WOMEN WHO CODE®
/data-science

Build a Data Science Project From Scratch - SESSION 4

JUNE 30
10 PM ET

Bisi Josh-Falade
Research Student

womenwhocode.com/datascience/events

Session Agenda

- Machine Learning - An Introduction
- Differences between Supervised, Unsupervised & Reinforcement Learning
- Regression vs Classification
- Machine Learning vs Deep Learning vs Artificial Intelligence
- Importance of Baseline Models
- Session use case & colab work

Machine Learning - An Introduction

- Branch of ***Artificial Intelligence*** Arthur Samuel (IBM, 1959)
  - Building algorithms that can learn from data and improve their performance over time “without being explicitly programmed”
  - Building & training a model on a dataset and using that model to make predictions or decisions on new data
- 3 main types: supervised learning, unsupervised learning, and reinforcement learning
  - Supervised Learning involves training a model on labeled data to make predictions on new data.
  - Unsupervised Learning involves finding patterns and relationships in data without any pre-existing labels.
  - Reinforcement Learning involves training a model to make decisions based on feedback and rewards.

Machine Learning - An Introduction

- Application in various fields, including:
  - Finance,
  - Healthcare,
  - Real Estate,
  - Retail and
  - Transportation
- Rapidly growing in importance as the amount of data generated globally, continues to increase.
- Serves to optimize, increase accuracy and simplify decision-making for seemingly countless problems in a broad range of industries and organizations.

Supervised vs Unsupervised vs Reinforcement Learning

Criteria	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Definition	The machine learns by using labeled data	The machine is trained on unlabeled data without any guidance	An agent interacts with its environment by taking actions and learning from errors & rewards
Type of Problems	Regression & Classification	Association & Clustering	Reward-based
Type of Data	Labeled Data	Unlabeled Data	No predefined Data
Training	External Supervision	No Supervision	No Supervision
Approach	Maps labeled inputs to known outputs	Understands patterns and discovers the outputs	Follows a trial-and-error method

WOMEN WHO CODE®
/data-science

Regression vs Classification

Regression	Classification
Task of predicting continuous quantity	Task of predicting a discrete class label
Aims to predict output value using training data	Aim is to group the output into a class
A Regression problem requires predicting a quantity/value	In a Classification problem, data is labelled into one of two or more classes
If it involves a real number or continuous value, then it is a regression problem	If it involves a discrete or categorical variable, then it is a classification problem
A regression problem with multiple input variables is called a Multivariable Regression Problem	A classification problem involving 2 classes is called a Binary Classification problem while a problem involving more classes is called a Multi-Classification problem
Example: House-pricing prediction	*Example: Email Spam classification*

WOMEN WHO CODE®
/data-science

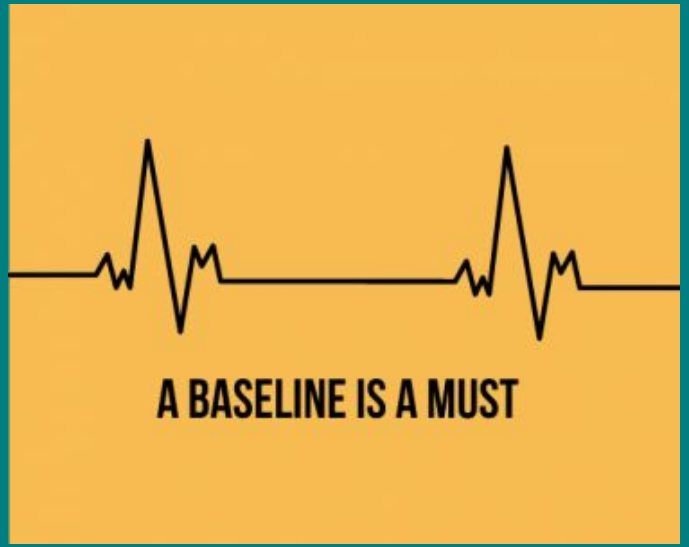
ML vs DL vs AI

Artificial Intelligence (AI)	Machine Learning (ML)	Deep Learning (DL)
Development of machines and systems that can perform tasks that would normally require human intelligence	Machine learning is a subset of AI that involves building algorithms that can learn from data	Subset of machine learning that involves building deep neural, complex networks of interconnected nodes that can learn to recognize patterns in data
Systems are designed to operate autonomously, adapt to changing environments, and make decisions based on incomplete or uncertain information.	Includes Supervised, Unsupervised, Reinforcement Learning	To identify features and patterns in data at different levels of abstraction, allowing them to achieve higher levels of accuracy and performance than traditional machine learning algorithms.
Includes machine learning, deep learning, natural language processing, robotics, and computer vision.	Application areas: robotics, prediction, market segmentation	Application areas: image and speech recognition.

WOMEN WHO CODE®
/data-science

What is a Baseline model?

A Baseline Model serves as a starting point for comparison in Machine Learning tasks. It represents the minimum level of performance that can be achieved without complexity, sophisticated techniques or fine-tuning.



WOMEN WHO CODE®
/data-science

Importance of Baseline Models

Baseline Models are used for determining the consistency and ascertaining the reliability of a trained model. They serve as a benchmark, with which to compare the actual model that is built and iterate quickly if/where necessary. Common baseline models are:

Baseline Regression Models	Baseline Classification Models
Mean or Median	Mode
Linear Regression	Logistic Regression
Conditional Mean/Business Logic	Conditional Mode/Business Logic

Join us on Slack to ask questions and keep the discussion going!

Use the channel:

#build-a-ds-project

Time for hands-on coding!

WOMEN WHO CODE®
/data-science