DATA SCIENCE

What is Data Science?

“Sexiest Job in the 21st century?”

- Data science is the field of exploring, manipulating, and analyzing data and using data to make answer questions or recommendation.

- Statistics, Visualization, Deep Learning, Machine Learning, are important Data Science Concepts.

- Data Science is not a discipline traditionally taught at the universities.

- Continuous learning of new tools and patience to clean and analyze the data is the secret skills to success in the field of Data Science.

Data Scientists

- Someone who finds solutions to problems by analyzing big or small data using appropriate tools and then tells stories to communicate findings to the relevant stakeholders.

- **Qualities:**
  - Curious,
  - Judgemental,
  - Argumentative,
  - Story teller, meaning, find somethings from data and tell whole world about your findings from data.

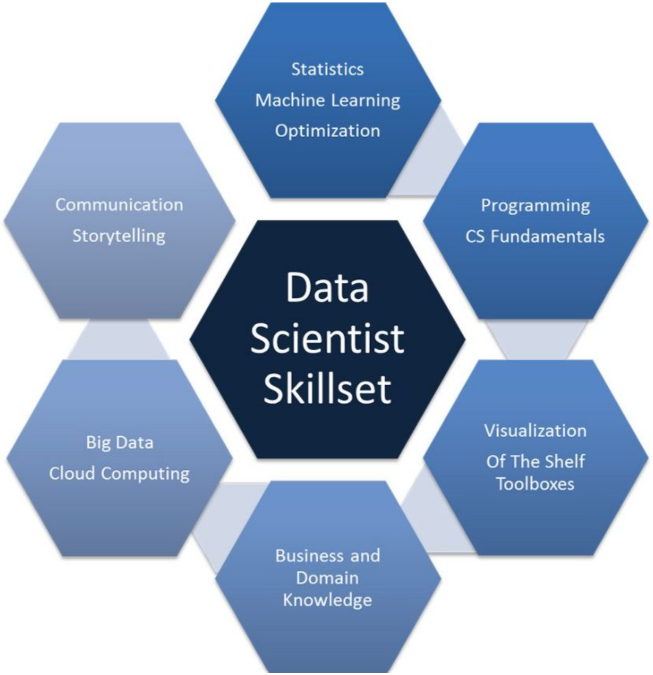
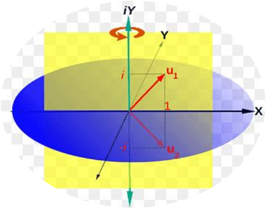


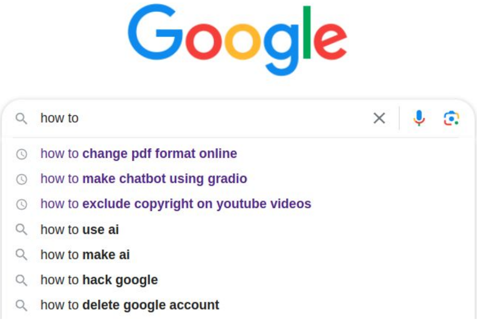
Image Source

For Successful Data Science Careers

- **Learn how to program**
- **Learn some math** (calculus, Linear Algebra)
- **Take a course in probability**
- **Learn some statistics**
- **And then play with the data**

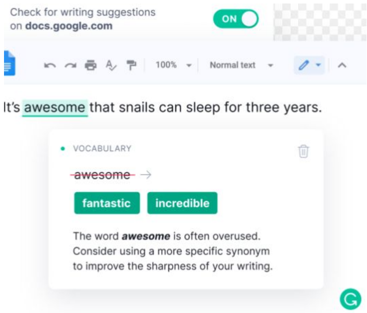


Applications



how to

- how to **change pdf format online**
- how to **make chatbot using gradio**
- how to **exclude copyright on youtube videos**
- how to **use ai**
- how to **make ai**
- how to **hack google**
- how to **delete google account**



Check for writing suggestions on **docs.google.com** ON

100% Normal text

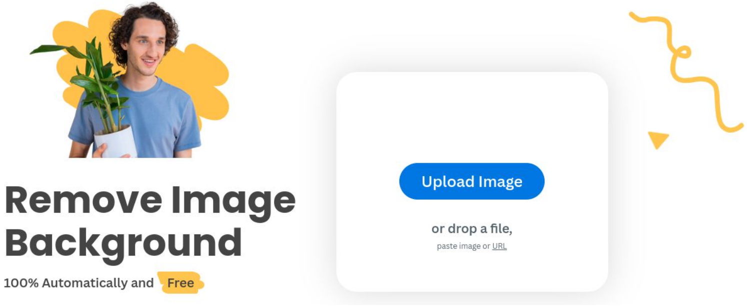
It's awesome that snails can sleep for three years.

VOCABULARY

~~awesome~~ →

fantastic incredible

The word ***awesome*** is often overused. Consider using a more specific synonym to improve the sharpness of your writing.



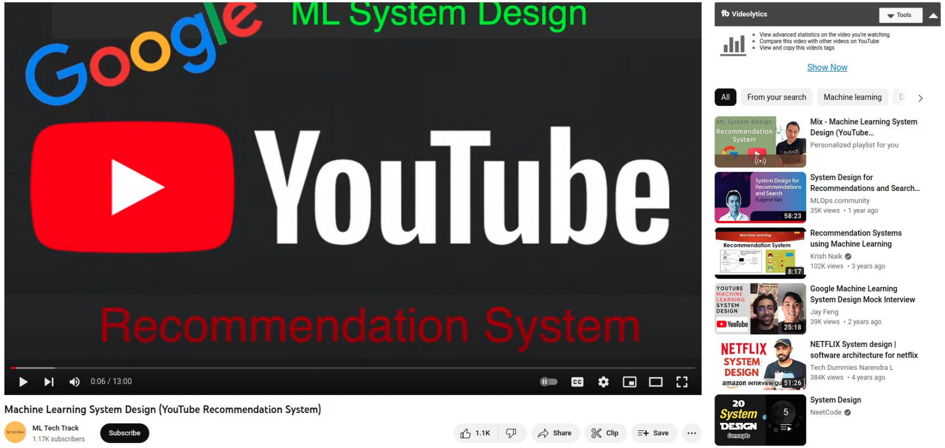
Remove Image Background

100% Automatically and Free

Upload Image

or drop a file,

paste image or URL



0:06 / 13:00

Machine Learning System Design (YouTube Recommendation System)

ML Tech Track
1.17K subscribers
Subscribe

1.1K Share Clip Save

Videolytics Tools

- View advanced statistics on the video you're watching
- Compare this video with other videos on YouTube
- View and copy this video's tags

Show Now

All | From your search | Machine learning

Mix - Machine Learning System Design (YouTube...
Personalized playlist for you

System Design for Recommendations and Search...
MLOps.community
35K views • 1 year ago

Recommendation Systems using Machine Learning
Krish Naik
102K views • 3 years ago

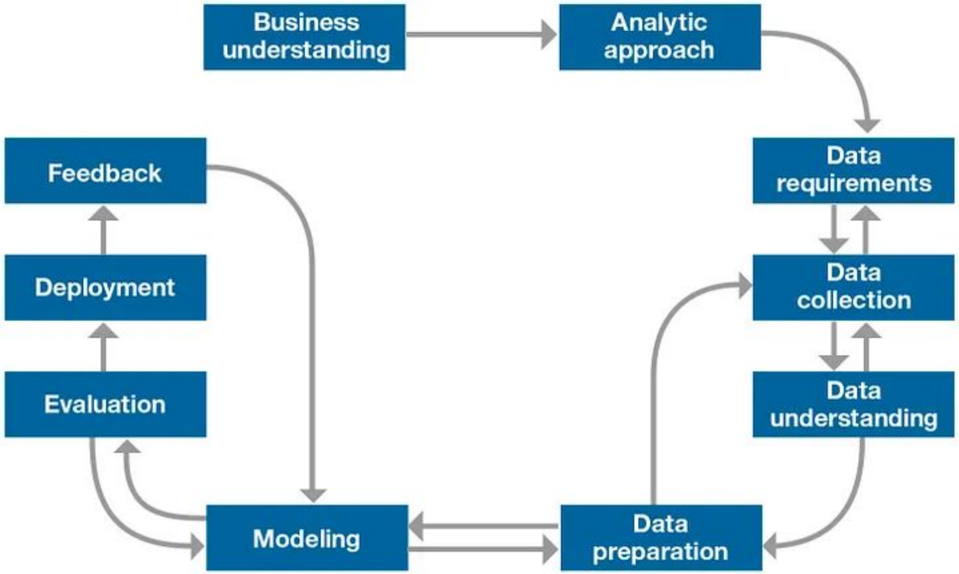
Google Machine Learning System Design Mock Interview
Jay Feng
39K views • 2 years ago

NETFLIX System design | software architecture for netflix
Tech Dummies Narendra L
384K views • 4 years ago

System Design
NeetCode

Data Science Methodology (CRISP-DM)

1. From Problem to Approach
2. From Requirements to Collection
3. From Understanding to Preparation
4. From modeling to Evaluation
5. From Deployment to Feedback



Source: *cognitive class.ai*

From Problem to Approach

1. **Business Understanding**
   - Helps to clarify customer goals.
   - Ask a lot of questions to the customer about every aspect of the problems.
   - **Question**: What does the business need?
   - **Outcome**: List of Business requirements

2. **Analytic Approach**
   - Express Business problem in the context of statistical and machine-learning techniques.
   - **Outcome:** Decision on whether to use,
     - Descriptive Statistics
     - Predictive Statistics
       - Forecasting, Classification, Regression, etc

Source: *cognitive class.ai*

From Requirements to Collection

3. **Data Requirements**
   - Identify the necessary data content, formats, and sources for initial data collection.
   - Identify internal and external data sources that provide relevant information. These sources can include databases, APIs, spreadsheets, text files, and more.
   - Collaborate with Domain experts and stakeholders to define the scope of the data requirements.
   - Document the data requirements clearly to guide subsequent phases of the project.

4. **Data Collection**
   - Gather data from various sources that align with the defined data requirements.
   - Extract the data using appropriate tools (like web scraping) or programming languages like SQL or Python.
   - Document the data collection process, including details about sources, extraction methods, and any transformations applied.

Source: *cognitive class.ai*

From Understanding to Preparations

5. **Data Understanding**
   - Helps to answer the questions ***“Is the data you’ve collected representative of the problem to be solved?”***
   - Different Techniques:
     - Descriptive Statistics
     - Exploratory Data Analysis:
       - histogram, scatter plots, correlation analysis etc
     - Verify Data Quality:
       - check for missing values, outliers, data type etc
6. **Data Preparation**
   - This step is essential for Modeling section of CRISP-DM.
   - Steps:
     - Feature Engineering
     - Data cleaning
     - Data Transformation
     - Data Integration
     - Data splitting
     - And many more……

Source: *cognitive class.ai*

From Modeling to Evaluation

7. **Modeling**
   - Focuses on developing the models that are either descriptive or predictive.
   - This step is based on the analytic approach that was taken, either statistically driven or machine learning driven.
   - Descriptive Modeling describes the real world events and the relationships between factors responsible for them.
     - Business Reporting in the form of graphs, charts, and dashboards.
   - Predictive model is a process that uses data mining and probability to forecast outcomes.
     - Example: whether an email is spam or not.
8. **Evaluation**
   - Modeling and Evaluation done iteratively.
   - Evaluation is performed during model development and before the model is deployed.
   - Evaluation answers the question:
     - ***Does the model used really answer the initial questions?***

Source: *cognitive class.ai*

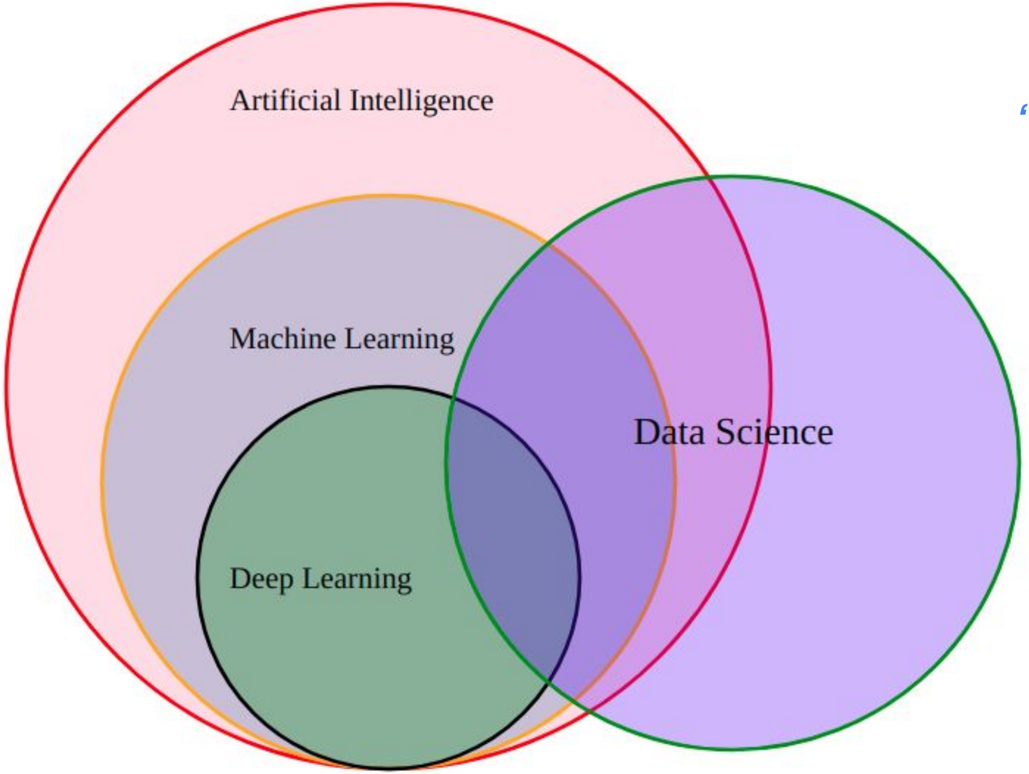
From Deployment to Feedback

9. **Deployment**
   - Once model is evaluated, it is deployed and put to the ultimate test.
   - Initially it may be rolled out to a limited group of users or in a test environment before making available to use across the world.

10. **Feedback**
    - Feedback from the customers/users is collected.
    - Data scientist collects the feedback to decide if they should improve the model that's because the process from modeling to feedback is highly iterative.

Source: *cognitive class.ai*

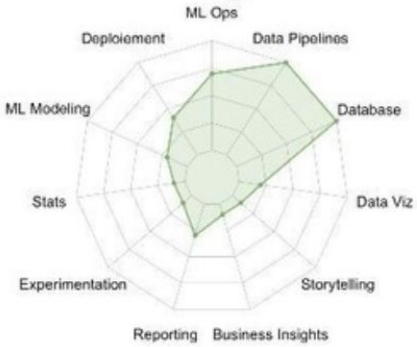
Terminologies



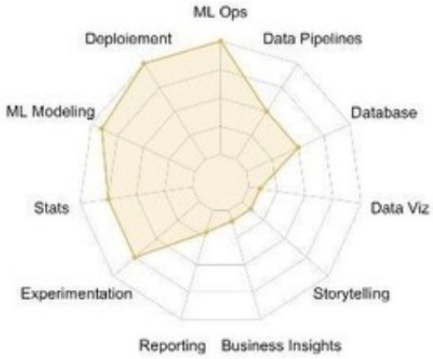
“Garbage In Garbage Out”

Types of Data Professionals

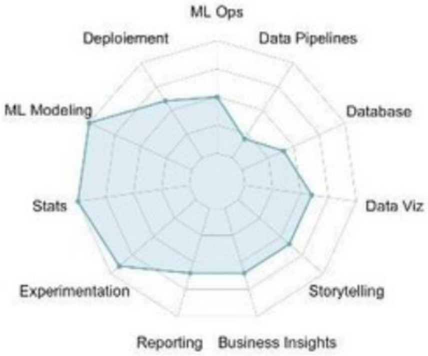
Data Engineer



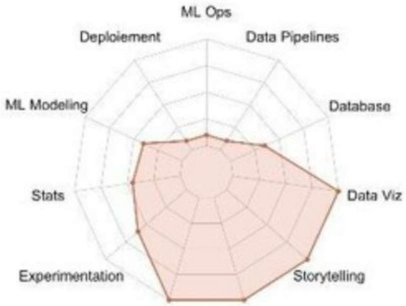
ML Engineer



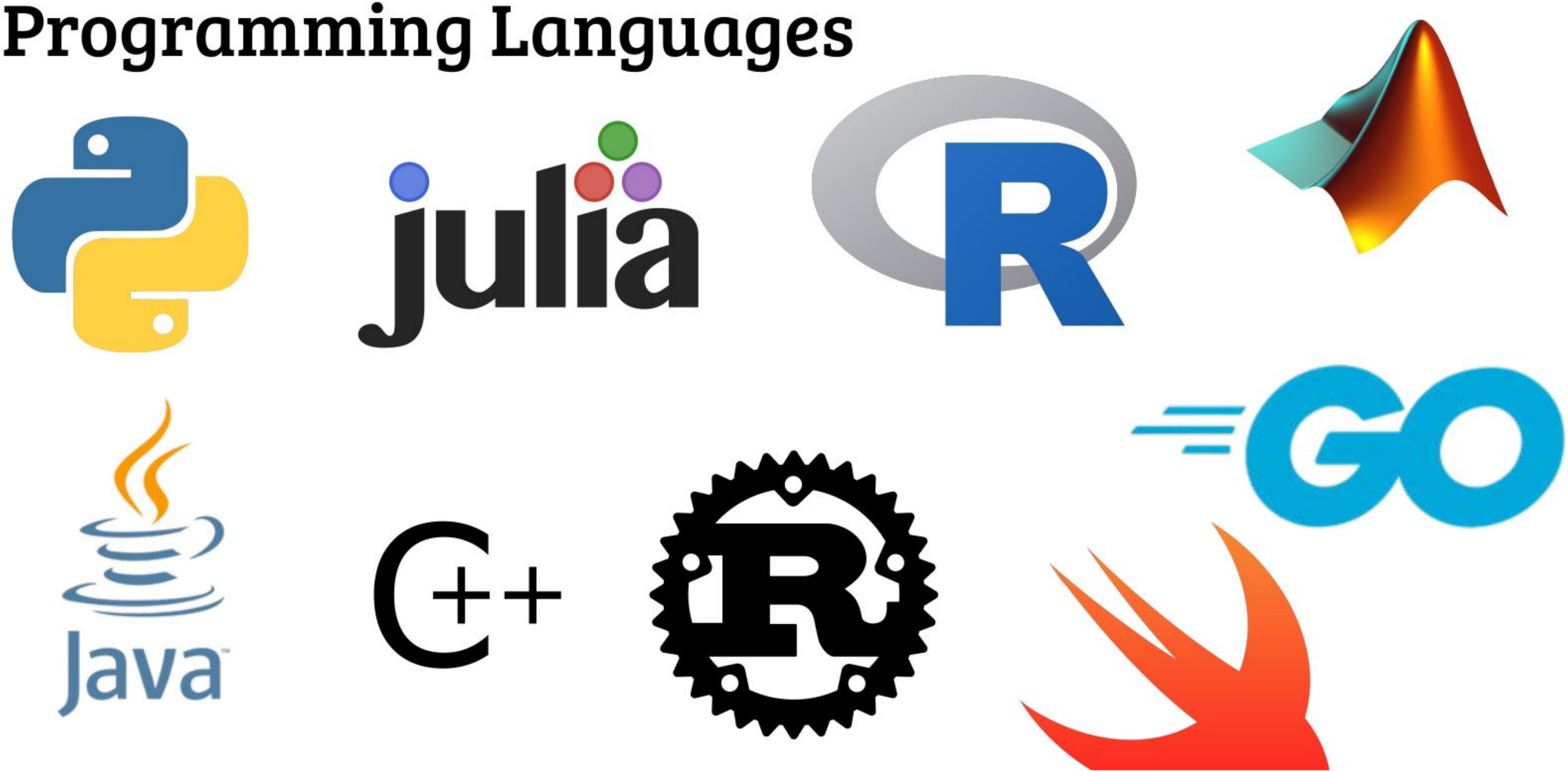
Data Scientist



Data Analyst



Programming Languages



Python Frameworks

Data Sources

kaggle

UCI Machine Learning Repository

aws

AWS Registry

NASA | EARTHDATA

NOAA

DATA.GOV

NASA | EARTHDATA

Cloud Platform (Open Sources)

colab

kaggle

IBM Cloud

- https://towardsdatascience.com/data-science-methodology-101-ce9f0d660336