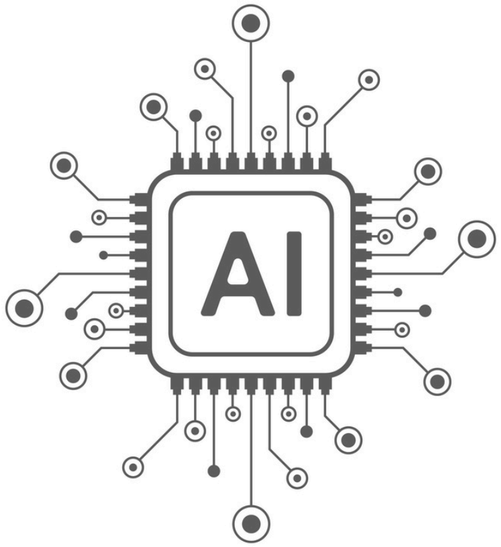


DSS 740
Analytics with Machine Learning

Sid Chakravarty
Adjunct Professor
Haub School of Business
Department of Decision and System Sciences
Saint Joseph's University
schakravarty@sju.edu

Course Housekeeping Items: Lecture and Office Hour Schedule

Class Meetings	Office Hours
Wednesdays 7:45 pm - 9:00 pm ET	Fridays 8:00 pm - 9:00 pm ET

- Meeting invite links are available on Canvas.
- All meetings will be recorded.
- Everyone needs to have their camera on during lectures.
- If you are unable to attend a lecture or office hour, let me know in advance.

SJU Saint Joseph's University

My Learning Philosophy: Make classroom mistakes, so you succeed in the real world!

Discussions [12%]

Application-oriented discussions will allow everyone to go beyond theory.

Homework [30%]

Homework assignments will focus on Python implementation. Partial credit will be awarded.

Quiz [8%]

Quizzes will test your understanding of a topic. Partial credit will be awarded.

Project [25%]

A comprehensive project will be assigned to test the knowledge gained throughout the course. The focus will be on application and interpretation.

Final Exam [20%]

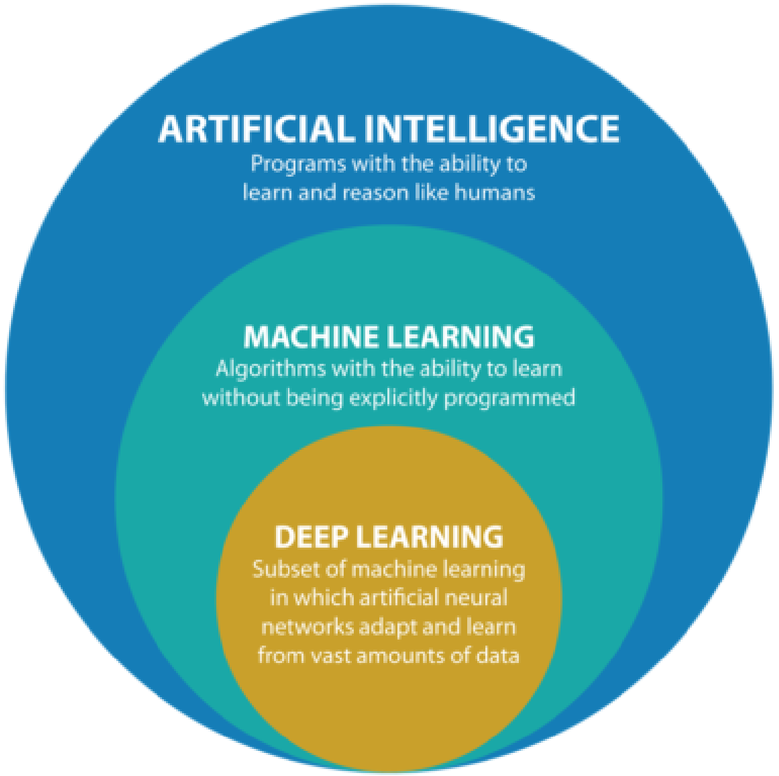
A comprehensive exam on topics covered in modules 1 - 6 (NLP will not be included in the final exam).

Class Participation [5%]

Participation in class meetings and office hours through insightful comments or questions related to the lecture is expected.

- Use of any technology to generate code snippets or complete code is not allowed.
- You may refer to Chat-GPT to learn about a topic, but must never use the Chat-GPT output directly in the assignment.
- If you get stuck, ask questions. Use Slack to help each other be successful.
- Partial credit will be awarded to those who try.

Is it Artificial Intelligence or Machine Learning?



Ability to distinguish

Ability to create new content

	Discriminative	Generative
Natural Language Processing	Text Classification Spam Detection	LLMs Chatbots
Computer Vision	Object Detection Facial Recognition	DALL-E (Image Generation)

Aggarwal, S. (2024, April 9). Deep Learning: the latest trend in AI and ML | Qubole. Qubole. https://www.qubole.com/blog/deep-learning-the-latest-trend-in-ai-and-ml

SJU Saint Joseph's University

Can machines truly learn, or are they sophisticated pattern matchers?

Do machines understand?

If a model can distinguish a dog from a cat, does it know what it means to have a Dog?

Can machines rationalize?

Given a situation, can a machine truly justify its actions?

Who is responsible?

If a machine makes a mistake, who needs to own up to that mistake?

SJU Saint Joseph's University

What is Machine Learning?

Arthur Samuel

"Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed."

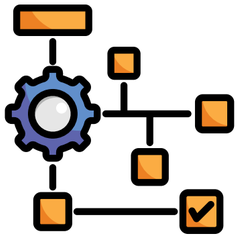
Tom Mitchell

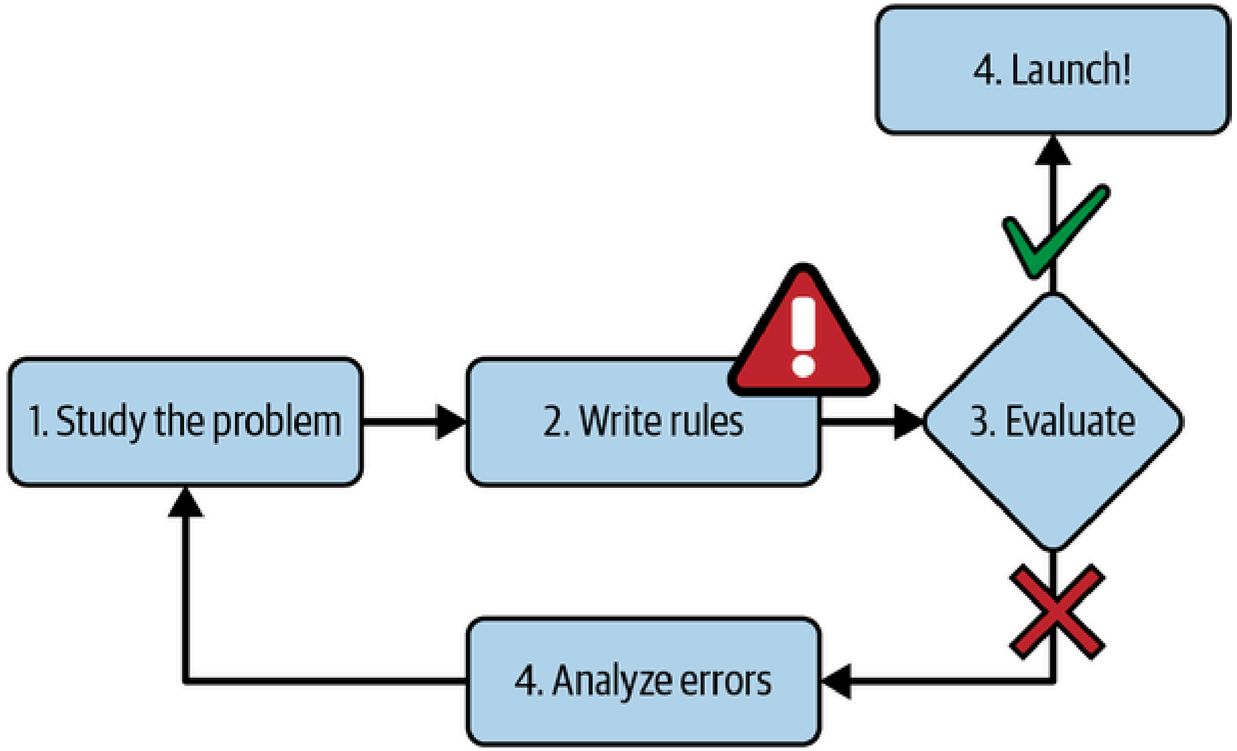
"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E"

Additional notes

- **Arthur Samuel**: He was a computer scientist who graduated from MIT and worked at Bell Labs, the University of Illinois, IBM, and Stanford University. He coined the term 'Machine Learning' and built the first checkers playing AI agent using a technique called Reinforcement Learning and Minimax algorithm.
- **Tom Mitchell**: He is a computer scientist who has made tremendous contributions in the field of Machine Learning. He earned degrees from Stanford University and MIT, and his book 'Machine Learning' has become a seminal text.

What is Machine Learning?

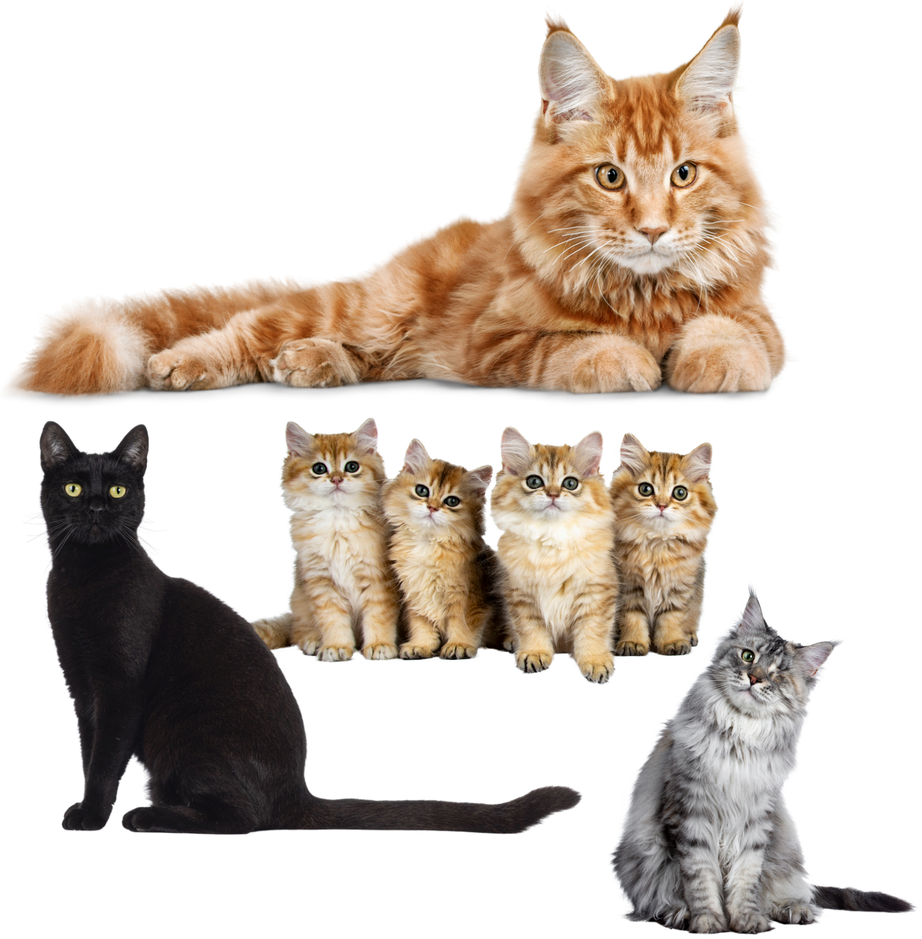




Géron, A. (n.d.). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition. O'Reilly Online Learning. https://learning.oreilly.com/library/view/hands-on-machine-learning/9781098125967/ch01.html#what_is_machine_learning

SJU Saint Joseph's University

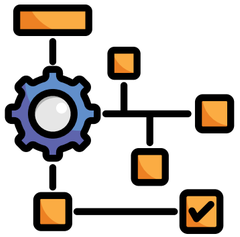
Identify at least 10 ways to distinguish cats and dogs in your respective teams.

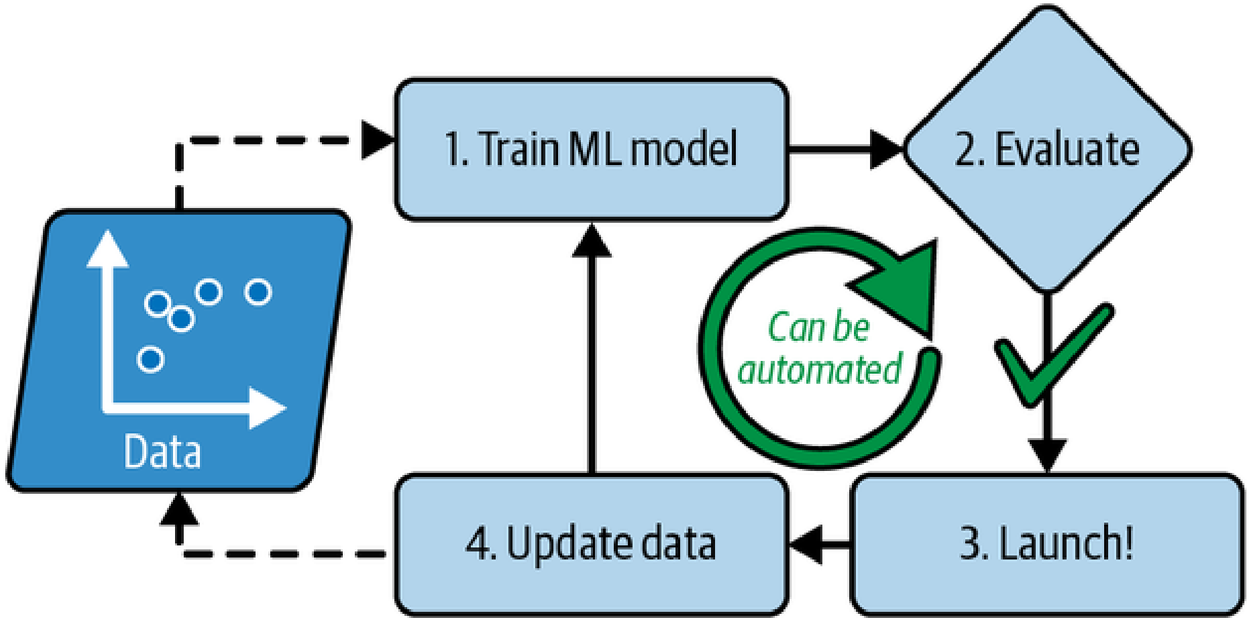


Now, let’s identify the common traits and put in a spreadsheet

	Color	Height	Fur	Sound	Likes sleeping	Behavior	Eyes	Is it a Dog or a Cat?
Animal 1								
Animal 2								
…								
Animal*N*								

So, what is Machine Learning?



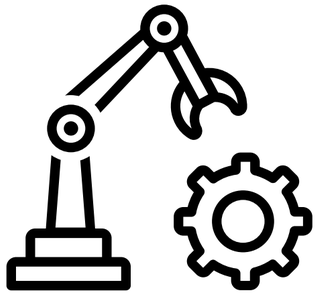


Géron, A. (n.d.). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition. O'Reilly Online Learning. https://learning.oreilly.com/library/view/hands-on-machine-learning/9781098125967/ch01.html#what_is_machine_learning

SJU Saint Joseph's University

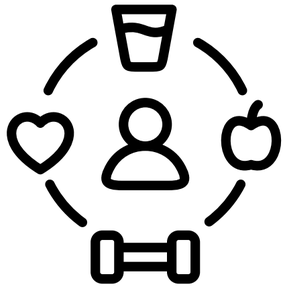
Let's look at some of the most *EXCITING* and PROMISING examples of AI deployment in the real world?

Manufacturing

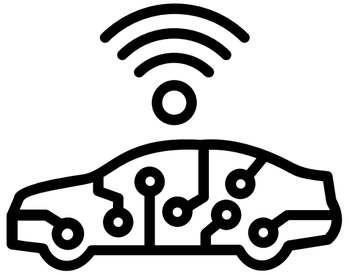


Healthcare

Lifestyle

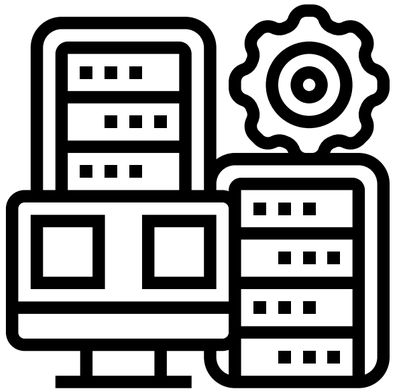


Autonomous Driving

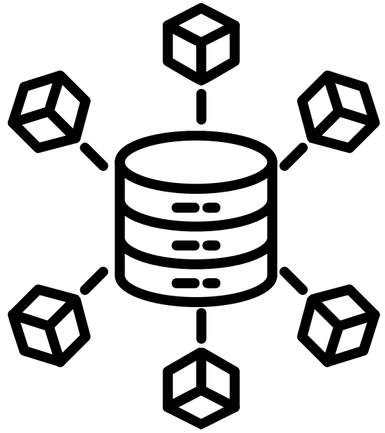


SJU Saint Joseph's University

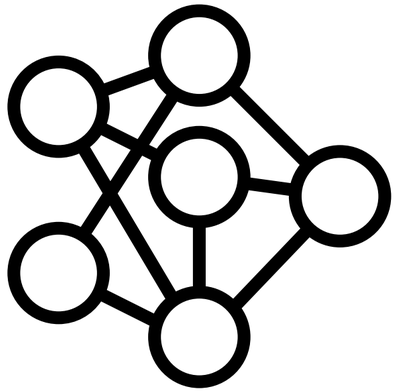
Why is AI so popular nowadays?



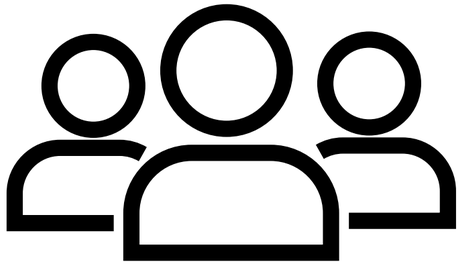
High-end Technology Hardware



Availability of Data



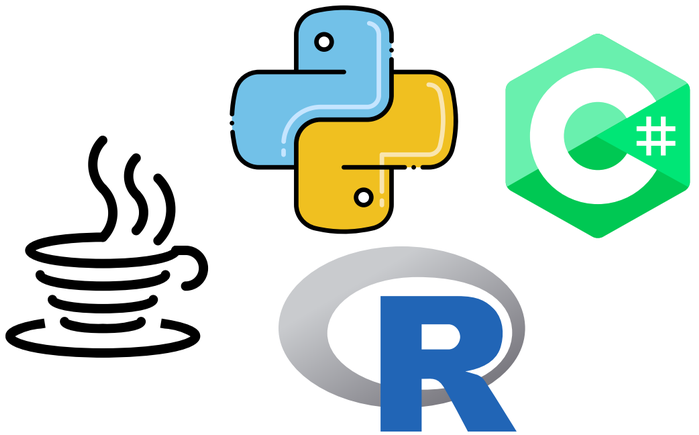
Sophisticated Algorithms



People / Leadership

Several languages can be used to build AI models

Language	Execution Speed	Ease of Use	Available AI Libraries
Python	Average	Very Easy	Excellent
R	Average	Average	Good
Java	Fast	Average	Good
C# / C++	Very Fast	Difficult	Average



OnCode. (2024, November 26). Top 5 most used programming languages in AI - Oncode Agency. Oncode Agency. https://oncode.ca/en/blog/top-5-most-used-programming-languages-in-ai/

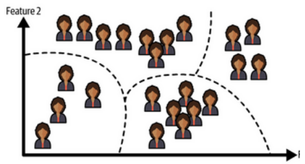
Machine learning systems can be classified in three different ways

Training Supervision

How is the machine getting trained?

Supervised

Unsupervised



Semi-Supervised

Supervised: The training data includes the desired results. The algorithm tries to figure out the patterns in the data that leads to the desired result

Unsupervised: The data does not have a label (desired result). The algorithm tries to learn the patterns based on how similar or farther apart the data points are.

Semi-Supervised: The algorithm relies on limited data labels to create labels for the remaining data and then a supervised algorithm helps classify the data or predict outcome.

Batch vs. Online Learning

Can the machine learn incrementally?

Batch Learning: To learn, the system must use all available data. Once the system has been trained, it is launched into production and runs without learning anymore. In production, the system applies what it already has learned. This is also known as *offline* learning.

Online Learning: The system is fed data instances sequentially, either individually or in mini-batches, resulting in faster learning. This technique allows the system to change extremely rapidly

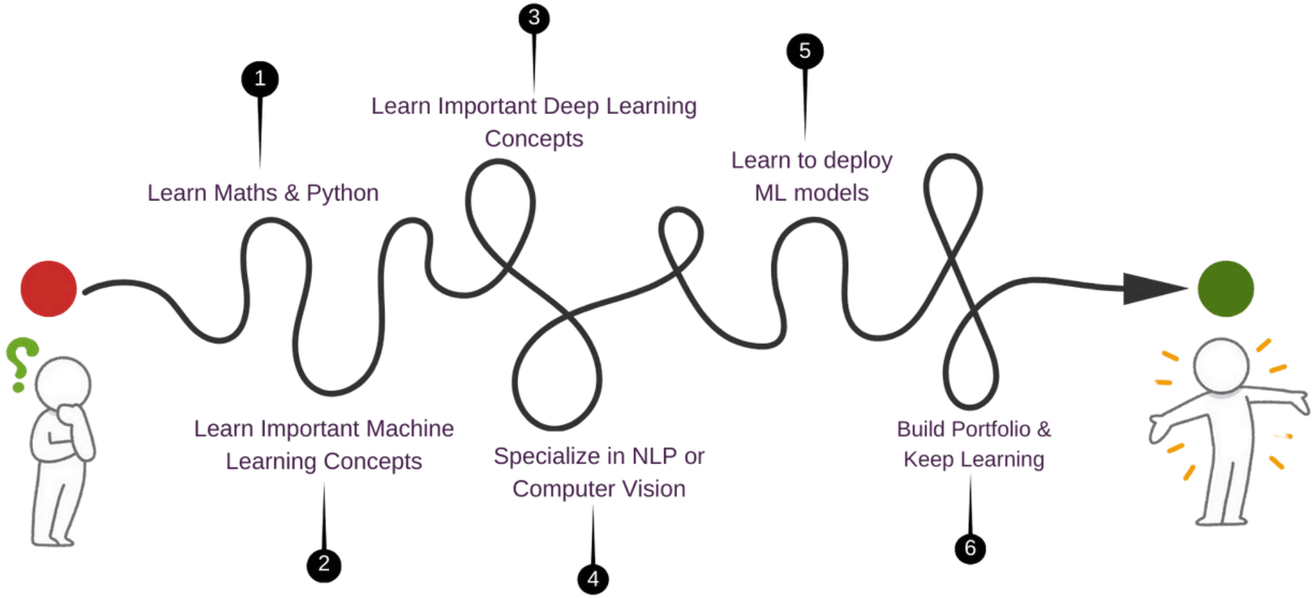
Instance-based vs. Model-based Learning

How does the machine learn and infer?

Instance-Based Learning: The system tries to learn the examples by heart, and then generalize it using some sort of a similarity score. E.g., K-Nearest Neighbor.

Model-Based Learning: In this approach, the solution tries to learn different parameters that can accurately explain the underlying data patterns. For example, a linear regression model tries to fit a straight line that can explain how the data is behaving by minimizing the error.

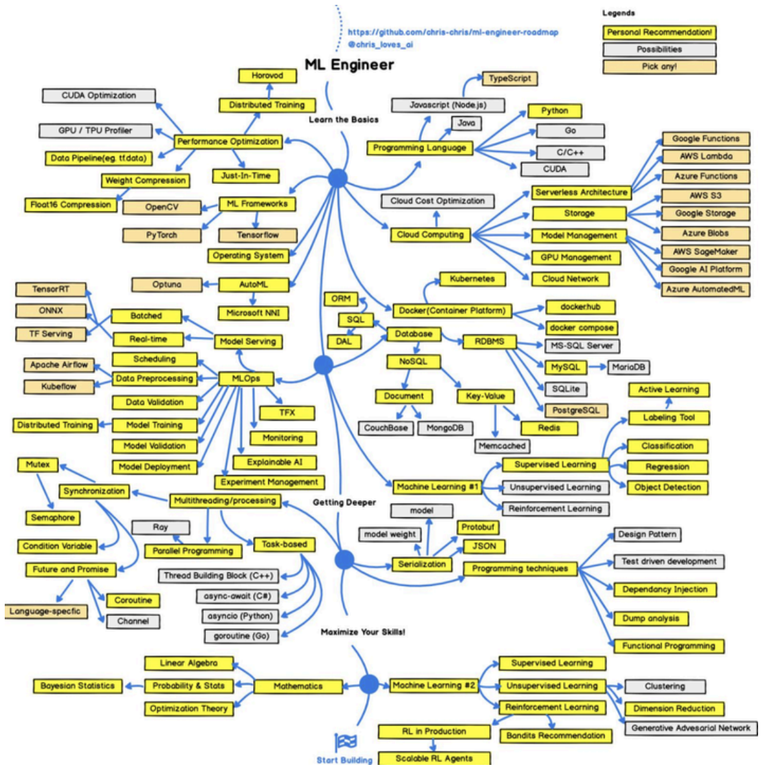
The world of Machine Learning is quite broad, but we have a roadmap to master it …



Brownlee, J., PhD. (2025, January 17). The Roadmap for Mastering Machine Learning in 2025. Machine Learning Mastery. Retrieved February 12, 2025, from https://machinelearningmastery.com/roadmap-mastering-machine-learning-2025/

SJU Saint Joseph's University

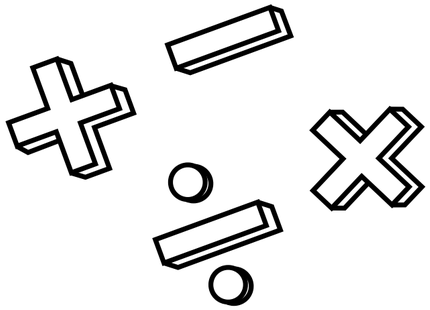
… if only it were that simple …



ML Engineer Roadmap | Kaggle. (n.d.). https://www.kaggle.com/discussions/getting-started/174107

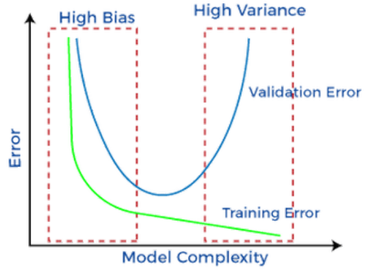
SJU Saint Joseph's University

… but, since we only have 7 weeks, let's see what we can accomplish



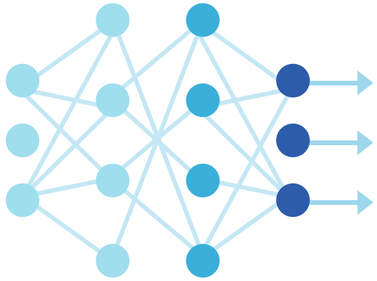
Math Overview

A very light intro to Linear Algebra, Calculus, and Statistics



ML Foundations

Data preparation
Feature Engineering and Feature Selection
Bias-Variance Tradeoff



ML Algorithms

Supervised algorithms
Unsupervised algorithms
Hyperparameter tuning
Cross-validation
Model evaluation

Brief NLP Introduction

Text Pre-processing
Text vectorization
Text classification

SJU Saint Joseph's University

Specifically, we will cover the following topics in weeks 1 - 4

Week	Math	ML Foundations	ML Algorithms	NLP
1	Vector and Matrix Operations Gradient Descent Fundamental Statistics			
2				
3				
4				

SJU Saint Joseph's University

Specifically, we will cover the following topics in weeks 5 - 8

Week	Math	ML Foundations	ML Algorithms	NLP
5				
6				
7				
8				

SJU Saint Joseph's University

In this course, we will apply the following Python packages and related technologies

- Conda/Pip environments for creating and managing Python environments
- Python Base Modules, such as os
- NumPy for numerical operations and data manipulation
- Pandas for loading, analyzing, and pre-processing data
- Matplotlib and Seaborn for visualizing data
- Scikit-Learn for creating ML models
- NLTK for text-processing

If time permits, we will explore the following

- Weights & Biases for model performance tracking
- Optuna for hyper-parameter tuning
- OpenCV for processing images

Anisha

Sam

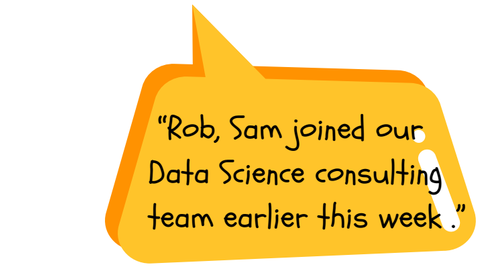
Rob

Our Awesome
Data Scientists

Imagine you joined a new consulting firm earlier this week in the Data Science team. This is your dream job!

Aah! Freshly brewed coffee and my GTX 4080 powered laptop ... Bring it on!!!

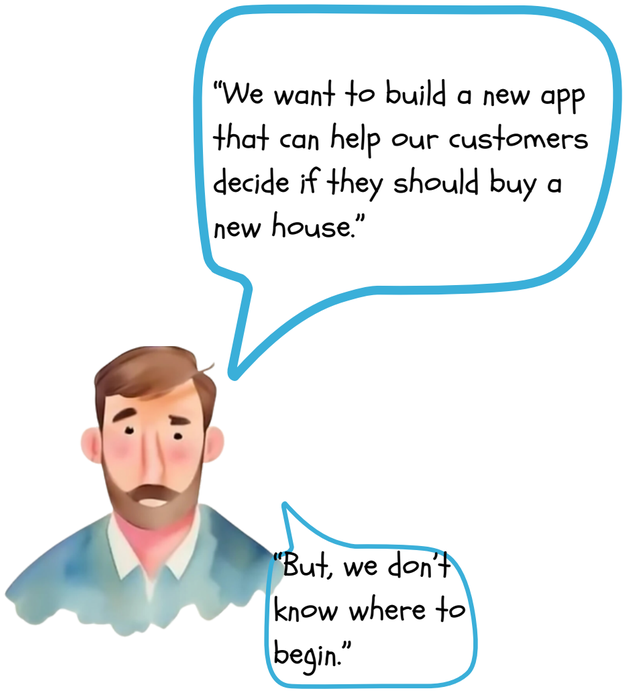

"Sam, meet Rob. He is our new client."


"Rob, Sam joined our Data Science consulting team earlier this week ."


"Nice to meet you, Sam. "


"Nice to meet you as well!"
"I look forward to working with you."


"Rob, how can we help you and your team be successful?"

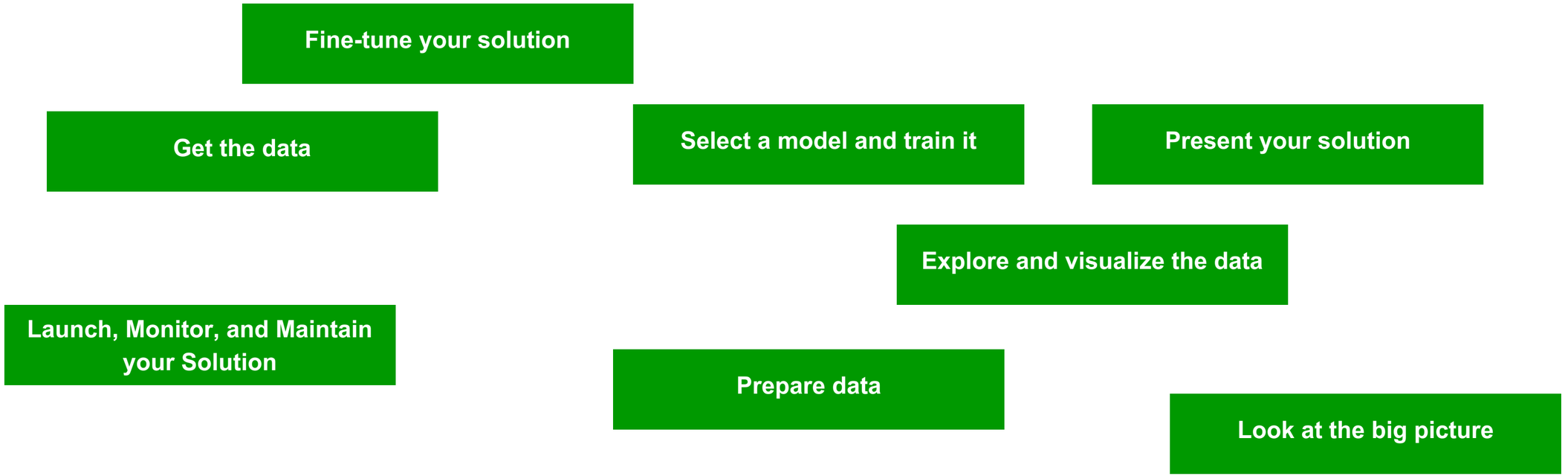

"We want to build a new app that can help our customers decide if they should buy a new house."
"But, we don't know where to begin."


"That's a great idea! We can certainly help."
"Do you have a list of requirements?"


""That's part of the problem. We are going in circles."
"Don't worry. We have an excellent team. They will help you get organized."

SJU Saint Joseph's University

While studying at Saint Joseph's University, you learned about initiating Ml projects.



… but you can't seem to remember the exact order

SJU Saint Joseph's University

Now that you have ordered the topics correctly, identify the steps

#	Topic	Steps
1	Look at the big picture	
2	Get the data	
3	Explore and visualize the data	
4	Prepare the data	
5	Select a model and train it	
6	Fine-tune your model	
7	Present your solution	
8	Launch, Monitor, and Maintain	

SJU Saint Joseph's University

Key Steps: 1 - 4

#	Topic	Steps
1	Look at the big picture	• Define the problem in business terms. How will the solution be used? • Is there an existing solution? How does it differ from what is being proposed? • What kind of Machine Learning solution is this - Supervised, Unsupervised, or Semi-Supervised? • How should the performance be measured? How is the performance aligned to the business objective? • List all assumptions and is there a way to verify those assumptions?
2	Get the data	• List the data you might need. How much? What kind of data is it (e.g., Time-Series, Geographical, Image, Video)? Are there any legal obligations we should be aware of in collecting or using the data? • Where is the data? Do you have access to it? • Do we need to convert the data before it can be used in a model?

3	Explore and visualize the data	• What do you know about the data?
		  ◦ Are there any missing data?
		  ◦ Are there any outliers?
		  ◦ Are the features / variables related somehow? If so, should we include them together?

4	Prepare the data	• How do you clean the data? • What features are important? Do we need to create any new features? • Do we need to scale the data?

SJU Saint Joseph's University

Key Steps: 5 - 8

#	Topic	Steps
5	Select a model and train it	• What models do you want to train? Why? • How would you measure and compare the performance of each of the models?
6	Fine-tune your model	• Is there a way to improve model performance through feature engineering or hyperparameter tuning? • Can you combine different models to form an ensemble?
7	Present your solution	• Document your approach. • Create a nice presentation, highlighting the big picture first. • Explain how your solution helps achieve the overall business objective.
8	Launch, Monitor, and Maintain	• How would you scale and deploy the solution? • How will you measure model performance over time? • What steps will you take if model performance degrades? Why do you think model performance might degrade?

… but wait …