Machine Learning & Gen AI

Live Classes	Guided Learning

Week 1: Python Advanced & OOPS
• Packages & Modules,
• Iterators
• Generators
• Classes & Objects
• Exceptions
 	
Week 2: AI and Machine Learning Overview
• What is ML ?
• Supervised vs Unsupervised
• Regression vs classification
• Anomaly Detection
• Neural Networks
• NLP
• Various Business scenarios where ML is being used
 	
Week 3 4: Data Processing & Visualization & Feature Engineering
• Data Libraries: Numpy, Pandas
• Visualization libraries: Matplotlib, seaborn
• Data Sourcing
• Exploratory Data Analysis
• Feature engineering
• Feature reduction
• Feature bucketing
	Documents, Videos and Assignments will be provided for below topics. Adequate time will be given to the below topics according to your semester calendar
	
	Python Fundamentals - Self Learning
	• Data Structures: List, Tuples, Sets, Dictionary
	• Functions & argument variations, decorators
	
	SQL Advanced - Self Learning
	• Schema
	• Joins
	• GroupBy
	• Index
	• Window Functions
	• Subquery
	• Case when
	• Connecting & Querying from Python
	• Use SQL in python on pandas:
	  ○ Pandasql
	  ○ duckdb
	Math Foundations for Data Science - Self Learning
	• Calculus
	• Linear Algebra

- Feature Importance

Week 4 to 7: Machine Learning Algorithms

- Supervised ML Models:
  - Linear regression
  - Logistic Regression
  - SVM
  - Naive Bayes
  - KNN
  - Decision Tree
  - Bagging -> Random Forest
  - Boosting -> AdaBoost, GBM, XGBoost, etc.
  - *<Mini Project> Titanic Survival / Breast Cancer*
- Unsupervised ML Models:
  - Isolation Forest (Anomaly Detection)
  - Principal Component Analysis
  - Clustering (K-Means)
  - Capacity, Overfitting, underfitting
- Model Training:
  - Train Test Split
  - Scoring
- Hyper parameters tuning:
  - RandomSearch
  - GridSearch
  - Hyperopt
- Loss Functions:
  - Logg loss
- Probability
- Statistics

Advanced DSA using Python

- Stack
- Queue
- Linked List

- Cross Entropy
- MAE, MSE, etc.

- Regularization:
  - L1 vs l2
- Overfitting vs Underfitting
- Model Evaluation metrics:
  - Accuracy
  - AUC
  - Precision
  - Recall
  - F1 Score
  - Lift, etc.
- Model Deployment and scoring:
  - Batch vs API

Week 8 to 10: Deep Learning:

- What is a Neural Network
- Forward and backward propagation
- Activation Functions
- Gradient Descent
  - Vanishing and exploding
- Weights & Bias
- Regularization
  - Dropout
- Batch vs Mini Batch vs Stochastic
  - Batch Normalisation
- Optimizers
  - GD with Momentum
  - RMSProp
  - Adam
- Autoencoders
  - Anomaly Detection
- PyTorch

- *<Mini Project>* Deep Learning model building and training using PyTorch for a classification Problem

Week 11 to 13: NLP

- What is NLP ?
- Various methods to transform textual data into numerical data:
  - Bag of Words, TF-IDF, etc.
  - Word2vec, doc2vec, etc.
- Data Preprocessing in NLP
- RNN, LSTM, GRU
- *<Mini Project>* A sentiment classification model in PyTorch

Week 14 to 16: Generative AI

- Attention & Transformers.
- BERT
  - *<Mini Project>*
- Generative Models
  - How does GPT-2 work?
  - How do decoder models predict the sentences? Different decoding methods
    - Greedy search:
    - Breedy search:
    - Sampling
    - Tok-k sampling
    - Top-p sampling.
- Sentence Transformer
  - What is semantic search ?
  - What is vector db?

○ KNN(Brute Force) vs Approx Nearest Neighbor(ANN) concept in Information retrieval (IR) for a RAG system? ■ IVF ■ LSH ■ HNSW • Prompt Engineering: ○ Best practices when writing a prompt ○ Zero-shot vs few-shot ○ What is chain of thoughts ? • Langchain, RAG ○ What is memory buffer in langchain for chat models ? ○ Models and different tasks supported by langchain ○ *<Mini Project> Build a RAG model for Q/A* • Fine Tuning of LLMs Week 17 to 19: End-to-End Project	