Released by OpenAI in November 2022
- ▶ Generative AI chatbot
- ▶ Rapid worldwide popularity
- ▶ 1 million users in 5 days
- ▶ Netflix took 3.5 years for same user count
- ▶ 100 million monthly active users by January 2023
- ▶ Fastest-growing application in history

Midjourney: Image Generation Model

- ▶ Developed by Midjourney Inc.
- ▶ Released in July 2022
- ▶ Architecture details undisclosed
- ▶ High-quality image generation
- ▶ Wide variety of styles and genres

LLaMA

- ▶ February 2023: Meta releases LLM "LLaMA"
- ▶ LLaMA: 65-billion parameter model
- ▶ Trained on extensive text and code dataset

Significance of LLaMA Release

- ▶ One of the largest public LLMs
- ▶ Suited for complex and challenging tasks
- ▶ Open source, initially for research purposes
- ▶ Model weights leaked online, accessible to all
- ▶ Sparked development of numerous open source LLMs

Anthropic Claude

The screenshot shows a user interface for generating product names. At the top, there is a title bar with a back arrow, intro, roles, tools, prep, and refs. Below the title bar, the main content area has a header "Fit Shaker: Fast and Healthy". On the right side of the header are "Chat" and "Edit Chat" buttons. To the right of the header is a three-dot menu icon. The main content area contains several input fields and suggestions:

- "Product names: HomeShaker, Fit Shaker, QuickShake, Shake Maker"
- "Product description: A pair of shoes that can fit any foot size."
- "Seed words: adaptable, fit, omni-fit"
- "Product names: M" (with a small circular icon)

Below these fields, a message says "Here are some suggested product names based on the seed words:" followed by a list:

- OmniFit
- AdaptiShoe
- UniSole
- One Size Fits All
- FlexiFit

At the bottom left is a text input field with placeholder "Write a message..." and a blue send button with a white arrow. At the very bottom of the interface is a footer with the text "(Ref: The Complete Prompt Engineering for AI Bootcamp (2023))".

Created by Anthropic
<https://console.anthropic.com/>
or API Uses Constitutional AI rather than RLHF
Constitutional AI trains to follow a set of high-level principles or rules, such as a constitution, that specify the desired behavior and outcomes of the system. RLHF uses human feedback, such as ratings, preferences, or corrections, to optimize a language model or an agent's policy using reinforcement learning

Github Copilot: Breakthrough Coding Assistant

- ▶ OpenAI introduced Github Copilot in 2021
- ▶ Built on GPT-3 architecture
- ▶ Fine-tuned on millions of public code lines
- ▶ Auto-completes and suggests code
- ▶ Supports multiple programming languages

Low/No Code Platform : Knime



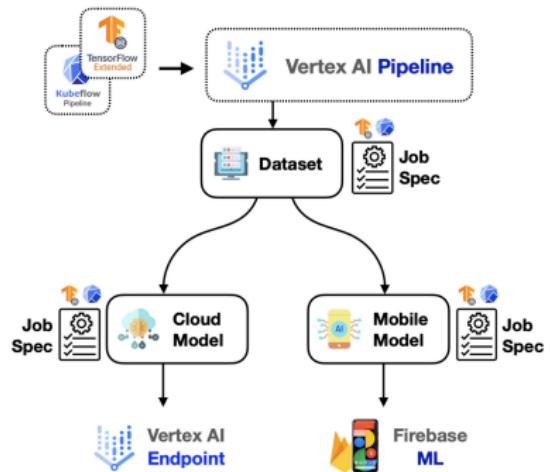
- ▶ Open-source low/no code platform for data analytics, reporting, and integration
- ▶ Offers visual programming interface

Low/No Code Platforms: Weka



- ▶ Open-source low/no code platform for data mining and machine learning
- ▶ Provides tools for data preprocessing, classification, regression, clustering, and association rules

Cloud Platforms: GCP



- ▶ Google Cloud Platform (GCP)
- ▶ Cloud computing services by Google
- ▶ Offers various AI and machine learning services such as Gen AI Studio, Vertex AI, etc.

Cloud Platforms; Azure



- ▶ Microsoft Azure
- ▶ Cloud computing services by Microsoft
- ▶ Provides AI and machine learning services like Azure Machine Learning, Cognitive Services, etc.

Cloud Platforms: AWS



- ▶ Amazon Web Services (AWS)
- ▶ Cloud computing services by Amazon
- ▶ Offers AI and machine learning services such as Amazon SageMaker, Amazon Comprehend, etc.

Career in AI: But . . . , But . . . , How to prepare?



What got you here, won't get you there!!
- Marshal Goldsmith

So, Well, you can't prepare!!
not everything, but certainly, specifically ...

Are you suitable?

Advantages:

- ▶ Domain Expertise
- ▶ Maturity, Communication, Soft Skills
- ▶ Problem Solving

Dis-advantages:

- ▶ Lost touch with Mathematics
- ▶ Un-Learning and Re-Learning inertia
- ▶ Starting from scratch? Seniority?

Why do you want to Switch?

- ▶ $\$_1\$_2\$_3 \dots \$_n$?
- ▶ Will remain in fashion forever?
- ▶ Hate my current job? no growth?

What's in it for me?

Current + ML combo?

- ▶ First : DON'T QUIT!!!
- ▶ Don't lose advantage due to domain expertise
- ▶ ML just another problem solving technique, IF DATA IS AVAILABLE
- ▶ Can you leverage domain expertise and apply ML there, a good/smooth transition?

Learning Path, Roadmap

Resources

- ▶ First : try Free Online resources, see how much you grasp
- ▶ No expensive (read, fees in lakhs) certification courses, to start with
- ▶ Test waters, gain some understanding of yourself then decide.

Transition to Data Science: Start Playing the Role

- ▶ Wish to be a Data Scientist? Start playing that role today.
- ▶ Take specific actions to embody the desired role.
- ▶ Tone of the suggestion: Begin playing the coveted role immediately.

Transition to Data Science : Build Foundation

- ▶ Take courses in necessary mathematics, programming, ML, and DL.
- ▶ Engage in assignments to solidify foundational knowledge.
- ▶ Lay the groundwork for a strong understanding of key concepts.

Transition to Data Science : Kaggle Competitions

- ▶ Participate in Kaggle competitions across various domains.
- ▶ Explore NLP, Image Processing, Time-Series, and more.
- ▶ Gain practical experience and exposure to diverse challenges.

Transition to Data Science : Specialize and Apply

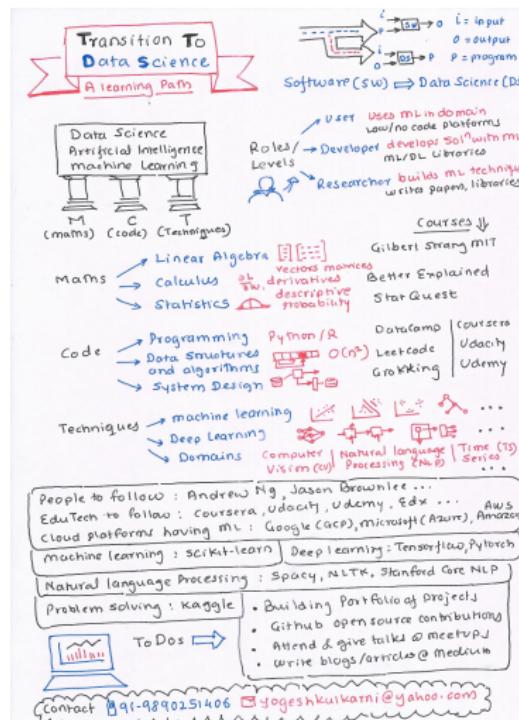
- ▶ Choose a specific area, e.g., NLP, and go deep into it.
- ▶ Apply your expertise to problems from different domains (legal, medical, etc.).
- ▶ Develop a comprehensive and specialized skill set.

Transition to Data Science : Build a GitHub Portfolio

- ▶ Showcase your work, courses, and projects on GitHub.
- ▶ Portfolio serves as a self-assessment tool and demonstrates your grasp.
- ▶ Discuss it during interviews, providing concrete evidence of your skills.
- ▶ Your GitHub repo is your real resume – proxies like education and gender matter less.



My Sketchnote



(Ref: How to become a Data Scientist? - Yogesh Kulkarni)

Summary Steps

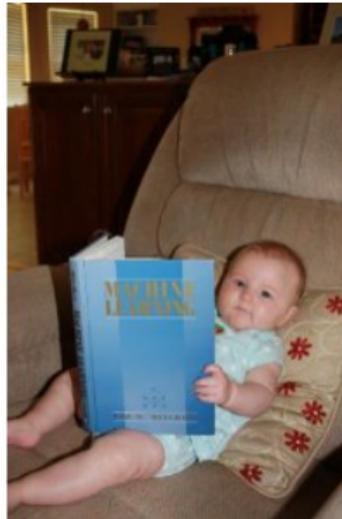
Prep:

- ▶ Mathematics: Statistics, Calculus, Linear Algebra
- ▶ Programming: Python, Data Structure & Algorithms, Tools
- ▶ ML/DL: algorithms & frameworks

Practice: Kaggle, Hackathons, projects on Github, blogs, Meetups-talks, etc.

Where do I start?

- ▶ Do a full workshop!!
- ▶ Coursera : Andrew Ng
Machine Learning Course
<https://goo.gl/fDTwSE> (uses
Octave!!!)
- ▶ Book: Programming Collective
Intelligence



Analytics Vidhya Learning Path 2017

- ▶ An year long schedule
- ▶ Mostly free resources
- ▶ Followed it myself
- ▶ Separate paths for:
 - ▶ Beginner: Not much experience in programming but just college maths
 - ▶ Transitioner: Decent experience programming, but no ML and just college maths
 - ▶ Intermediate: Knows ML, comfortable with programming and maths.

<https://www.analyticsvidhya.com/blog/2017/01/the-most-comprehensive-data-science-learning-plan-for-2017/>



More Generally . . . For Career . . .

Ikigai



(Source: How To Find Your Ikigai And Transform Your Outlook On Life And Business - Chris Myers)

Generic Gyan

Specific Knowledge

- ▶ Unique, rare combination
- ▶ Un-trainable, un-scalable
- ▶ Acquired through apprenticeship

Leverage

- ▶ Permission-ed: Capital, Labor
- ▶ Un-permission-ed: Content, code
- ▶ Marginal cost of duplication

Wealth is a positive sum game

Foreword by
TIM FERRISS

THE ALMANACK OF NAVAL RAVIKANT

A guide to wealth and happiness
ERIC JORGENSEN

Own story

- ▶ Current jobs were not available 10 yrs back
- ▶ Pick difficult problems/domain, Job, RnD
- ▶ If not now, can change later, Mid career change, harder, but possible.

YHK

Career in AI: References

References

- ▶ How to become a Data Scientist? - Yogesh Kulkarni
- ▶ Finding Your Persona - Yogesh Kulkarni
- ▶ Mid-Career Transitions into ML-AI, with Yogesh Kulkarni - Choose To Thinq
- ▶ Learning Plan 2017 for beginners in data science - Analytics Vidhya
- ▶ Mid-Career Transitions into ML-AI, with Yogesh Kulkarni - Choose To Thinq <https://www.youtube.com/watch?v=IQzWosVzkM4>
- ▶ What is Data Science? - SimpliLearn
- ▶ Roadmap: How to Learn Machine Learning in 6 Months - Zach Miller, Senior Data Scientist at Metis
- ▶ Tetiana Ivanova: How to become a Data Scientist in 6 months
- ▶ How to switch career to data science from non computer science background - Codebasics
- ▶ Step by step roadmap for machine learning engineer - Codebasics

Thanks ...

- ▶ Search "**Yogesh Haribhau Kulkarni**" on Google and follow me on LinkedIn and Medium
- ▶ Office Hours: Saturdays, 2 to 5pm (IST); Free-Open to all; email for appointment.
- ▶ Email: yogeshkulkarni at yahoo dot com



(Generated by Hugging Face QR-code-AI-art-generator,
with prompt as "Follow me")