apply machine learning using Python tools and focus on machine learning modeling techniques, such as classification, regression, and clustering.

Modules:

Module 1: Introduction to Machine Learning

Module 2: Linear and Logistic Regression

Module 3: Building Supervised Learning Models

Module 4: Building Unsupervised Learning Models

Module 5: Evaluating and Validating Machine Learning Models

Module 6: Final Exam and Project

Ai compreses:

Generative ai,ml,dl,nlp,computer vision.

An overview of ml:

Machine learning teaches computers to learn from data, identify patterns, and make decisions without receiving any explicit instructions from a human being.

ML algorithms use computational methods to learn information directly from data without depending on any fixed algorithm.

1.supervised learning models train on labeled data to learn how to make inferences on new data, enabling them to predict otherwise unknown labels

2.Semi-supervised learning trains on a relatively small subset of data that is already labeled, and iteratively retrains itself by adding new labels that it generates with reasonably high confidence.

3.Unsupervised learning works without labels by finding patterns in data.

4.Reinforcement learning simulates an artificially intelligent agent interacting with its environment and learns how to make decisions based on feedback from its environment.

machine learning technique depends on several factors:

1.classification technique is used to predict the class or category of a case.

2.The regression technique is used to predict continuous values.

3.Clustering groups of similar case or makes a group of item into a same cluster.

4.The association technique is used to find items or events that often co-occur.

5.Anomaly detection is used to discover abnormal and unusual cases. for credit card fraud detection.

6.Sequence mining is used to predict the next event.

7.Classification reduction is used to reduce data size, particularly the number of features needed.

8.recommends new items to them, such as books or movies.

two methods are used in supervisied data in machine learning:

1.classification

2.regression

1.Classification categorizes input data into one of several predefined categories or classes.

2.Regression predicts continuous numerical values from input data.

Clustering is one of many unsupervised machine learning techniques for data modeling. It's used to group data points or objects that are somehow similar.

machine learning is applied to predict diseases, analyze consumer behavior, and recognize images

Process of lifecycle of a ml:

the model lifecycle is iterative, which means that we tend to go back and forth between these processes.[we can go front or back phase in any point]

1.prblm def

2.data collection

3.data preparation

4.model development & evaluation

5.model deployment

2.Data collection and preparation is called the Extract, Transform, and Load, or ETL process.

The ETL process involves collecting data from various sources, then cleaning, transforming, and storing it into a new single place.

taking an ex as a beauty product usage- how life cycle works

Data collection:

user data: demographics,history etc

product data: inventory,items,popularity.

other data: saved,liked,searched products

some major transforming by wrangling, aggregating, joining, merging, and mapping the data onto one central source. This reduces the need to deal with multiple databases every time we need to pull data.

3. data from multiple sources will contain errors, different formatting, and missing data. This process overlaps with the Data Collection process as they can be done in tandem.



4a. leveraging the  technique called content-based filtering.

This technique finds the similarity between products, based on product content.

technique called **Collaborative Filtering** that uses the user’s data. creating similarities between two users based on how they view a product.

4b. Model Evaluation step;

the initial stages of Model Evaluation will involve me tuning the model and

doing some testing on the data set I had kept earlier for testing.

5.model development: performance to make sure it continues to do the job that the business requires. Future iterations may include retraining the model based on new information

in order to expand its capabilities.

**Data Scientist vs AI Engineer**