Workshop Name:
Intro to Data Science w/ Python

Workshop Host:
Tashi T. Gurung

What is Data Science?

Artificial Intelligence (AI)

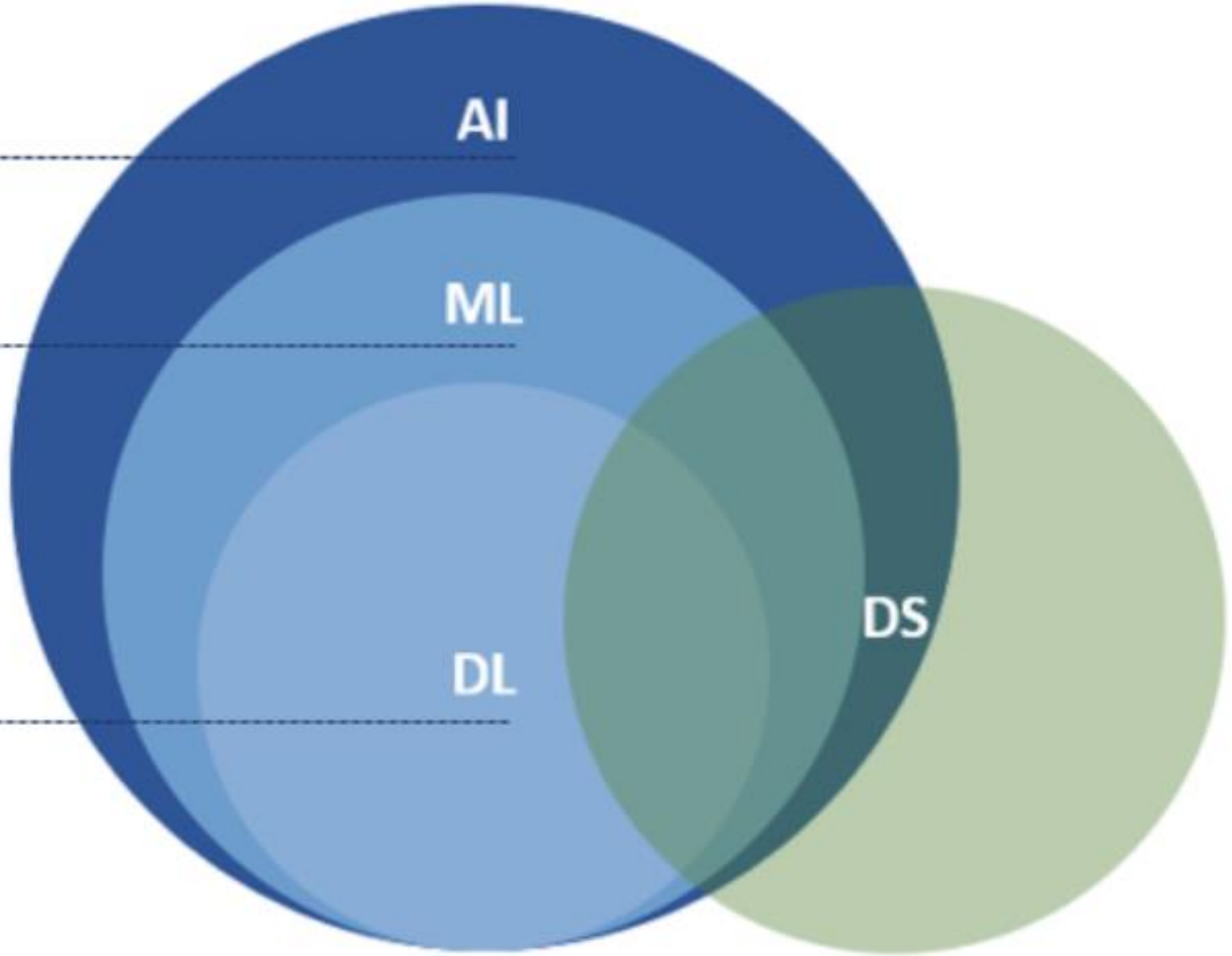
Machine Learning (ML)

Deep Learning (DL)

ARTIFICIAL INTELLIGENCE machines simulate the human thinking process through methods ranging from simple if-then statements to complex models.

MACHINE LEARNING machines analyse vast amounts of data searching for patterns to answer a very specific question. ML improves its decisions based on experience.

DEEP LEARNING is based on deep neural networks similar to the networks of the human brain. These machines work with deep models (i.e. models with several layers).



What is Data Science?

- Google Search:
  Data Science is an inter-disciplinary field, that uses [different tools] *scientific methods, processes, algorithms and systems* to extract knowledge and insights from *many structural and unstructured* data.

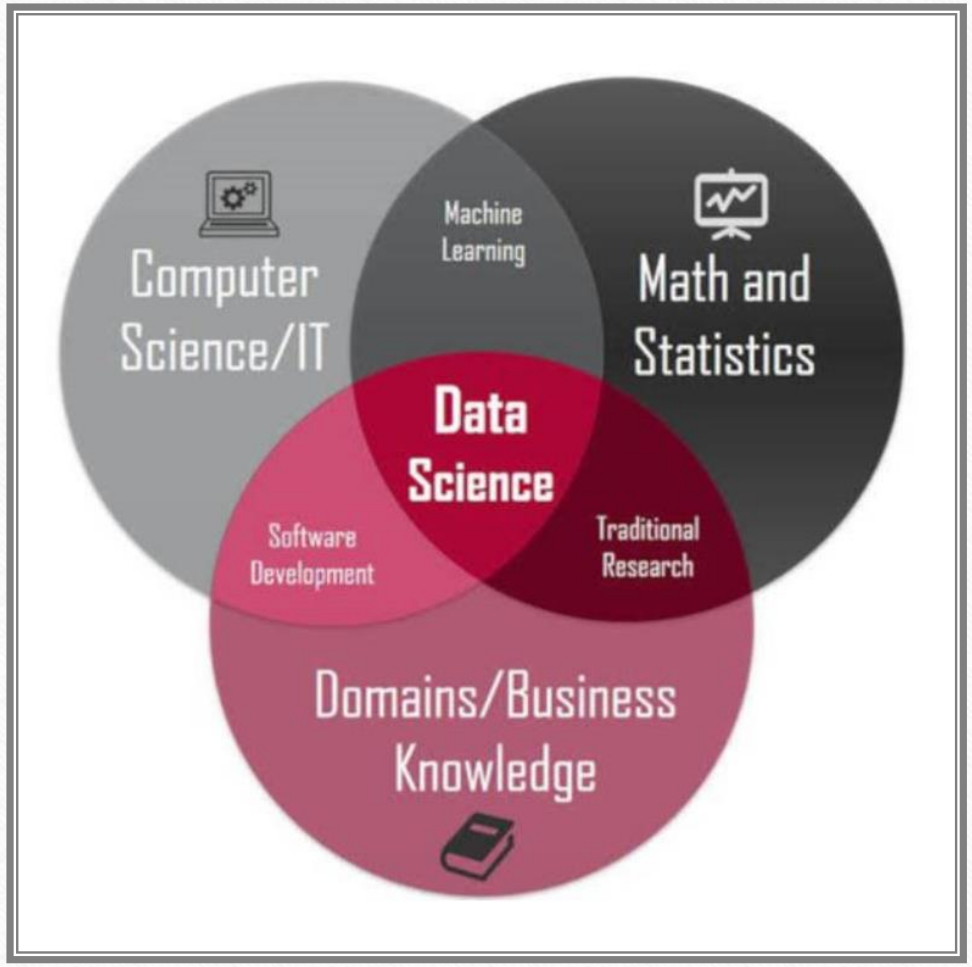
data scientist
Search term

statistician
Search term

+ Add comparison

Worldwide | 2004 - present | All categories | Web Search

Interest over time



Intersection of

- Computer Science
- Statistics
- Business

What are the different roles within the field of Data Science?

- Data Analyst
- Business Analyst
- Business Intelligence
- Data Scientist
- Data Engineer
- Data Architect
- Statistician
- Machine Learning Engineer

The Data Science Umbrella

Data Analyst

Data Scientist

Machine Learning Engineer

D[illegible]ngineer

Data Analyst

Responsibilities

-Curate the data
-Visualize and report data

Tools

- Excel
- SQL
- Tableau

Skills

- Analytics
- Communication & Visualization

Business Facing

- Sometimes

Salary

$65,000

Data Scientist

Responsibilities

-Source data
-Analyze data
-Run experiments
-Build models
-Recommend solutions
-Storytell

Tools

- Python
- R
- SQL

Skills

- Analytics
- Communication
- Story-tell
- Model building
- Math
- Coding

Business Facing

- Yes

Salary

$120,000

Data Engineer

Responsibilities

-Build data pipelines and warehouse
-Manage scalability of data products

Tools

- Java
- C++
- Kubernetes
- Hadoop
- Spark
- Python

Skills

- Coding
- Model implementation

Business Facing

- Not Really

Salary

$110,000

Python vs R

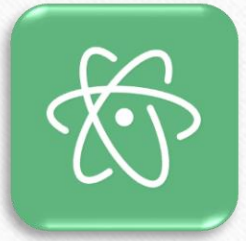
- R is great for statistical modeling
- Used in research extensively, loved by academics
- Great for time-series analysis

- Python is a general-purpose language.
- More widely used in the industry
- Better for integration and deployment
- Better for Deep Learning

Data Science w/ Python

Integrated Development Environment (IDE)

- Software application
- Popular IDEs:
  - PyCharm
  - VS Code
  - Atom
  - Sublime Text



Jupyter Notebooks

- Web applications that allow you to write and run code
- Run code cell-wise

Jupyter Notebooks

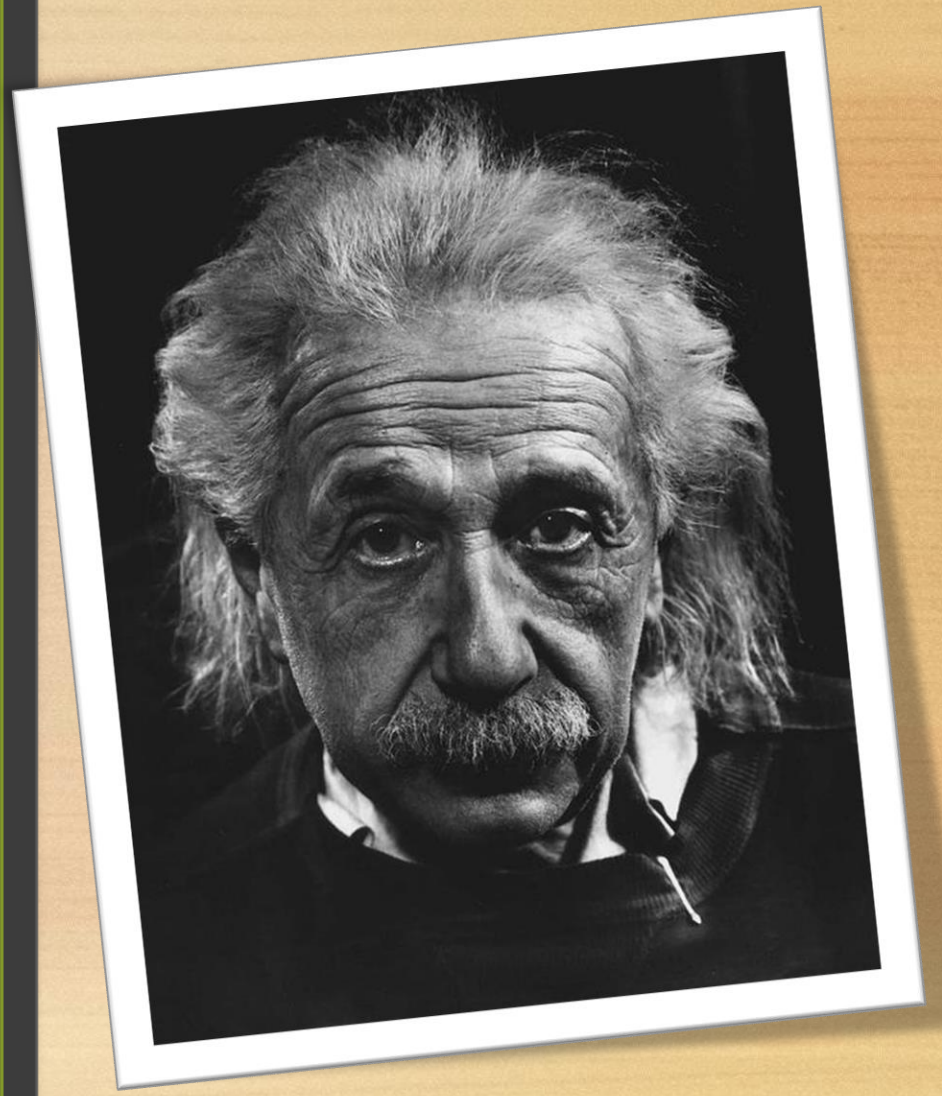
Anaconda Navigator

- Install in local machine

Google Colab

- Run Online

Questions?



CODING TIME !!!

Resources:

- **Harvard Business Review:**
  - Data Scientist: The Sexiest Job of the 21st Century(Oct 2012)

References

Images

- AI vs ML vs DL:
  https://www.abrisconsult.com/artificial-intelligence-and-data-science/
- Data Science Umbrella:
  https://mathamagicians.co/2020/04/06/differences-between-a-data-analyst-data-scientist-and-data-engineer/
- Intersection of image:
  https://dzone.com/articles/why-you-should-learn-data-science-with-python-in-2