🤖 📊 **12-Hour Workshop: Machine Learning with Statistics**
"From Statistical Foundations to Predictive Models"

📅 **2-Day Schedule (12 Hours Total)**
Ideal for a full-day workshop (with breaks) for 2-day split session.

Day 1: Statistics for Machine Learning (4 Hours)
🕐 **Session 1 – Descriptive Statistics (2 Hours)**

- Types of Data: Qualitative vs Quantitative
- Central Tendency: Mean, Median, Mode
- Dispersion: Range, Variance, Standard Deviation
- Distribution: Normal, Skewness, Kurtosis
- Hands-on:
    o Use pandas, numpy, and matplotlib to explore datasets
    o Visualize distributions with seaborn (hist, boxplot, violin)

🕐 **Session 2 – Inferential Statistics (2 Hours)**

- Sampling, Population vs Sample
- Probability Basics
- Hypothesis Testing: Null vs Alternate
- Confidence Interval
- p-value and Significance
- Hands-on:
    o Perform t-tests, z-tests with scipy.stats
    o Interpret results on real-world datasets

Day 2: Machine Learning Fundamentals (4 Hours)
🕐 **Session 3 – Supervised Learning: Regression (2 Hours)**

- Types of Machine Learning
- Introduction to Linear Regression
- Cost Function, Gradient Descent
- Model Evaluation: MAE, MSE, R^2 Score
- Hands-on:
    o Use scikit-learn to implement Linear Regression
    o Predict house prices, salary, or marks

🕐 **Session 4 – Supervised Learning: Classification (2 Hours)**

- Logistic Regression Concepts
- Decision Trees: Gini, Entropy
- Evaluation Metrics: Accuracy, Confusion Matrix, Precision, Recall, F1-score
- Hands-on:
    o Use scikit-learn for Logistic Regression & Decision Tree
    o Apply on IRIS or Titanic dataset
    o Visualize decision boundaries

🕐 **Day 3: Applied ML with Statistics (4 Hours)**
Session 5 – Feature Engineering & Data Preprocessing (2 Hours)
- Handling Missing Values
- Encoding Categorical Variables (Label, OneHot)
- Feature Scaling: Standardization & Normalization
- Correlation & Multicollinearity
- Hands-on:
  - Clean raw datasets using pandas
  - Create correlation heatmaps using seaborn
  - Apply scaling with StandardScaler, MinMaxScaler

🕐 **Session 6 – Model Selection & Mini Project (2 Hours)**
- Train/Test Split, Cross-validation
- Underfitting vs Overfitting
- Introduction to Random Forest
- Hands-on:
  - Build and compare multiple models
  - Perform a mini-project:
    - Predict loan approval / student grades / customer churn
  - Submit results and evaluate performance

🎯 **Workshop Deliverables**
- Certificate of Participation
- Sample Project Code & Datasets
- Machine Learning with Stats Roadmap PDF
- Interview Questions Cheat Sheet
- Colab notebooks for all sessions

🌟 **Learning Outcomes**
By the end of this workshop, students will:
- Understand core statistics and how they support ML
- Build and evaluate ML models using real datasets
- Perform data preprocessing, visualization, and model tuning
- Be ready to take on ML internships or academic projects