AI & Data Scientist – 100+ Lab Exercises (Basic, Intermediate, Advanced)

☐ Basic Level (30+ Exercises)

Goal: Build a strong base in programming, statistics, and data handling.

Python Programming Essentials

- Variables, loops, and conditional statements in Python.
- Functions, classes, and object-oriented principles.
- List/dictionary/set/tuple manipulations.
- Reading/writing files and data parsing.
- Exception handling and logging in Python.

Mathematics for Data Science

- Linear algebra: vectors, matrices, dot/cross products.
- Probability: distributions, Bayes theorem.
- Statistics: mean, median, standard deviation, IQR.
- Derivatives and gradients with visual plots.
- Matrix decompositions: SVD, eigenvalues/eigenvectors.

Data Handling & Cleaning

- Use Pandas for data preprocessing.
- Handle nulls, outliers, duplicates.
- Convert categorical features to numeric.
- Merge, group, pivot, and filter dataframes.
- Exploratory Data Analysis (EDA) on CSV/Excel data.

Visualization

- Histograms, boxplots, scatterplots (Matplotlib/Seaborn).
- Correlation matrix and pairplots.
- Trend and line plots for time-series data.
- Interactive charts using Plotly.

- Geospatial mapping using Folium.

☐ Intermediate Level (40+ Exercises)

Goal: Master core ML/AI concepts with practical applications.

Machine Learning Models (Supervised)

- Implement Linear & Logistic Regression.
- Train Decision Trees, Random Forests, XGBoost.
- Model evaluation: confusion matrix, precision, recall, ROC.
- Hyperparameter tuning using GridSearchCV.
- Use scikit-learn pipelines for model building.

Unsupervised Learning

- K-Means and DBSCAN clustering on real datasets.
- PCA for dimensionality reduction and visualization.
- Anomaly detection with Isolation Forest.
- Market Basket Analysis with Apriori.
- Hierarchical clustering and dendrogram analysis.

Natural Language Processing

- Text cleaning, stemming, lemmatization.
- Convert text to vectors using BoW, TF-IDF.
- Build sentiment classifiers using Naive Bayes.
- Topic modeling using LDA.
- Named Entity Recognition using spaCy.

Deep Learning Fundamentals

- Build simple neural networks using Keras/TensorFlow.
- Activation functions: ReLU, sigmoid, softmax.
- Train an image classifier on MNIST.
- Use dropout and batch normalization.
- Plot model accuracy/loss graphs.

☐ Advanced Level (40+ Exercises)

Goal: Engineer intelligent solutions, deploy models, and leverage cutting-edge techniques.

Advanced AI Models

- Implement CNNs for image recognition (e.g., CIFAR-10).
- RNN and LSTM models for time-series and text data.
- Build attention-based transformers from scratch.
- Use BERT for text classification tasks.
- Apply transfer learning using pre-trained models.

Generative AI & Foundation Models

- Build a simple GAN to generate synthetic data.
- Generate text using GPT and fine-tuned transformers.
- Create image generation pipelines using Stable Diffusion APIs.
- Build a multimodal AI model (text+image).
- Prompt engineering for ChatGPT and Claude models.

Time-Series Analysis

- Use ARIMA and SARIMA for forecasting.
- Implement Prophet by Facebook for time-series trends.
- Feature engineering for temporal features.
- Anomaly detection in financial time-series.
- Build dashboards with seasonal trends.

Big Data & Distributed AI

- Data manipulation using PySpark.
- Model training using MLlib on large datasets.
- Parallelized model evaluation in Spark.
- Build ML pipelines in Azure/AWS/GCP.
- Ingest data from Kafka, store in Hadoop, analyze in Spark.

AI Deployment & MLOps

- Build REST APIs for AI models using FastAPI.
- Containerize models using Docker.
- Implement model monitoring with Prometheus & Grafana.

- Automate pipelines using CI/CD tools (GitHub Actions).
- Setup MLFlow for experiment tracking.

Capstone Projects

- AI-powered recommendation engine (e-commerce).
- Fraud detection engine for financial transactions.
- Medical diagnosis prediction system.
- Generative chatbot using fine-tuned LLMs.
- AI for predictive maintenance in manufacturing.

☐ Optional Tools & Platforms

- Python, Jupyter, Colab, TensorFlow, PyTorch, Scikit-learn
- Power BI, Tableau, Excel, BigQuery, Spark
- HuggingFace Transformers, OpenAI APIs, MLFlow
- Docker, Kubernetes, GitHub Actions, DVC