DATA SCIENCE WITH AI

DETAILED SYLLABUS

- 3.5 Months of Practical & Module Training
- 1.5 Months Advance AI & ML Training
- 0.5 Month Deep Learning using Keras and TensorFlow Training
- 0.5 Month Interview & Resume Building Preparation
- 26 Weeks: 230+ Hours

Note: 6 Months Internship (Internship will start from the first month of the program)

Phase 1 – Practical & Module Training (14 Weeks = 42 Classes)
(3.5 months)

Weeks 1–2 (6 classes): Python (Core + Advanced), Data Structures

- **Class 1**: Orientation, Python installation, IDE setup, Syntax, Variables, Data Types
- **Class 2**: Operators, Typecasting, Input/Output, String operations
- **Class 3**: Lists, Tuples, Sets – indexing, slicing, operations
- **Class 4**: Dictionaries – keys, values, methods, nested dicts
- **Class 5**: Control Structures – if, elif, else, nested conditions
- **Class 6**: Loops – for, while, nested loops, Comprehensions

Project 1: Student Management System (CLI App)
- Build a small console-based app using Python.
- Features: Add students, store marks in a dictionary, calculate average, and print reports.
- Skills applied: Strings, Lists, Dicts, Loops, Functions.

Weeks 3–4 (6 classes): Statistics & Probability

- **Class 7**: Intro to Statistics – population vs sample, data types
- **Class 8**: Descriptive Statistics – Mean, Median, Mode, Variance, SD
- **Class 9**: Probability Basics – rules, conditional probability
- **Class 10**: Probability Distributions – Normal, Binomial, Poisson
- **Class 11**: Hypothesis Testing, p-values, Confidence Intervals
- **Class 12**: Correlation vs Regression, Sampling Techniques

Project 2: *Dice & Card Probability Simulator*
- Simulate dice rolls and card draws.
- Calculate probabilities of outcomes (e.g., rolling doubles, getting an Ace).
- Validate theoretical probability vs simulated probability.

Weeks 5–6 (6 classes): NumPy & Pandas
- **Class 13**: NumPy – Arrays, vectorization, indexing, slicing
- **Class 14**: NumPy – Broadcasting, Functions, Random numbers
- **Class 15**: Pandas – Series, DataFrame basics, indexing
- **Class 16**: Pandas – Cleaning (missing values, duplicates, outliers)
- **Class 17**: Pandas – GroupBy, Merge, Join, Concat
- **Class 18**: Pandas – Advanced Operations + Case Study

Project 3: *Retail Sales Data Analysis*
- Dataset: Sales transactions (CSV).
- Clean data (missing values, duplicates).
- Compute total sales by category, region.
- Find top-selling products.

Weeks 7–8 (6 classes): Data Visualization
- **Class 19**: Matplotlib – Basic plots (line, bar, scatter, pie)
- **Class 20**: Matplotlib – Styling, subplots, annotations
- **Class 21**: Seaborn – Distplot, Countplot, Boxplot
- **Class 22**: Seaborn – Heatmaps, Pairplot, Violin plots
- **Class 23**: Plotly – Interactive visualizations
- **Class 24**: Visualization Project – Exploratory Dashboard

Project 4: *COVID-19 Data Dashboard*
- Dataset: COVID cases by country.
- Use Matplotlib/Seaborn to plot daily cases, deaths, recovery trends.
- Use Plotly for interactive country-wise dashboard.

Weeks 9–10 (6 classes): SQL & Database Management
- **Class 25**: SQL Basics – DDL, DML, Joins
- **Class 26**: SQL Queries – Filtering, Aggregations, GroupBy
- **Class 27**: Advanced SQL – Subqueries, Views, Indexes
- **Class 28**: Case Study – Business Queries using SQL
- **Class 29**: Connecting SQL with Python (SQLite/MySQL)
- **Class 30**: Mini Project – Database + Python Integration

Project 5: *Library Management System with SQL*
- Create DB with Books, Users, Issue Logs.
- Run queries: most borrowed books, active users.
- Connect with Python to fetch results.

Weeks 11–12 (6 classes): Exploratory Data Analysis (EDA)
- **Class 31**: Understanding Data, Data Cleaning Workflow
- **Class 32**: Outlier Detection, Handling Missing Values
- **Class 33**: Feature Engineering – Encoding, Scaling
- **Class 34**: Univariate & Bivariate EDA
- **Class 35**: Multivariate Analysis, Correlation Heatmaps
- **Class 36**: EDA Project – Business Dataset

Project 6: Titanic Survival Analysis

- Dataset: Titanic (Kaggle).
- Clean missing age, cabin data.
- Analyze survival rate by gender, class, age group.
- Visualize using heatmaps & barplots.

Weeks 13–14 (6 classes): Case Studies + Integration

- **Class 37**: Case Study – Python + Pandas + SQL
- **Class 38**: Case Study – EDA + Visualization
- **Class 39**: Case Study – Data Pipeline Creation
- **Class 40**: Case Study – Analytics with SQL + Python
- **Class 41**: Mini Project – End-to-End Data Handling
- **Class 42**: Review & Assessment (Phase 1 Wrap-up)

Project 7: Customer Churn Prediction (Pre-ML)

- Dataset: Telecom churn dataset.
- Perform EDA, cleaning, feature encoding.
- Find patterns in churners vs non-churners.
- Deliver insights (without ML yet).

Phase 2 – Advance AI & ML Training (7 Weeks = 21 Classes)

(1.5 months)

Weeks 15–16 (6 classes): Supervised Learning

- **Class 43**: Intro to ML – Types, Challenges, Workflow
- **Class 44**: Regression – Simple & Multiple Linear Regression
- **Class 45**: Classification – Logistic Regression
- **Class 46**: Evaluation Metrics – Accuracy, Precision, Recall, F1
- **Class 47**: ROC, AUC, Confusion Matrix, PR Tradeoff
- **Class 48**: Mini Project – Regression + Classification

Project 8: *House Price Prediction (Regression)*

- Dataset: Housing dataset (Kaggle/Sklearn).
- Build Linear Regression model.
- Evaluate with RMSE, MAE, R^2.

Weeks 17–18 (6 classes): Unsupervised Learning

- **Class 49**: Clustering – K-Means, Elbow Method
- **Class 50**: Hierarchical & DBSCAN Clustering
- **Class 51**: PCA – Dimensionality Reduction
- **Class 52**: Feature Engineering – Feature Selection/Importance
- **Class 53**: Anomaly Detection using Gaussian Mixtures
- **Class 54**: Mini Project – Customer Segmentation

Project 9: *Customer Segmentation (Clustering)*
- Dataset: Mall Customers.
- Apply K-Means & PCA.
- Visualize clusters of customers based on income & spending.

Week 19 (3 classes): Ensemble Models
- **Class 55**: Bagging – Voting Classifier, Random Forest
- **Class 56**: Boosting – AdaBoost, XGBoost, Gradient Boosting
- **Class 57**: Hyperparameter Tuning – GridSearchCV, RandomizedSearchCV

Project 10: *Loan Default Prediction*
- Dataset: Loan data.
- Train Random Forest & XGBoost.
- Compare accuracy, precision-recall.

Week 20 (3 classes): NLP
- **Class 58**: NLP Basics – Tokenization, Lemmatization, Stopwords
- **Class 59**: Text Classification + Sentiment Analysis with Sklearn/Keras
- **Class 60**: Mini Project – Sentiment Analysis on Tweets

Project 11: *Twitter Sentiment Analysis*
- Dataset: Tweets (positive/negative).
- Clean text (tokenization, stopwords).
- Build Logistic Regression or simple NN model.

Week 21 (3 classes): Computer Vision
- **Class 61**: Intro to Images, OpenCV Basics
- **Class 62**: Hands-on CNN – Image Classification (MNIST/CIFAR)
- **Class 63**: Mini Project – Object Detection/Segmentation

Project 12: *Handwritten Digit Recognition (MNIST)*
- Train a CNN on MNIST dataset.
- Achieve >95% accuracy.
- Visualize predictions.

Phase 3 – Deep Learning with Keras & TensorFlow (2 Weeks = 6 Classes)
(2 weeks)
Week 22 (3 classes): Neural Networks
- **Class 64**: ANN Basics – Neurons, Perceptron, MLPs
- **Class 65**: Activation Functions, Optimizers, Loss Functions
- **Class 66**: Build & Train ANN with Keras Sequential API

Project 13: *Image Classification with Transfer Learning*
- Dataset: Cats vs Dogs (Kaggle).
- Use pretrained ResNet/VGG with Keras.
- Fine-tune and deploy model.

Week 23 (3 classes): CNN + RNN + LSTM
- **Class 67**: CNN Architectures (LeNet, AlexNet, VGG, ResNet)
- **Class 68**: RNN, LSTM, GRU – Sequence Modeling
- **Class 69**: Transfer Learning + Project – Image/Text Model

Phase 4 – Major project + Career Prep (3 Weeks = 9 Classes)
(3 weeks)

Week 24 (3 classes): Major Project – Problem Definition & Data Collection
- **Class 70**: Project Introduction – Problem Scoping
- **Class 71**: Data Collection & Cleaning
- **Class 72**: Exploratory Analysis for project Dataset

Week 25 (3 classes): Major project – Model Building
- **Class 73**: Model Selection & Training
- **Class 74**: Evaluation & Optimization
- **Class 75**: Deployment (Streamlit/Flask basics)

Week 26 (3 classes): Career Prep
- **Class 76**: Resume Building Workshop (AI-focused resumes)
- **Class 77**: Mock Technical Interviews
- **Class 78**: Mock HR Interviews