Data Science syllabus and project building.

📌 Section 1: Introduction to Data Science

1️⃣ What is Data Science?

- Definition and Scope
- History and Evolution
- Difference between Data Science, Data Analytics, Data Engineering, and Business Analytics
- Applications of Data Science in Industries

2️⃣ Data Science Process

- Problem Definition
- Data Collection
- Data Cleaning and Preparation
- Exploratory Data Analysis (EDA)
- Feature Engineering
- Model Building
- Model Evaluation
- Deployment
- Model Monitoring

3️⃣ Tools & Technologies in Data Science

- Programming Languages: Python, R
- IDEs: Jupyter, VS Code, Spyder
- Libraries and Frameworks:
    - **Python**: Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, TensorFlow, Keras, PyTorch
    - **R**: dplyr, ggplot2, caret

- Database: SQL, MongoDB
- Big Data Tools (overview): Hadoop, Spark
- Cloud Platforms (overview): AWS, GCP, Azure

📌 **Section 2: Programming for Data Science (Python Focused)**

- Python Basics (covered in previous syllabus)
- Numpy (Numerical Python)
    - Arrays creation and manipulation
    - Array slicing, indexing
    - Mathematical operations
    - Broadcasting
- Pandas (Data Manipulation)
    - Series and DataFrame
    - Reading and writing data (CSV, Excel, JSON)
    - Handling missing data
    - Data filtering, sorting, merging, grouping
- Matplotlib and Seaborn (Data Visualization)
    - Basic plots: Line, Bar, Histogram, Scatter, Pie
    - Customizing Plots
    - Subplots
    - Seaborn specialized plots: Boxplot, Heatmap, Pairplot, Violinplot

---

📌 **Section 3: Statistics for Data Science**

1️⃣ **Descriptive Statistics**

- Measures of Central Tendency (Mean, Median, Mode)
- Measures of Dispersion (Range, Variance, Standard Deviation)
- Percentiles and Quartiles

2️⃣ **Probability**

- Basic Probability Theory
- Conditional Probability
- Bayes' Theorem
- Probability Distributions:
    - Normal Distribution
    - Binomial Distribution
    - Poisson Distribution

3️⃣ **Inferential Statistics**

- Sampling Techniques
- Hypothesis Testing
    - Null and Alternative Hypothesis
    - t-test, z-test, ANOVA
    - Chi-Square Test
- Confidence Intervals
- Correlation and Covariance

📌 **Section 4: Data Preprocessing and Cleaning**

- Handling Missing Data
- Handling Duplicates
- Outlier Detection and Treatment
- Data Encoding (Label Encoding, One-Hot Encoding)
- Feature Scaling (Normalization, Standardization)
- Feature Selection and Extraction

📌 **Section 5: Exploratory Data Analysis (EDA)**

- Data Understanding
- Univariate, Bivariate, and Multivariate Analysis
- Data Visualization Techniques

- Detecting patterns, relationships, anomalies
- Heatmaps, Pairplots, Histograms, Boxplots

📌 Section 6: Machine Learning Fundamentals

1️⃣ Introduction to Machine Learning

- Types of Machine Learning:
    - Supervised Learning
    - Unsupervised Learning
    - Reinforcement Learning

2️⃣ Supervised Learning

- Linear Regression
- Logistic Regression
- Decision Trees
- Random Forest
- Support Vector Machines (SVM)
- K-Nearest Neighbors (KNN)
- Naïve Bayes

3️⃣ Unsupervised Learning

- K-Means Clustering
- Hierarchical Clustering
- Principal Component Analysis (PCA)

4️⃣ Model Evaluation Metrics

- Confusion Matrix
- Accuracy, Precision, Recall, F1-Score
- ROC-AUC Curve
- Cross Validation (K-Fold, Stratified K-Fold)

📌 **Section 7: Deep Learning (AI)**

- Introduction to Deep Learning
- Artificial Neural Networks (ANN)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)
- LSTM (Long Short-Term Memory)
- TensorFlow and Keras basics
- Activation Functions
- Loss Functions
- Optimizers

📌 **Section 8: Natural Language Processing (NLP)**

- Text Preprocessing (Tokenization, Stopwords Removal, Lemmatization, Stemming)
- Bag of Words (BoW) Model
- TF-IDF (Term Frequency - Inverse Document Frequency)
- Word Embeddings (Word2Vec, GloVe)
- Sentiment Analysis
- Text Classification
- Named Entity Recognition (NER)

📌 **Section 9: Time Series Analysis**

- Introduction to Time Series Data
- Time Series Components: Trend, Seasonality, Noise
- Moving Averages, Exponential Smoothing
- ARIMA, SARIMA Models
- Stationarity and Differencing
- Forecasting

📌 Section 10: Big Data & Cloud Integration (Optional / Advanced)

- Introduction to Big Data
- Hadoop Ecosystem
- Apache Spark and PySpark
- Working with AWS S3, Lambda for Data Storage and Deployment
- Cloud ML services overview

📌 Section 11: Project Management and Deployment

- Working with Git and GitHub
- Creating Dashboards (Streamlit, Dash)
- API Integration
- Model Deployment (Flask, FastAPI)
- Using Docker for ML Model Deployment
- Model Monitoring and Maintenance

📌 Suggested Projects

Beginner Projects

- Exploratory Data Analysis on Titanic Dataset
- Predicting House Prices
- Movie Recommendation System

Intermediate Projects

- Credit Card Fraud Detection
- Customer Segmentation using Clustering
- Sentiment Analysis of Tweets

Advanced Projects

- Real-time Object Detection using CNN
- Stock Price Prediction
- Resume Parser with NLP

- Time Series Forecasting for Electricity Demand
