

Table of Contents: Python, Machine Learning, Deep Learning, Computer Vision, NLP

I. Python Programming for Data Science

- 1. Python Basics (Syntax, Variables, Data Types)
- 2. Control Structures (If, Loops)
- 3. Functions and Lambda Expressions
- 4. Data Structures (Lists, Tuples, Sets, Dictionaries)
- 5. File Handling and Exceptions
- 6. Object-Oriented Programming (OOP)
- 7. Modules and Packages
- 8. Working with Libraries (NumPy, Pandas)
- 9. Data Manipulation and Aggregation
- 10. Data Visualization with Matplotlib & Seaborn

II. Data Science Fundamentals

- 1. Introduction to Data Science
- 2. Data Science Lifecycle
- 3. Data Collection Techniques
- 4. Exploratory Data Analysis (EDA)
- 5. Data Cleaning and Preprocessing
- 6. Feature Engineering and Selection
- 7. Data Visualization Tools and Techniques
- 8. Model Evaluation Metrics

III. Machine Learning (ML)

A. Supervised Learning

- Regression Algorithms: Linear Regression, Polynomial Regression, Ridge / Lasso / Elastic Net, Decision Tree Regression, Random Forest Regression, SVR, XGBoost, LightGBM, CatBoost
- Classification Algorithms: Logistic Regression, K-Nearest Neighbors (KNN), Naive Bayes (Gaussian, Multinomial, Bernoulli), Decision Trees / Random Forest, SVM, SGDClassifier, Gradient Boosting, XGBoost, LightGBM, CatBoost, LDA / QDA / MLP

B. Unsupervised Learning

- Clustering Algorithms: K-Means, DBSCAN, Hierarchical, Mean Shift, GMM
- Dimensionality Reduction: PCA, t-SNE, UMAP, ICA, Autoencoders

C. Semi-Supervised Learning

- Label Propagation / Spreading

- Self-Training / Co-Training
- Semi-Supervised SVM

D. Reinforcement Learning

- Basics and Terminology
- Value-Based Algorithms: Q-Learning, DQN, Double DQN, Dueling DQN
- Policy-Based Algorithms: REINFORCE, PPO
- Actor-Critic Algorithms: A2C, A3C, DDPG, TD3, SAC

E. Ensemble Methods

- Bagging - Random Forest, Bagged Trees
- Boosting - AdaBoost, XGBoost, LightGBM, CatBoost
- Stacking / Blending

F. Anomaly Detection

- One-Class SVM, Isolation Forest
- Local Outlier Factor (LOF), Elliptic Envelope
- Autoencoders for Anomaly Detection

IV. Deep Learning (DL)

- 1. Neural Networks (Perceptron, MLP)
- 2. Activation Functions (ReLU, Sigmoid, Tanh, etc.)
- 3. Optimization Algorithms (SGD, Adam, RMSprop)
- 4. CNNs – Convolutional Neural Networks
- 5. RNNs – Recurrent Neural Networks
- 6. LSTMs / GRUs
- 7. GANs – Generative Adversarial Networks
- 8. Autoencoders (Denoising, Sparse, Variational)
- 9. Transformers & Attention Mechanisms
- 10. Transfer Learning
- 11. Model Deployment Techniques

V. Computer Vision (CV)

- 1. Image Classification
- 2. Object Detection (YOLO, SSD, Faster R-CNN)
- 3. Image Segmentation (U-Net, Mask R-CNN)
- 4. Facial Recognition & Landmark Detection
- 5. Optical Character Recognition (OCR)
- 6. Pose Estimation
- 7. Image Super-Resolution / Enhancement
- 8. Image Augmentation Techniques
- 9. Depth Estimation and 3D Reconstruction

VI. Natural Language Processing (NLP)

- 1. Text Preprocessing (Tokenization, Stopwords, Stemming)
- 2. Vectorization Techniques (TF-IDF, Word2Vec, GloVe, FastText)
- 3. Text Classification
- 4. Named Entity Recognition (NER)
- 5. Sentiment Analysis
- 6. Machine Translation
- 7. Text Summarization (Extractive/Abstractive)
- 8. Question Answering Systems
- 9. Chatbots & Conversational AI
- 10. Transformers in NLP (BERT, GPT, T5, RoBERTa)
- 11. Large Language Models (LLMs) & Fine-Tuning

VII. Tools and Frameworks

- 1. Python Libraries (Pandas, NumPy, Scikit-learn, etc.)
- 2. Deep Learning Frameworks (TensorFlow, PyTorch, Keras)
- 3. NLP Libraries (NLTK, SpaCy, HuggingFace Transformers)
- 4. CV Libraries (OpenCV, PIL, Detectron2, etc.)
- 5. Data Labeling & Annotation Tools
- 6. Model Tracking (MLflow, Weights & Biases)
- 7. Cloud Platforms (AWS, GCP, Azure)

VIII. Projects and Use Cases

- 1. Stock Market Prediction
- 2. Fraud Detection
- 3. Autonomous Driving
- 4. Face Recognition System
- 5. Product Recommendation Engines
- 6. Sentiment Analysis of Tweets
- 7. Medical Diagnosis via Imaging
- 8. Customer Churn Prediction
- 9. AI Chatbots for Customer Support

IX. Model Deployment & MLOps

- 1. REST APIs with Flask / FastAPI
- 2. Dockerizing ML Applications
- 3. CI/CD for ML
- 4. Model Versioning
- 5. Monitoring and Logging
- 6. A/B Testing for Models
- 7. Edge AI & On-device Inference

Data Science & AI Projects with Objective and Dataset

Data Science Fundamentals

1. 1. Customer Segmentation

- Objective: Group customers based on purchasing behavior.
- Dataset: Mall Customers Dataset - Kaggle

2. 2. Sales Forecasting

- Objective: Predict future sales using historical data.
- Dataset: Retail Sales Forecasting Dataset - Kaggle

Machine Learning

3. 1. Loan Default Prediction

- Objective: Predict whether a borrower will default on a loan.
- Dataset: Lending Club Loan Data - Kaggle

4. 2. Diabetes Prediction

- Objective: Classify if a patient has diabetes based on health indicators.
- Dataset: PIMA Indian Diabetes Dataset - UCI

Deep Learning

5. 1. Handwritten Digit Recognition

- Objective: Classify handwritten digits from images.
- Dataset: MNIST Dataset - Yann LeCun

6. 2. Image Captioning

- Objective: Generate captions for input images.
- Dataset: Flickr8k Dataset - Kaggle

Computer Vision

7. 1. Face Mask Detection

- Objective: Detect face masks in real-time video streams.
- Dataset: Face Mask Dataset - Kaggle

8. 2. Vehicle Detection

- Objective: Detect and classify vehicles on roads.
- Dataset: Open Images Dataset V6 - Google

Natural Language Processing (NLP)

9. 1. Spam Email Classifier

- Objective: Classify emails as spam or not.
- Dataset: SpamAssassin Public Corpus

10. 2. Named Entity Recognition

- Objective: Identify entities like names, dates in text.
- Dataset: CoNLL-2003 NER Dataset

MLOps & Deployment

11. 1. Model Serving with FastAPI

- Objective: Deploy ML model as API using FastAPI.
- Dataset: Custom trained model

12. 2. Dockerize ML Workflow

- Objective: Containerize ML training & inference pipeline.
- Dataset: Any public ML dataset