

Data Science Lectureflow

Module 1) Introduction to Data Science	2
<ul style="list-style-type: none"> SESSION 1 — What is Data Science + Foundations + Importance & Applications What is Data Science? Foundations: Statistics Programming Domain Knowledge Importance of Data Science Data-Driven Decision Making Innovation Competitive Advantage Applications in Industries Healthcare Predictive analytics Personalized medicine Medical imaging Session 2 - Data Science Lifecycle Problem Definition Data Collection (DBs, APIs, scraping) Data Preparation Data Exploration Modeling (intro) Evaluation (intro) What is Generative AI? Definition Examples: ChatGPT, DALL·E, Stable Diffusion How GenAI relates to DS Enhancing Data Quality Synthetic data Minority class oversampling Accelerating Model Development Feature engineering automation Model selection assistance 	
Module 2) PYTHON FOR Data Science	30
<ul style="list-style-type: none"> SESSION 1 — Introduction to Python & Installation Topics: What is Python? Why it's essential for analytics Installing Python, Anaconda, Jupyter Notebook Understanding .py vs .ipynb Writing your first Python cell Demo: Print "Hello Analyst" in Jupyter Notebook SESSION 2 — Python Syntax, Keywords, Variables & Comments Topics: Indentation rules Naming conventions Comments (# and multiline) Variables: dynamic typing Demo: Create variables for sales, region, discount SESSION 3 — Data Types in Python Topics: int, float, str, bool type() function Type casting Demo: Convert numeric strings to floats for analysis SESSION 4 — Strings Topics: String methods: upper, lower, replace, split Indexing & slicing Formatting strings Demo: Clean product names like "samsung-Mobile" SESSION 5 — Lists & Tuples Topics: List operations: append, extend, pop Tuple immutability When to use lists vs tuples Demo: Create a list of customer IDs, update and iterate SESSION 6 — Dictionaries & Sets Topics: Key-value pairs Adding, updating, deleting Set operations (union, intersection) Demo: Create product ? price dictionary SESSION 7 — Conditional Statements Topics: if, elif, else Nested conditions Demo: If revenue > target ? print "Achieved" SESSION 8 — Loops Topics: for loop while loop break / continue Enumerate Demo: Loop through sales list ? calculate total SESSION 9 — Functions Topics: def Arguments & return Default arguments Demo: Define function: calculate_discount(price, rate) SESSION 12 — Functions (Part 2: Lambda, Map, Filter, Reduce) Topics: Lambda functions Higher-order functions map() filter() reduce() (from functools) Demo: Use map() to clean product names Use reduce() for total sales SESSION 13 — Recursion + Scope Topics: Recursive functions Function call stack Local, global scope Lifetime of variables Demo: Recursive factorial OR recursive sum 	

- SESSION 14 — File Handling (Text, CSV, JSON) Topics: open(), modes Reading and writing text files CSV reading JSON parsing with json module pathlib basics Demo: Load JSON file ? extract key values
- SESSION 15 — Error & Exception Handling Topics: try except else finally raise custom exceptions Demo: Build safe division program that handles ZeroDivisionError
- SESSION 16 — Iterators & Generators Topics: iter() next() enumerate() zip() generator functions with yield Demo: Create generator ? return next sales value each call
- SESSION 17 — Modules & Packages Topics: import basics Built-in modules: math, os, datetime, json, statistics Creating custom module Virtual environment (venv) Installing packages via pip Demo: Create custom utils.py module
- SESSION 18 — Regular Expressions (Regex) Topics: re.findall() re.search() re.sub() Patterns: email, phone, number, date Demo: Extract all valid emails from text file
- SESSION 19 — OOP Basics (Classes, Objects, init) Topics: class definition object creation attributes methods constructor __init__() self Demo: Class: Customer(name, age, purchase_history)
- SESSION 20 — Encapsulation & Abstraction Topics: Private attributes Getters/setters Abstraction basics Demo: Create class with private _salary
- SESSION 21 — Inheritance & Polymorphism Topics: Single inheritance Method overriding Polymorphism Real-world example Demo: Class: Employee ? Manager ? override bonus method
- SESSION 19 — APIs: Introduction ? What is an API? ? JSON structure ? REST: GET, POST, PUT, DELETE ? Headers, params, rate limits
- SESSION 20 — APIs Using Requests (Part 1) ? requests.get(), requests.post() ? .status_code, .json() ? Query parameters (params={})
- SESSION 21 — APIs Using Requests (Part 2) ? POST with JSON/data ? Parsing responses ? Saving API data to CSV/JSON
- SESSION 22 — Advanced APIs ? requests.Session() ? Basic async concepts ? Auth: API key, token, Bearer ? OAuth (intro)
- SESSION 23 — API Mini Projects ? Weather API ? Crypto prices ? NASA / COVID dataset ? Multi-API projects
- SESSION 24 — Case Study: Binance Crypto Price Tracker (Part 1) ? Fetch top 10 crypto prices ? Save to CSV ? Add price change %, high/low
- SESSION 25 — Case Study: Binance Crypto Price Tracker (Part 2) ? Most volatile coin ? Coins below average price ? Total traded volume ? Ranking ? Automate hourly script ? Handle rate limits & errors

Module 3) - Data collection, Cleaning, Visualization and Analysis

10

- SESSION 1 — NumPy Foundations (Part 1: Arrays & Creation) Topics Covered: What is NumPy & why it's essential for Data Science Installing & importing NumPy Creating NumPy arrays np.array() np.arange() np.linspace() np.zeros(), np.ones() np.full(), np.eye() Array attributes shape, ndim, size, dtype, itemsize Conversion between Python lists ? NumPy arrays Understanding contiguous memory layout
- SESSION 2 — NumPy Intermediate (Part 2: Indexing, Slicing, Broadcasting) Topics Covered: Indexing in 1D, 2D, and nD arrays Slicing arrays: [:], [start:end], [start:end:step] Selecting rows

& columns Boolean indexing Fancy indexing Broadcasting rules Scalar-to-array broadcasting
 Array-to-array broadcasting Practical examples: Masking out negative values Broadcasting for feature scaling

- SESSION 3 — NumPy Advanced (Part 3: Matrix Ops & Manipulation) Topics Covered: Mathematical operations Addition, subtraction, multiplication, division Element-wise vs matrix multiplication Matrix operations dot(), matmul() transpose, T Statistical functions mean, median, std, var, sum Linear algebra module np.linalg.inv(), det(), eig() Array manipulation reshape(), flatten(), ravel() concatenate(), stack(), split()
- SESSION 4 — Data Collection using Pandas (Part 1) Topics Covered: What is pandas & why DS heavily depends on it Types of data files CSV, Excel, TSV JSON Importing datasets pd.read_csv() pd.read_excel() Handling separators (comma, semicolon, tab) Reading large files chunksize parameter iterating through file chunks Detecting data issues early df.info() df.describe(include='all')
- SESSION 5 — Data Collection using Pandas (Part 2) Topics Covered: Reading data from databases MySQL / PostgreSQL pd.read_sql() / pd.read_sql_query() Web-based extraction (Intro only) Reading from URLs JSON APIs ? DataFrame conversion Loading multiple datasets Merge concepts Concatenation File path handling with pathlib
- SESSION 6 — Data Cleaning (Part 1: Missing Values) Topics Covered: Identifying missing data isnull(), notnull(), isna() Strategies to handle missing data dropna() parameters (how, axis, thresh)fillna() with mean, median, mode Forward fill, backward fill Missingness mechanisms MCAR, MAR, MNAR (conceptual) Practical examples: Cleaning customer demographic dataset
- SESSION 7 — Data Cleaning (Part 2: Outliers & Data Types) Topics Covered: Outlier detection Boxplot method Z-score method IQR method Outlier treatment Winsorization Log transform Capping and flooring Data type conversions astype() Converting objects ? numeric Fixing bool/Category/int mixups Practical cleanup tasks: Date parsing Fixing incorrect string formats
- SESSION 8 — Data Cleaning (Part 3: Duplicates, Typos, Consistency) Topics Covered: Detecting duplicates duplicated(), value_counts() Removing duplicates drop_duplicates() Fixing typos in categorical columns Using replace() Fuzzy matching overview Ensuring consistency Capitalization, trimming whitespace Unifying category formats (e.g., Yes/No ? 1/0)
- SESSION 9 — Working with Date & Time Data Topics Covered: Converting strings to datetime pd.to_datetime() Extracting datetime components day, month, year, weekday, quarter Setting datetime as index Resampling (daily, weekly, monthly) Handling timezone issues Creating new time-based features
- SESSION 10 — Data Visualization: Matplotlib (Part 1) Topics Covered: Introduction to Matplotlib Basic plots line plot scatter plot bar chart histogram Figure & axis concepts Customizing charts titles, labels legends colors Saving charts to files
- SESSION 11 — Data Visualization: Matplotlib (Part 2) Topics Covered: Subplots plt.subplots() grid of subplots Plot styling linestyle, linewidth markers Multi-axis charts Plotting categorical vs numeric data Annotation & text on charts
- SESSION 12 — Data Visualization with Seaborn (Part 1) Topics Covered: Introduction to Seaborn Axes-level vs figure-level APIs Distribution plots histplot kdeplot distplot (legacy) Categorical plots boxplot violinplot countplot Styling & themes (darkgrid, whitegrid)
- SESSION 13 — Data Visualization with Seaborn (Part 2) Topics Covered: Pairplot for quick multivariate analysis Heatmaps for correlation analysis relplot, catplot, displot Jointplot scatter + hist

regression line fit Visualizing group trends bar + confidence intervals

- SESSION 14 — Exploratory Data Analysis (EDA) (Part 1) Topics Covered: What is EDA & why it's critical Univariate analysis numerical vs categorical Bivariate analysis correlation scatterplots groupby analysis Multivariate analysis pairplot heatmap Identifying data patterns Detecting feature relationships
- SESSION 15 — Exploratory Data Analysis (EDA) (Part 2: Auto-EDA + Case Study Setup) Topics Covered: Auto EDA tools Pandas Profiling Sweetviz D-Tale Interpreting EDA reports Case Study Introduction: Superstore Sales Analysis (Kaggle Dataset) Tasks introduced for next module sections: Sales summary Profitability analysis Customer segment behavior Discount vs profit patterns City/region performance

Module 4) Working with Databases - Data Science

30

- SESSION 1 — Introduction to Databases & Installation Topics: What is a Database, RDBMS, SQL DB Engines (MySQL, PostgreSQL, Oracle) Installing MySQL / SQL Server Using SQL Workbench or DB Browser Trainer Depth: Explain why SQL is the backbone of analytics. Demo: Install MySQL ? Create database analytics_db.
- SESSION 2 — Basic SELECT & FROM Topics: SELECT column1, column2 SELECT * Renaming columns (AS keyword) Comments in SQL Demo: SELECT * FROM employees;
- SESSION 3 — WHERE Clause & Operators Topics: WHERE =, < >, >=, <= Logical operators AND, OR, NOT Demo: Filter employees with salary > 50000 and in 'IT' dept.
- SESSION 4 — Wildcards & Pattern Matching (LIKE, BETWEEN, IN) Topics: LIKE ("R%") BETWEEN (ranges) IN (multiple values) Demo: Find all customers whose names start with 'M'.
- SESSION 5 — DISTINCT, ORDER BY, LIMIT Topics: DISTINCT Sorting (ASC, DESC) Limiting rows Demo: Top 10 highest selling products.
- SESSION 6 — Aggregate Functions Topics: SUM(), COUNT(), AVG(), MIN(), MAX() Rounding numbers (ROUND) Demo: Total revenue, average salary, highest order value.
- SESSION 7 — GROUP BY & HAVING (Critical Analytics Topic) Topics: GROUP BY basics HAVING vs WHERE Aggregation filters Demo: Total revenue per region ? show only regions with revenue > 1M.
- SESSION 8 — JOINS: INNER, LEFT, RIGHT (Foundation Session) Topics: Relationship between tables Primary / Foreign Keys INNER JOIN LEFT JOIN RIGHT JOIN Demo: Join Orders + Customers ? Customer Segment ? Revenue.
- SESSION 9 — ADVANCED JOINS: FULL OUTER, SELF JOIN, CROSS JOIN Topics: FULL OUTER SELF JOIN CROSS JOIN (Cartesian) When to use each Demo: Self join employees table ? manager relationship.
- SESSION 10 — UNION & UNION ALL Topics: Combine multiple tables Difference between UNION vs UNION ALL Compatible column structures Demo: Combine online + offline sales tables.
- SESSION 11 — Subqueries (Scalar, Row, Table) Topics: Subquery in SELECT Subquery in WHERE Subquery in FROM (derived tables) Demo: Find employees earning above company average salary.

- SESSION 12 — Common Table Expressions (CTEs) Topics: WITH clause Temporary result sets Recursive CTE intro Demo: WITH SalesCTE AS (...) ? Calculate YoY growth.
- SESSION 13 — Window Functions (Part 1: OVER, PARTITION BY) Topics: OVER() PARTITION BY Row-level calculations Demo: Find each employee's salary vs department average.
- SESSION 14 — Window Functions (Part 2: ROW_NUMBER, RANK, DENSE_RANK) Topics: ROW_NUMBER RANK DENSE_RANK Demo: Top 3 selling products per region.
- SESSION 15 — Date Functions Topics: DATE, YEAR(), MONTH(), DAY(), NOW() DATE_ADD, DATE_SUB DATEDIFF Demo: Calculate customer churn age (days since last visit).
- SESSION 16 — String Functions Topics: CONCAT, LOWER, UPPER SUBSTR, LEFT, RIGHT REPLACE, TRIM Demo: Clean inconsistent product codes.
- SESSION 17 — CASE WHEN (Conditional Logic) Topics: CASE WHEN ELSE END Categorizing data on the fly Binning values Demo: Classify customers into High/Medium/Low value.
- SESSION 18 — Views Topics: CREATE VIEW Updating views View limitations Good practices for analytics views Demo: Create view for Dashboard reporting (Daily Sales).
- SESSION 19 — Indexing & Query Optimization Topics: Why queries become slow Index types EXPLAIN PLAN Avoiding full table scans Demo: Run slow query ? add index ? measure improvement.
- SESSION 20 — Import/Export Data (CSV, Excel, Python, Power BI) Topics: Load CSV to SQL Export query results Connect Excel to SQL Connect Power BI to SQL Demo: Connect Excel ? Run SQL queries ? Load pivot.
- SESSION 22 — Stored Procedures (Overview) Topics Covered: What stored procedures can automate Input parameters Small working example Benefits for automation
- SESSION 23 — Data Cleaning in SQL (Part 1) Topics Covered: Identifying NULLs Using COALESCE Removing duplicates via ROW_NUMBER() Identifying inconsistent entries
- SESSION 24 — Data Cleaning in SQL (Part 2) Topics Covered: Outlier detection using aggregate functions CASE statements for conditional cleaning Replacing incorrect categories Fixing inconsistent text formats
- SESSION 22 — SQL + Python Integration Topics: Using pandas read_sql() Jupyter Notebook SQL magic Combining SQL + Python for EDA Demo: Run SQL query ? Load into Pandas ? Visualize.
- SESSION 29 — SQLAlchemy ORM (Part 1) Topics Covered: What ORM means Defining models as Python classes Mapping attributes to columns Creating tables from Python
- SESSION 30 — SQLAlchemy ORM (Part 2) Topics Covered: Session management Querying ORM models Adding & committing objects Relationship loading strategies
- SESSION 31 — NoSQL Overview Topics Covered: SQL vs NoSQL differences Schema-less design When NoSQL is the right choice Overview of Redis, MongoDB, Cassandra, Neo4j
- SESSION 32 — MongoDB Basics Topics Covered: Documents & collections Inserting JSON documents find(), findOne() Projection (selecting fields) Operators: \$gt, \$lt, \$in, \$eq
- SESSION 33 — MongoDB Aggregation Pipeline Topics Covered: Stages: \$match, \$group, \$sort, \$limit Grouping large JSON datasets Creating aggregated summaries Indexing basics for speed
- Case- Study Dataset: Zomato Bangalore Restaurants – Kaggle Kaggle+ Goal As a Data Analyst for Zomato Bangalore, use SQL to help business teams answer key questions around restaurant performance, pricing, ratings, customer preferences, and market segmentation. The outcome should guide marketing, partnerships, and strategic expansion decisions.

- SESSION 34 — Database Design & Optimization Topics Covered: Schema design Normal forms: 1NF, 2NF, 3NF De-normalization Star schema (fact + dimension tables) Index types Reading execution plan Query optimization tips

Module 5) - Essential Statistics and Mathematics for Data Science

25

- SESSION 1 — Introduction to Linear Algebra (Vectors Basics) Topics Covered: What is linear algebra & why it matters in ML Understanding vectors Vector definition & dimensions Representing features as vectors Vector operations Addition & subtraction Scalar multiplication Dot product Geometric interpretation Similarity concepts Vector norms L1 norm L2 norm Use-cases: Distance calculations Cosine similarity intro
- SESSION 2 — Matrices (Part 1: Foundations) Topics Covered: What is a matrix? Rows, columns, and shapes Matrix notation & terminology Matrix addition & subtraction Scalar multiplication of matrices Practical examples: Storing datasets as matrices
- SESSION 3 — Matrices (Part 2: Multiplication & Operations) Topics Covered: Matrix multiplication rules Dimension compatibility Row × column concept Identity matrix Zero matrix Matrix transpose Real-world example: Transforming datasets using matrix operations
- SESSION 4 — Matrices (Part 3: Inverse & Determinant) Topics Covered: Determinant: concept & calculation Matrix inverse When inverse exists Importance in ML (solving systems) Relationship: $Ax = b$ Singular matrices Application in linear regression (analytical solution)
- SESSION 5 — Matrix Decompositions Topics Covered: Eigenvalues & eigenvectors Geometric meaning Derivation intuition Singular Value Decomposition (SVD) U, ?, V matrices Why decompositions matter Dimensionality reduction PCA foundation
- SESSION 6 — Applications of Linear Algebra in Data Science Topics Covered: Representing datasets as matrices Matrix multiplication for feature transformations Dimensionality reduction (PCA basics) Neural networks: weight matrices activation computations
- SESSION 7 — Intro to Statistics (Foundations) Topics Covered: What is statistics? Types of statistics Descriptive Inferential Types of data Numerical vs categorical Discrete vs continuous Understanding random variables Frequency distribution tables
- SESSION 8 — Measures of Central Tendency Topics Covered: Mean Sample vs population mean Median For skewed data Mode When to use what Example comparisons Outlier influence on measures
- SESSION 9 — Measures of Dispersion Topics Covered: Variance Standard deviation Range Coefficient of variation Why dispersion matters in ML Real-world examples (income inequality, ratings spread)
- SESSION 10 — Shape & Distribution (Skewness, Kurtosis) Topics Covered: Skewness Positive vs negative Kurtosis Leptokurtic, mesokurtic, platykurtic Interquartile Range (IQR) Understanding distributions using boxplots & histograms
- SESSION 11 — Sampling (Part 1) Topics Covered: Why sampling is required Population vs sample Simple random sampling Sampling bias Sampling distributions
- SESSION 12 — Sampling (Part 2) + Central Limit Theorem Topics Covered: Law of large numbers Central Limit Theorem (CLT) Normal approximation Importance for inferential statistics
- SESSION 13 — Hypothesis Testing (Part 1: Foundations) Topics Covered: What is hypothesis testing? Null hypothesis (H_0) Alternative hypothesis (H_1) Type I error Type II error p-value

interpretation Confidence intervals

- SESSION 14 — Hypothesis Testing (Part 2: t-test, z-test) Topics Covered: When to use z-test When to use t-test One-sample t-test Two-sample t-test Real DS cases: Compare average spending of two groups
- SESSION 15 — Hypothesis Testing (Part 3: Chi-square & F-test) Topics Covered: Chi-square test Categorical independence F-test Variance comparison Applications: AB testing Feature relationship validation
- SESSION 16 — Correlation & Causation Topics Covered: What is correlation? Pearson correlation coefficient (r) Interpretation rules Causation vs correlation Confounding variables Scatterplot interpretation
- SESSION 17 — Probability Basics Topics Covered: Experiment, event Sample space Mutually exclusive events Independent events Classical vs empirical probability Real-world examples
- SESSION 18 — Probability Rules Topics Covered: Addition rule Multiplication rule Complement rule Conditional probability (intro) Tree diagram examples
- SESSION 19 — Conditional Probability (Detailed) Topics Covered: Conditional probability formula Dependence vs independence Practical cases Disease testing Marketing campaign response prediction
- SESSION 20 — Bayes' Theorem Topics Covered: Statement & formula Prior, likelihood, posterior Real-world examples: Spam filtering Medical test accuracy
- SESSION 21 — Discrete Probability Distributions Topics Covered: Bernoulli Distribution Binomial Distribution Poisson Distribution Real-world modeling examples Customer arrivals Click-through rates
- SESSION 22 — Continuous Probability Distributions Topics Covered: Normal distribution Properties Empirical rule (68–95–99.7%) Uniform distribution Exponential distribution Applications in DS (anomaly detection, survival analysis)
- SESSION 23 — Applications of Probability in ML Topics Covered: Predictive modeling foundations Naive Bayes classifier Markov chains (intro) Probabilistic graphical models (overview) Distribution assumptions in ML algorithms
- SESSION 24 — Integration of Linear Algebra + Statistics + Probability Topics Covered: Covariance and correlation matrices Matrix form of linear regression PCA eigenvalue relationships Probability distributions in ML evaluation Using statistics during EDA
- SESSION 25 — Case Study: Customer Spending & Behavior Analysis Dataset: Customer Personality Analysis (Kaggle) Topics Covered: Data loading & inspection Income distribution analysis Average & standard deviation of spending Relationship between income & spending Hypothesis tests: Married vs single spending Mean income vs >50,000 Probability-based questions: $P(\text{spending} > 1000 \mid \text{income} > 80,000)$ Covariance matrix for numeric features Skewness & kurtosis analysis Preparing

Module 6) - Supervised Machine Learning

20

- SESSION 1 — Introduction to Machine Learning Topics Covered: What is Machine Learning? Definition How ML differs from traditional programming Three categories of ML Supervised (labelled data) Unsupervised (unlabelled data) Reinforcement Learning (agent-based) ML workflow overview Data ? Preprocessing ? Modeling ? Evaluation ? Deployment ML in real-world applications

Banking (credit scoring) E-commerce (recommendations) Manufacturing (predictive maintenance)
 Hands-on

- SESSION 2 — Core ML Concepts Topics Covered: Features & labels Independent vs dependent variables Dataset splitting Train set Test set Validation set Overfitting Definition Real examples Why it happens Underfitting Causes Simple vs complex models Bias-variance tradeoff Graphical explanation Impact on predictions Hands-on: Train-test split using sklearn
- SESSION 3 — Data Preprocessing (Part 1: Missing Values & Encoding) Topics Covered: Why preprocessing is essential Handling missing values Drop vs impute Imputation strategies Mean Median Mode Constant value Encoding categorical variables Label encoding One-hot encoding Ordinal encoding Practical lab: Cleaning a dataset with mixed types Handling categorical + numeric features
- SESSION 4 — Data Preprocessing (Part 2: Scaling, Selection, Outliers) Topics Covered: Feature scaling StandardScaler MinMaxScaler When scaling is required Distance-based models (KNN, SVM) Feature selection Correlation-based RFE (Recursive Feature Elimination) Tree-based feature importance Outlier detection & handling IQR Z-score Practical lab: Scaling + selection + cleaning pipeline
- SESSION 5 — Regression Algorithms (Part 1: Regression Foundations) Topics Covered: What is regression? When to use regression Assumptions in regression Simple Linear Regression Relationship between two variables Mathematical formula $y = mx + c$ Cost function Visualizing regression line Hands-on: Fitting simple regression with sklearn Plotting scatter + regression line
- SESSION 7 — Regression Model Evaluation Topics Covered: Evaluation metrics: MSE (Mean Squared Error) RMSE MAE R² score Residual plots Interpreting errors Overfitting detection via evaluation metrics Hands-on: Comparing models using metrics
- SESSION 6 — Regression Algorithms (Part 2: Multiple Linear Regression) Topics Covered: Multiple inputs ? single output Matrix representation of regression Coefficients & intercept Interpretation of regression coefficients Multicollinearity Detecting using VIF Hands-on: Fitting multi-regression model Feature importance interpretation
- SESSION 8 — Regression Algorithms (Part 3: Polynomial Regression) Topics Covered: Why linear models sometimes fail Adding polynomial features Quadratic, cubic terms Underfitting vs overfitting in polynomial models Regularization necessity at higher degrees Hands-on: PolynomialFeatures() Comparing linear vs polynomial regression
- SESSION 9 — Regularization Techniques (Part 1: L1 & L2) Topics Covered: Regularization necessity Overfitting prevention L1 Regularization (Lasso) Shrinking coefficients to zero Feature selection property L2 Regularization (Ridge) Penalizing large weights ElasticNet Combining L1 & L2 Hands-on: Fit Lasso, Ridge, ElasticNet Compare coefficient shrinkage
- SESSION 10 — Classification Algorithms (Part 1: Introduction + Logistic Regression) Topics Covered: What is classification? Binary vs multi-class Logistic Regression Sigmoid function Log-odds interpretation Decision boundary Threshold-based predictions Hands-on: Logistic regression on a binary dataset Plotting sigmoid
- SESSION 11 — Classification Algorithms (Part 2: KNN + Naive Bayes) Topics Covered: K-Nearest Neighbors Distance metrics (Euclidean, Manhattan) Choosing k Pros & cons Naive Bayes Bayes theorem Gaussian NB Multinomial NB Practical examples: Text classification Basic image classification

- SESSION 12 — Classification Algorithms (Part 3: Decision Trees) Topics Covered: Tree structure Splitting criteria Gini Entropy Depth, leaf nodes, pruning Interpretability advantages Hands-on: Visualizing decision tree Feature importance
- SESSION 13 — Classification Algorithms (Part 4: Random Forest) Topics Covered: Bagging concept Random feature selection Ensemble learning benefits Avoiding overfitting Feature importance from forests Hands-on: Building a random forest model Tuning n_estimators, max_depth
- SESSION 14 — Classification Algorithms (Part 5: Support Vector Machines) Topics Covered: What is a hyperplane? Maximum margin Linear vs non-linear separability Kernel trick intro RBF Polynomial Hands-on: SVM classification Kernel comparisons
- SESSION 15 — Classification Algorithms (Part 6: Boosting) Topics Covered: Boosting concept Weak ? strong learners Popular boosting methods AdaBoost Gradient Boosting Real-world use-cases Hands-on: Fit a GradientBoostingClassifier
- SESSION 16 — Classification Evaluation Metrics (Part 1) Topics Covered: Confusion matrix TP, TN, FP, FN Accuracy Precision Recall F1 score Hands-on: Manual calculation using confusion matrix
- SESSION 17 — Classification Evaluation Metrics (Part 2) Topics Covered: ROC curve AUC score Precision–Recall curve Log Loss When accuracy becomes misleading (imbalanced data) Hands-on: Plotting ROC & PR curves
- SESSION 18 — Model Evaluation & Improvement (Part 1: Cross Validation) Topics Covered: Need for cross-validation K-fold CV Stratified sampling Reducing variance with CV Hands-on: Using cross_val_score() Selecting best CV fold value
- SESSION 19 — Model Evaluation & Improvement (Part 2: Hyperparameter Tuning) Topics Covered: GridSearchCV RandomizedSearchCV Choosing search parameters Preventing overfitting during tuning Handling imbalanced data Class weights SMOTE oversampling Hands-on: Hyperparameter tuning on RF or SVM
- SESSION 20 — Case Study: Employee Attrition Prediction (End-to-End) Dataset: IBM HR Analytics Attrition Dataset (Kaggle) Topics Covered: Load and inspect dataset Treat missing values & outliers Encode categorical variables Feature scaling Train-test split Train Logistic Regression Train Random Forest Compare models Feature importance extraction ROC-AUC comparison Handle imbalance using SMOTE Apply cross-validation Hyperparameter tuning using GridSearchCV Final model selection Insights & am

Module 7) Unsupervised Machine Learning

5

- SESSION 1 — Introduction to Unsupervised Learning Topics Covered: What is Unsupervised Learning? Definition How it differs from supervised learning Types of Unsupervised Learning Clustering Dimensionality reduction Association rules (brief mention) When to use Unsupervised Learning No labeled data Hidden pattern discovery Real-world applications Customer segmentation Anomaly detection Recommendation systems Image compression Dataset characteristics Structu
- SESSION 2 — K-Means Clustering (Full Deep Dive) Topics Covered: What is clustering? Goal ? grouping similar items K-Means fundamentals Centroids Cluster assignment Updating centroids Algorithm workflow (step-by-step) Choose k Randomly initialize centroids Assign points to nearest centroid Update centroids Repeat until convergence Distance metrics Euclidean Manhattan (overview) Initialization methods Random KMeans++ Choosing optimal k Elbow Method

- SESSION 3 — DBSCAN & Density-Based Clustering Topics Covered: What is DBSCAN? Density-based clustering Identifying noise & outliers Concepts Epsilon (?) Min samples Core points Border points Noise points DBSCAN Algorithm Steps Find core points Expand cluster Mark noise When to use DBSCAN Non-spherical clusters High-noise datasets Large density variations Comparison vs K-Means K-Means fails on irregular shapes DBSCAN handles noise well P
- SESSION 4 — Dimensionality Reduction & Clustering Evaluation Metrics Topics Covered: Dimensionality Reduction Why dimension reduction? Curse of dimensionality Visualization Noise removal PCA (Principal Component Analysis) Variance explanation Covariance matrix Eigenvalues & eigenvectors Principal components Explained variance ratio Hands-on PCA with sklearn Plotting 2D projections Visualizing PCA-transformed clusters Clustering Evaluation Internal Metrics
- SESSION 5 — Case Study: Customer Segmentation (End-to-End) Dataset: Customer Segmentation Tutorial Dataset (Kaggle) Topics Covered: Phase 1 — Data Understanding Load & inspect dataset Identify key features: Age Annual income Spending score Detect missing values Basic summary statistics Outlier check Phase 2 — EDA Income vs Spending Score scatterplot Distribution plots Heatmap for correlations Removing or treating outliers Phase 3 — K-Means Clustering

MODULE 8 — STREAMLIT: BUILDING & DEPLOYING DATA APPS

10

- SESSION 1 — Introduction to Streamlit Topics Covered: What is Streamlit? Purpose and advantages for Data Science Why Streamlit instead of Flask/Django Streamlit application structure Python file-based app. Installing Streamlit pip install streamlit Checking version Understanding the Streamlit workflow Reruns on code save Reactive execution model Creating your first app st.title(), st.write(), st.text() Walkthrough of the Streamlit UI Center pane (app view) Sidebar Rerun button Settings Han
- SESSION 2 — Streamlit Fundamentals (Part 1: Display Elements) Topics Covered: Display Elements st.title(), st.header(), st.subheader() st.markdown() for rich formatting st.text(), st.code(), st.write() Display dictionaries, lists, DataFrames Images & Media st.image() st.audio(), st.video() Layout & Page Formatting Line breaks via markdown Using HTML/CSS inside markdown (safe) Hands-on: Build a basic dashboard page layout with titles, tables, and markdown descriptions.
- SESSION 3 — Streamlit Fundamentals (Part 2: Input Widgets) Topics Covered: Input Widgets st.text_input() st.number_input() st.selectbox() st.multiselect() st.radio() st.slider() st.checkbox() st.button() Capturing User Input Variables storing input Using input to update UI dynamically Form Handling st.form() and form_submit_button() Hands-on: Create an interactive input-based page where users enter name, age, and a choice ? display summary.
- SESSION 4 — Data Handling & Display (Pandas Integration) Topics Covered: Upload files st.file_uploader() Handling CSV, Excel Reading uploaded data into a DataFrame Displaying data st.dataframe() st.table() Conditional formatting preview Basic data exploration df.describe() output st.write(df.info()) Downloading processed files st.download_button() Hands-on: Build a mini “Data Explorer” where users upload CSV ? preview table ? download filtered version.
- SESSION 5 — Visualizations in Streamlit (Matplotlib, Seaborn, Plotly) Topics Covered: Matplotlib Integration Rendering plots using st.pyplot() Line, bar, scatter, histogram Seaborn Integration heatmap, countplot, boxplot, pairplot Plotly Charts Interactive charts px.scatter(), px.line(), px.bar() Hover tooltips UI Integration Placing plots inside columns Toggling plots with radio buttons Hands-

on: Build a visualization page where the user selects chart type ? chart updates dynamically.

- SESSION 6 — Building ML Apps (Part 1: Structure & Form Inputs) Topics Covered: Designing ML app structure Sidebar inputs Main panel outputs Loading pretrained ML models pickle joblib Creating input forms for prediction Numeric fields Selectbox encoding Running model.predict() on submit Handling invalid inputs Hands-on: Create a simple linear regression prediction app.
- SESSION 7 — Building ML Apps (Part 2: Full Classification App) Topics Covered: Creating feature inputs for classification Dropdowns for categorical features Sliders for numeric features Converting user input ? DataFrame Data preprocessing inside Streamlit Scaling Encoding Displaying prediction results st.metric() st.success() st.warning() Hands-on Project: Build a simple “Iris Flower Classifier” Streamlit app.
- SESSION 8 — Enhancing Apps (UX + State Management + Components) Topics Covered: UI Enhancements st.progress() st.spinner() st.balloons() st.toast() Layout Enhancements Columns Tabs Expanders Sidebars Streamlit Session State st.session_state Persisting user choices Preserving data across reruns Custom Styling HTML/CSS hacks Adding custom fonts & colors Hands-on: Add a sidebar menu + progress bar + tabs into an existing app.
- SESSION 9 — Deployment of Streamlit Apps Topics Covered: Deploy on Streamlit Cloud Creating account Pushing to GitHub Linking GitHub repo Deploying app Environment requirements requirements.txt runtime.txt Alternative Deployment Hugging Face Spaces Render Heroku (buildpack approach) AWS EC2 basics Version Control Git & GitHub basics Commit ? push ? deploy workflow Hands-on: Deploy the Iris classifier app live on Streamlit Cloud.
- SESSION 10 — Capstone: Build & Deploy a Real ML Streamlit App Project Options: 1?? Regression App (e.g., House Price Prediction) 2?? Classification App (e.g., Diabetes Prediction) 3?? NLP App (e.g., Sentiment Analysis) 4?? Computer Vision App (basic image classification) Topics Covered During Project: Designing UI flow Creating sidebar navigation Adding input validation Displaying charts + results Model file loading Putting it all together Deploying to Streamlit Cloud Final testing & debug

Module 9) Neural Network and Deep Learning

30

- SESSION 1 — Introduction to Deep Learning Topics Covered: What is Deep Learning? Difference from ML Why deep learning emerged (GPU, data availability) Biological neuron vs artificial neuron Where DL is used today Speech recognition Image recognition Recommendation systems Autonomous driving Deep Learning workflow overview Data ? Model ? Training ? Evaluation ? Deployment Hands-on Demo: Visual demonstration of a neural network with TensorFlow Playground.
- SESSION 2 — Perceptron & Multilayer Perceptron (MLP) Topics Covered: Perceptron architecture Weighted sum & activation Limitations of perceptron (XOR problem) Multilayer Perceptron Input ? Hidden ? Output Layers Forward propagation Why stacking layers helps learning Hands-on: Create a perceptron model in NumPy (without libraries).
- SESSION 3 — Neurons, Weights, and Biases (Deep Dive) Topics Covered: Anatomy of a neuron Weights Biases How weights determine decision boundary Initialization strategies: Zeros Random Xavier/He (conceptual) Why bias is essential

- SESSION 4 — Activation Functions (Part 1: Basics) Topics Covered: Why activation functions? Sigmoid Tanh ReLU Graphs & interpretations Issues: Vanishing gradient (sigmoid/tanh) ReLU dying problem Hands-on: Plot activation functions using matplotlib.
- SESSION 5 — Activation Functions (Part 2: Advanced) Topics Covered: Leaky ReLU Parametric ReLU Softmax for multi-class classification Hard sigmoid, GELU (overview) Choosing correct activation per layer
- SESSION 6 — Forward Propagation (Mathematical Explanation) Topics Covered: Matrix-based forward pass Dot products ? activations ? outputs Forward pass flow in multi-layer networks Visualization of activations through layers Hands-on: Implement forward propagation manually with NumPy.
- SESSION 7 — Loss Functions (Part 1) Topics Covered: What is a loss function? Mean Squared Error (MSE) Mean Absolute Error (MAE) Cross-Entropy Loss Binary Cross-Entropy Categorical Cross-Entropy Real-world examples
- SESSION 8 — Gradient Descent & Backpropagation (Part 1) Topics Covered: What is optimization? Gradients & partial derivatives Learning rate Local minima vs global minima Gradient Descent family overview
- SESSION 9 — Backpropagation (Part 2: Full Mechanics) Topics Covered: Chain rule Calculating gradients layer-by-layer Weight update rule Problems: Exploding & vanishing gradients Why deep networks struggled before ReLU Hands-on: Backprop demo using a 2-layer network. TensorFlow Overview: Developed by Google, TensorFlow provides extensive tools for designing and deploying deep learning models. Features: Tensor manipulation, automatic differentiation, and scalability. Keras API
- SESSION 10 — Regularization & Normalization Topics Covered: L1 & L2 regularization Dropout Batch Normalization Early stopping Practical benefits in DL models Hands-on: Add dropout & batchnorm to a simple Keras model.
- SESSION 11 — Optimizers (SGD, Momentum, RMSProp, Adam) Topics Covered: Need for optimization Understanding: SGD Momentum RMSProp Adam (industry standard) Learning rate schedules Hands-on: Train same model using different optimizers ? compare performance.
- SESSION 12 — Introduction to TensorFlow & Keras Topics Covered: Installing TensorFlow Understanding Keras layers & models Sequential vs Functional API Model summary & parameter count Device placement (CPU/GPU)
- SESSION 13 — Building Neural Networks in Keras (Part 1) Topics Covered: Creating layers Activation functions Dense networks Compiling models Training with fit() Training callbacks Hands-on: Build a 3-layer ANN for binary classification.
- SESSION 14 — Building Neural Networks (Part 2: Functional API) Topics Covered: Why Functional API? Multi-input & multi-output networks Advanced architectures Skip connections (conceptually)
- SESSION 15 — Model Evaluation & Monitoring Topics Covered: Training loss vs validation loss Overfitting detection Early stopping callback Performance metrics in DL Accuracy AUC Precision/recall for DL
- SESSION 16 — Hyperparameter Tuning & Model Improvement Topics Covered: Tuning layers, neurons Batch size vs epochs Optimizer parameter tuning Automated tuning (KerasTuner intro) Weight initialization tuning

- SESSION 17 — Saving & Loading Models Topics Covered: model.save() ? .h5 files Saving architecture + weights Checkpointing Loading models for inference Exporting for deployment Hands-on: Save & reload a trained neural network.
- SESSION 18 — Introduction to CNNs Topics Covered: Why CNNs for images? CNN architecture overview Local receptive fields Translation invariance Feature maps
- SESSION 20 — Pooling Layers Topics Covered: Max pooling Average pooling Why pooling helps Reducing dimensions Combining conv + pooling layers
- SESSION 19 — Convolution & Feature Extraction Topics Covered: Filters & kernels Stride, padding Convolution operation (manual example) Activation maps Visualizing filters Hands-on: Implement convolution with NumPy on a small matrix.
- SESSION 21 — Building CNNs in Keras Topics Covered: Conv2D MaxPooling2D Flatten Dense layers at the end BatchNorm + Dropout Compiling & training Hands-on: Build a CNN on CIFAR-10 or MNIST.
- SESSION 22 — Image Preprocessing & Augmentation Topics Covered: rescale resize grayscale ? RGB conversions Augmentation techniques: rotation zoom flip brightness shift shear Using ImageDataGenerator
- SESSION 23 — Transfer Learning & Pretrained CNNs Topics Covered: Why transfer learning? Pretrained models: VGG16 ResNet50 Inception Feature extraction vs fine-tuning Freezing layers Adding custom top layers Hands-on: Fine-tune VGG16 for a custom dataset.
- SESSION 24 — Introduction to RNNs Topics Covered: Why sequence models? Time-series, text, speech RNN architecture Hidden states Backpropagation Through Time (BPTT) Vanishing gradient issue
- SESSION 25 — LSTM (Long Short-Term Memory Networks) Topics Covered: Cell state Input/forget/output gates How LSTM solves vanishing gradients Applications: stock prediction text generation Hands-on: Build a simple LSTM for sequence prediction.
- SESSION 26 — GRU (Gated Recurrent Unit) Topics Covered: GRU gates Difference vs LSTM When GRU is preferred Performance comparison Hands-on: Train GRU vs LSTM on a small dataset.
- SESSION 27 — Case Study Setup: FER Data Understanding Topics Covered: Dataset loading Data size & classes Data imbalance Image dimensions Preprocessing steps Train/validation/test split
- SESSION 28 — Baseline CNN Model Topics Covered: Basic Conv2D stack Training baseline model Checking accuracy & problems Hands-on: Train basic CNN ? evaluate accuracy.
- SESSION 29 — Improving Model + Transfer Learning Topics Covered: Data augmentation to handle imbalance Fine-tuning using VGG16 / ResNet50 Feature extraction vs full fine-tuning Improving generalization Hands-on: Train a transfer-learning model for FER.
- SESSION 30 — Deployment in Streamlit Topics Covered: Saving final model Creating inference pipeline Creating upload interface Real-time predictions Displaying emotion results Deploying to Streamlit Cloud Final Deliverable: A full Facial Emotion Recognition App.

MODULE 10) — COMPUTER VISION (CV) FUNDAMENTALS + PROJECTS	10
--	-----------

- SESSION 1 — Introduction to Computer Vision & OpenCV Topics Covered: What is Computer Vision? Understanding images as numerical matrices Why CV is important in AI,

robotics, healthcare Installing OpenCV pip install opencv-python Reading images cv2.imread() Color channels: BGR vs RGB Displaying images cv2.imshow(), cv2.waitKey(), cv2.destroyAllWindows() Image paths & directory issues Basic image properties shape (height, width, channels) dtype resizing basics Image formats: JPG, PNG,

- SESSION 2 — Image Processing Basics (Part 1: Transformations) Topics Covered: Image resizing cv2.resize() Interpolation methods (INTER_LINEAR, INTER_AREA) Image rotation Rotation matrices cv2.warpAffine() Cropping Slicing image arrays Flipping Horizontal, Vertical, Both Color space conversions BGR ? RGB BGR ? GRAY BGR ? HSV Hands-on: Build an “image transformation station” script with multiple operations
- SESSION 3 — Image Processing Basics (Part 2: Filters) Topics Covered: Blurring techniques Averaging Gaussian blur Median blur Bilateral filter (edge-preserving) Edge detection Sobel filter Laplacian Canny edge detection Sharpening filters Kernel-based sharpening Noise types Gaussian Salt & pepper Hands-on: Apply all blurs on sample images & visually compare
- SESSION 4 — Drawing & Masks + Bitwise Operations Topics Covered: Creating blank images Drawing shapes Rectangle, Circle, Line, Polygon Adding text to images Masks Binary masks Creating circular / square masks Using masks to isolate objects Bitwise operations AND OR XOR Mask extraction Hands-on: Cut out specific areas of an image using masks
- SESSION 5 — Image Thresholding & Contours Topics Covered: Thresholding Techniques Global thresholding Adaptive thresholding Mean Gaussian Otsu’s thresholding Contours Finding contours cv2.findContours() Contour hierarchy Drawing contours Contour features Area Perimeter Approximation Bounding rectangles Hands-on: Detect objects in an image using contours
- SESSION 6 — Face & Object Detection with OpenCV Topics Covered: Haar Cascade classifiers face detection XML files Running detection cv2.CascadeClassifier() detectMultiScale() Tuning detection parameters scaleFactor minNeighbors Eye detection Integrating multiple cascades Hands-on: Detect face + eyes from webcam video feed
- SESSION 7 — Image Segmentation & Histogram Analysis Topics Covered: Histogram Analysis Plotting histograms Distribution of pixel intensities Histogram equalization cv2.equalizeHist() CLAHE (Contrast Limited Adaptive Histogram Equalization) Segmentation Techniques Threshold-based segmentation k-means image segmentation Watershed algorithm (intro) Hands-on: Segment foreground/background in simple images.
- SESSION 8 — Feature Detection & Matching Topics Covered: Corners & keypoints Harris corner detection Shi-Tomasi ORB (Oriented FAST and Rotated BRIEF) Keypoint extraction Descriptor computation Feature matching Brute-Force matcher FLANN-based matcher Application: Logo detection Object recognition Hands-on: Match keypoints between two images (e.g., cover photo variations).
- SESSION 9 — Video Processing (Frame-by-frame Pipelines) Topics Covered: Reading webcam/video files cv2.VideoCapture() Writing videos cv2.VideoWriter() Processing pipeline Frame reading Converting colors Detecting objects Overlaying shapes/text Real-time face detection pipeline FPS optimization tips Hands-on: Build a “Real-time Edge Detector Webcam App”
- SESSION 10 — Capstone Project: Image Classification Pipeline Goal: Build a complete classical CV + DL hybrid app. Dataset: Any small image dataset (Cats vs Dogs / Flowers / Handwritten digits) Topics Covered: Phase 1 — Data Collection Collect sample dataset Folder structure Load images with cv2 Phase 2 — Preprocessing Resize images Normalize pixel values Augment with OpenCV (rotate, flip, blur) Phase 3 — Feature Extraction Using ORB or raw pixels Flatten feature vectors

Phase 4 — ML model train

MODULE 11) — NATURAL LANGUAGE PROCESSING (NLP)

11

- SESSION 1 — Introduction to NLP & Text Structure Topics Covered: What is NLP? Understanding unstructured text data NLP vs traditional ML Where NLP is used Sentiment analysis Chatbots Machine translation Topic modeling Anatomy of text data Characters ? Words ? Sentences ? Documents Corpus, vocabulary NLP pipeline overview Data cleaning ? Preprocessing ? Feature extraction ? Modeling Hands-on: Load raw text dataset Print sample sentences Basic text overview
- SESSION 2 — Text Cleaning & Preprocessing (Part 1) Topics Covered: Basic cleaning techniques Converting to lowercase Removing punctuation Removing special characters Removing numbers Tokenization word_tokenize Split-based tokenization Stopwords removal NLTK stopword lists Custom stopwords Lemmatization WordNetLemmatizer Stemming Porter stemmer Differences with lemmatization Hands-on: Apply cleaning on text column Compare stemmed vs lemmatized outputs
- SESSION 3 — Text Cleaning & Preprocessing (Part 2: Advanced) Topics Covered: POS (Part of Speech) tagging Chunking (noun/verb phrase extraction) Named Entity Recognition (NER) Person Organization Location Handling emojis Mapping emojis to words Handling social media text URLs removal @mentions, #hashtags Handling repeated characters "coooooool" ? "cool" Hands-on: Extract NER tags from sample text Perform POS tagging and chunk extraction
- SESSION 4 — Text Vectorization (Part 1: Classic Methods) Topics Covered: Why text needs numeric representation Bag of Words Vocabulary creation Document-term matrix Advantages & limitations N-grams Unigrams, bigrams, trigrams Why bigrams improve context TF-IDF (Term Frequency – Inverse Document Frequency) Formula breakdown Why TF-IDF helps compared to BoW Sparse matrices overview Hands-on: Build BoW matrix Build TF-IDF matrix Visualize top words
- SESSION 5 — Text Vectorization (Part 2: Word Embeddings) Topics Covered: What are embeddings? Dense vector representations Word2Vec (conceptual) CBOW Skip-Gram GloVe embeddings Why embeddings outperform BoW Using pre-trained embeddings Google News Word2Vec GloVe Twitter Representing sentences using average embeddings Hands-on: Visualize word similarity with Word2Vec Perform analogy tasks ("king – man + woman = ?")
- SESSION 6 — Classical NLP Models Topics Covered: Naive Bayes for text Why NB works well on text Logistic regression for text classification SVM for high-dimensional vectors Feature engineering Word counts Sentiment lexicons Custom dictionaries Evaluation metrics Accuracy F1 Confusion matrix Hands-on: Train Naive Bayes on sentiment dataset Evaluate using confusion matrix
- SESSION 7 — Transformer-Based NLP (Introduction to Modern NLP) Topics Covered: Limitations of classical NLP Transformational shift in NLP What is a transformer? Encoder Decoder Attention mechanism Self-attention explanation Query Key Value Attention scores HuggingFace Transformers library Overview of famous models BERT RoBERTa DistilBERT GPT (conceptual) Hands-on (theory + small demo): Use pipeline() for sentiment analysis
- SESSION 8 — Using BERT for Text Classification Topics Covered: Tokenization for transformer models WordPiece tokenizer Attention masks Input formatting for BERT Fine-tuning vs feature extraction Training pipeline overview Batch creation GPU acceleration Model evaluation Hands-on:

Build a BERT-based text classifier Predict sentiment for unseen text

- SESSION 9 — Topic Modeling + Summarization Topics Covered: Topic Modeling Latent Dirichlet Allocation (LDA) Document-topic distribution Topic-word distribution Choosing number of topics Visualizing topics (pyLDAvis) Use-cases Customer feedback analysis Text Summarization Extractive summarization Text rank Abstractive summarization Transformer-based summary generators Hands-on: Build simple LDA model Generate summaries using transformers pipeline
- SESSION 10 — Case Study: Twitter Sentiment Analysis End-to-End Topics Covered: Phase 1 — Data Collection Twitter dataset (CSV or API-based) Text cleaning Remove emojis, hashtags, mentions, URLs Phase 2 — Preprocessing Tokenization Lemmatization Stopword removal Bigram/trigram generation Phase 3 — Feature Engineering TF-IDF Word embeddings Sentiment lexicons mapping Phase 4 — Modeling Train Naive Bayes Train Logistic Regression Compare classic ML vs BERT Phase 5 — Evaluation Confusion matrix

MODULE 12) — REINFORCEMENT LEARNING

13

- SESSION 1 — What is Reinforcement Learning? (RL Fundamentals – Part 1) Topics Covered: Definition of Reinforcement Learning How RL differs from Supervised & Unsupervised Learning Real-world RL applications: Game AI Robotics Self-driving cars Recommendation systems
- SESSION 2 — Key Components of RL (RL Fundamentals – Part 2) Topics Covered: Agent Environment State Action Reward Episodic vs continuous tasks Reward function design Cumulative returns concept (G, total future rewards)
- SESSION 3 — Exploration vs Exploitation (RL Fundamentals – Part 3) Topics Covered: Why exploration is required Pure exploration vs pure exploitation Balancing the trade-off ?-greedy policy (concept) Example: Multi-armed bandit intuition
- SESSION 4 — Markov Decision Process (MDP) – Part 1 Topics Covered: What is an MDP? Role of MDP in RL MDP elements: States Actions Transition probability Rewards Discount factor ? Deterministic vs probabilistic transitions
- SESSION 5 — Markov Decision Process (MDP) – Part 2 Topics Covered: Expected return Value functions: State-value function (V) Action-value function (Q) Policy (?): Deterministic Stochastic Optimal policy intuition
- SESSION 6 — Bellman Equations (Conceptual) Topics Covered: Bellman Expectation Equation Bellman Optimality Equation Why Bellman equations are fundamental to RL Intuition of dynamic programming
- SESSION 7 — Q-Learning (Tabular RL) — Part 1 Topics Covered: What is Q-Learning? Q-table structure Q(s, a) value interpretation Action selection strategies ?-greedy action policy Exploration decay
- SESSION 8 — Q-Learning (Tabular RL) — Part 2 Topics Covered: Q-value update rule, Learning rate ? Discount factor ? Episode vs timestep updates Convergence intuition Limitations of Tabular Q-learning
- SESSION 9 — Deep Q-Learning (DQN) — Introduction Topics Covered: Why Q-tables fail for large state spaces Using neural networks to approximate Q-values General architecture of DQN Input ? hidden layers ? Q-values output Replay buffer (experience replay) Target network concept
- SESSION 10 — Deep Q-Learning (DQN) — Concepts Continued Topics Covered: Stabilizing DQN training Training loop structure Updating target network Mini-batch sampling Loss function for

DQN Action selection in DQN Common challenges: catastrophic forgetting instability sample inefficiency

- SESSION 11 — Extensions of DQN (Optional Overview) Topics Covered: Policy Gradient methods (concept only) Actor-Critic methods PPO (Proximal Policy Optimization – high-level overview) When to use value-based vs policy-based RL Advantages & limitations
- SESSION 12 — Case Study Setup: CartPole Environment Topics Covered: Environment link (Gym Classic Control): https://www.gymlibrary.dev/environments/classic_control/cart_pole/ Understanding the environment: State representation: Cart position Cart velocity Pole angle Pole angular velocity Actions: 0 = Push left 1 = Push right Reward function: +1 per timestep Episode termination: Pole angle threshold Cart position threshold 500-step cap
- SESSION 13 — Case Study: CartPole — Implement RL Algorithms Topics Covered: Train using Tabular Q-Learning (conceptual demonstration) Implement DQN for CartPole Steps: Initialize replay buffer Train neural network for Q-value approximation Update target network Track episode rewards Evaluate performance curves Optional: Build a small Streamlit visualization app

MODULE 13 — ML DEPLOYMENT & LIFECYCLE MANAGEMENT	11
---	-----------

- SESSION 1 — Introduction to Model Deployment Topics Covered (expanded but unchanged): What is ML deployment? From Jupyter Notebook ? Production Why deployment is necessary Types of deployment: Local deployment Cloud deployment (AWS / Azure / GCP concepts only) Inference modes: Real-time inference Batch inference Serving ML models with REST APIs What is an API endpoint? Request ? Processing ? Response lifecycle
- SESSION 2 — Saving & Loading Models (Part 1) Topics Covered: Model persistence using: pickle joblib Model version files: .pkl, .joblib, .sav Serialization concepts: Why models must be serialized Loading models safely
- SESSION 3 — Saving & Loading Models (Part 2 – DL Models) Topics Covered: Saving deep learning models: TensorFlow/Keras: .h5, SavedModel format PyTorch: .pt, .pth Re-loading models for inference Verifying model performance after loading Common errors in loading models (path, environment mismatch)
- SESSION 4 — Building REST APIs for ML Models (Flask Basics) Topics Covered: What is Flask? Creating the first API endpoint Input/output formats: JSON payloads Form data Loading ML models inside Flask Returning predictions via JSON
- SESSION 5 — Building REST APIs for ML (FastAPI Basics) Topics Covered: Advantages of FastAPI Defining Pydantic models for validation Async endpoints for scalable ML apps Testing responses Hands-on: convert Flask API ? FastAPI API
- SESSION 6 — Testing APIs Using Postman & Curl Topics Covered: Introduction to Postman Sending POST requests Sending JSON payload Checking response codes Testing failure cases Using curl for API testing in terminal Logging API calls
- SESSION 7 — Model Versioning & Experiment Tracking (MLflow – Part 1) Topics Covered: Why version control is required in ML What changes between model versions? Dataset Features Hyperparameters Architecture Introduction to MLflow Tracking models Tracking experiments MLflow UI Logging metrics & artifacts

- SESSION 8 — Model Registry, Versioning Tools (MLflow – Part 2) Topics Covered: MLflow Model Registry Register model Promote model to “Production” Archive older models DVC for dataset versioning Local + cloud remote storage Cloud storage for models: AWS S3 Azure Blob Storage GCP Storage
- SESSION 9 — CI/CD Pipelines for ML (Part 1) Topics Covered: What is CI/CD in ML? Difference between DevOps & MLOps Automated ML workflows CI steps: Linting Testing Packaging CD steps: Auto-deployment to API GitHub Actions / GitLab CI concepts
- SESSION 10 — CI/CD Pipelines for ML (Part 2 – Cloud-Native) Topics Covered: Cloud-based MLOps frameworks (conceptual): AWS SageMaker Pipelines GCP Vertex AI Pipelines Azure ML Pipelines Trigger-based retraining Scheduling jobs Continuous delivery of model updates
- SESSION 11 — Model Monitoring & Maintenance + Case Study Topics Covered: Part A — Monitoring & Maintenance What to monitor: Drift detection Prediction errors Latency Data quality Logging predictions Alerts & triggers for re-training Using MLflow / Weights & Biases for monitoring Part B — Case Study / Project End-to-end model deployment using: Flask MLflow Cloud platform (optional) Tasks: Deploy initial version Observe drift Update to new version Test versioning & rollba

Module 14) Introduction to Gen AI
5

- SESSION 1 — Introduction to Generative Models
- SESSION 2 — Autoencoders & Variational Autoencoders (VAE)
- SESSION 3 — Generative Adversarial Networks (GANs)
- SESSION 4 — Diffusion Models (Stable Diffusion Intro)
- SESSION 5 — Applying Generative Models in Industry

MODULE 15 — TRANSFORMER ARCHITECTURE & LLMs
10

- SESSION 1 — Overview of LLMs
- SESSION 2 — Open-Source vs Closed-Source LLMs
- SESSION 3 — Running LLMs Locally vs Cloud
- SESSION 4 — Why Transformers Replaced RNNs & CNNs work
- SESSION 5 — Transformer Architecture (Encoder / Decoder / Encoder–Decoder)
- SESSION 6 — Multi-Head Self-Attention & Positional Encoding
- SESSION 7 — Feedforward Layers, Residual Connections & Layer Norm
- SESSION 8 — Vision Transformers (ViT)
- SESSION 9 — LLM Model Families (GPT, BERT, T5, Falcon, LLaMA)
- SESSION 10 — Zero-Shot, Few-Shot & Chain-of-Thought Reasoning + Case Study App

MODULE 16 — LANGCHAIN FOR GENAI APPLICATIONS
8

- SESSION 1 — Introduction to LangChain (Part 1)
- SESSION 2 — Introduction to LangChain (Part 2) — LangChain vs LlamaIndex
- SESSION 3 — Core Components of LangChain

- SESSION 4 — Practical Implementation (Part 1): Build a Simple QA Bot
- SESSION 5 — Practical Implementation (Part 2): Memory + Custom Tools
- SESSION 6 — Integrations (Part 1): Vector Databases & Retrieval
- SESSION 7 — Integrations (Part 2): APIs + Streamlit/Gradio UI
- SESSION 8 — Case Study: LangChain-Powered Internal Assistant

MODULE 17 — PROMPT ENGINEERING

7

- SESSION 1 — Prompt Fundamentals (Part 1)
- SESSION 2 — Prompt Fundamentals (Part 2): Crafting Structured Prompts
- SESSION 3 — Prompting Techniques: Zero-Shot, One-Shot, Few-Shot
- SESSION 4 — Chain-of-Thought (CoT) & Reasoning Prompts
- SESSION 5 — System, Role & Context Prompts
- SESSION 6 — Prompt Templates with LangChain
- SESSION 7 — Evaluating Prompts & Avoiding Hallucinations

MODULE 18 — RETRIEVAL-AUGMENTED GENERATION (RAG)

8

- SESSION 1 — Introduction to RAG: What It Is, Need & Benefits
- SESSION 2 — RAG Architecture: Retriever + Generator + Knowledge Base
- SESSION 3 — Data Ingestion & Cleaning for RAG Systems
- SESSION 4 — Chunking & Embeddings (OpenAI / Sentence Transformers)
- SESSION 5 — Vector Databases (ChromaDB, FAISS, Pinecone)
- SESSION 6 — Context Injection into Prompts (Retriever ? LLM)
- SESSION 7 — Building a Full RAG Pipeline (LangChain / LlamaIndex)
- SESSION 8 — Case Study: Enterprise Document Intelligence Assistant

MODULE 19 — FINE-TUNING & MODEL CUSTOMIZATION

8

- SESSION 1 — Why Fine-Tune? (vs RAG, Prompting & Other Methods)
- SESSION 2 — Prompt Tuning vs Fine-Tuning vs RAG (Comparison Framework)
- SESSION 3 — Fine-Tuning Approaches (Part 1): Full Fine-Tuning & LoRA/PEFT Theory
- SESSION 4 — Fine-Tuning Approaches (Part 2): QLoRA & Instruction Fine-Tuning
- SESSION 5 — Data Preparation & Tokenization for Fine-Tuning
- SESSION 6 — Training Using HuggingFace Trainer (Hands-On Architecture)
- SESSION 7 — Evaluating, Saving Models & Avoiding Overfitting/Data Leakage
- SESSION 8 — Case Study: Fine-Tuned Customer Support LLM (From the PDF)

MODULE 20 — AGENTIC AI

15

- SESSION 1 — Introduction to Agentic AI (Part 1)

- SESSION 2 — Introduction to Agentic AI (Part 2)
- SESSION 3 — Agent Architecture & Workflow (Part 1)
- SESSION 4 — Agent Architecture & Workflow (Part 2)
- SESSION 5 — Agent Architecture & Workflow (Part 3)
- SESSION 6 — Building a Basic Agent (Part 1)
- SESSION 7 — Building a Basic Agent (Part 2)
- SESSION 8 — Building a Basic Agent (Part 3) — Weather Assistant Example
- SESSION 9 — Agentic AI with LLMs (Part 1)
- SESSION 10 — Agentic AI with LLMs (Part 2)
- SESSION 11 — Agentic AI with LLMs (Part 3): Multi-Agent Systems
- SESSION 12 — Advanced Agentic Concepts (Part 1)
- SESSION 13 — Advanced Agentic Concepts (Part 2)
- SESSION 14 — Case Study: Building a Personal Assistant Agent (Part 1)
- SESSION 15 — Case Study: Building a Personal Assistant Agent (Part 2)