🎯 Machine Learning for Data Science – Full Roadmap

🩵 1. Fundamentals (Before ML)

- **Mathematics Basics:**
  - Linear Algebra (vectors, matrices)
  - Statistics (mean, median, variance, correlation)
  - Probability (conditional, Bayes theorem)
- **Python Libraries:**
  - NumPy, Pandas, Matplotlib, Seaborn
  - Scikit-learn (ML library)

💙 2. Data Preprocessing

- Data Cleaning (missing values, duplicates)
- Encoding (Label, One-Hot)
- Feature Scaling (Standardization, Normalization)
- Feature Engineering (new feature creation)
- Splitting data: train_test_split()

💜 3. Supervised Learning (With Labels)

🔹 **Regression Algorithms (for continuous output):**

- Linear Regression
- Polynomial Regression
- Ridge & Lasso Regression
- Decision Tree Regressor
- Random Forest Regressor

🔹 **Classification Algorithms (for categories):**

- Logistic Regression
- K-Nearest Neighbors (KNN)
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine (SVM)

- Naive Bayes

💚 4. Unsupervised Learning (Without Labels)

- Clustering:
  - K-Means
  - Hierarchical Clustering
  - DBSCAN
- Dimensionality Reduction:
  - PCA (Principal Component Analysis)

💛 5. Model Evaluation & Validation

- Metrics for Regression → MSE, RMSE, R^2
- Metrics for Classification → Accuracy, Precision, Recall, F1-score, ROC-AUC
- Cross-validation (KFold, GridSearchCV)
- Confusion Matrix & Heatmap

🧠 6. Feature Selection & Optimization

- Feature Importance
- Regularization (L1, L2)
- Hyperparameter Tuning
- Grid Search / Random Search

🧩 7. Advanced Machine Learning

- Ensemble Methods: Bagging, Boosting (XGBoost, AdaBoost, LightGBM)
- Time Series Forecasting (ARIMA, Prophet)
- Recommendation Systems (Collaborative Filtering)