Data Science & Machine Learning Handbook

1. Lecture Flow

Introduction to Data Science & ML
- What is Data Science?
- Difference between Data Science, AI, and ML
- Real-world applications

Data Preprocessing & Exploration
- Data Cleaning (Handling missing values, outliers)
- Feature Scaling & Normalization
- Data Visualization (Matplotlib, Seaborn, Plotly)

Machine Learning Basics
- Supervised vs Unsupervised Learning
- Regression vs Classification
- Bias-Variance Tradeoff

Model Building & Evaluation
- Train-Test Split, Cross-Validation
- Performance Metrics (Accuracy, Precision, Recall, F1-score, AUC-ROC)
- Overfitting vs Underfitting

2. Essential Handbook

Key Concepts
- Mean, Median, Mode (Measures of Central Tendency)
- Standard Deviation, Variance (Measure of Dispersion)
- Confusion Matrix (Classification Performance)
- Gradient Descent (Optimization Algorithm)
- Overfitting & Regularization (L1, L2 Regularization)

Libraries & Tools
- Pandas - Data Manipulation
- NumPy - Numerical Computing

- Scikit-learn - ML Algorithms
- TensorFlow / PyTorch - Deep Learning

3. Interview Questions

Conceptual Questions

1. Explain the difference between correlation and causation.
2. What is feature engineering? Why is it important?
3. What is the Curse of Dimensionality?
4. Explain the difference between bagging and boosting.
5. What is a confusion matrix, and how do you interpret it?

Coding Questions

1. Write Python code to implement Linear Regression.
2. How do you handle missing values in a dataset?
3. Implement k-Nearest Neighbors (KNN) in Python.

4. Assignments

Hands-on Practice

1. Data Preprocessing: Load a dataset, clean missing values, and visualize it.
2. Regression Model: Build a Linear Regression model and evaluate performance.
3. Classification Task: Train a Logistic Regression model on a dataset.
4. Feature Selection: Use different techniques (PCA, Chi-square test) to select the best features.
5. Hyperparameter Tuning: Implement Grid Search & Random Search for ML models.

End of Handbook