ML Model Development and Deployment Lifecycle

==========

1. Problem Definition

- Role & Responsibility: Understand the business problem, define objectives, success metrics.

- Tools: Jira, Confluence, Documentation tools.

2. Data Collection

- Role: Collect raw data from databases, APIs, logs.

- Tools: SQL, Python, APIs, Airflow, Kafka.

3. Data Loading & Management (Large Data)

- Role: Efficiently load and manage big datasets.

- Tools: Pandas, Dask, PySpark, BigQuery, Snowflake.

4. Exploratory Data Analysis (EDA)

- Role: Explore data patterns and insights.

- Tools: Pandas, Matplotlib, Seaborn, Plotly.

5. Data Preprocessing

- Role: Cleaning, encoding, outlier removal, missing value handling.

- Tools: Scikit-learn, Pandas, Imbalanced-learn.

6. Feature Engineering & Selection

- Role: Create and select optimal features.

- Tools: Scikit-learn, SHAP, LIME, FeatureTools.

7. Data Scaling & Normalization

- Role: Normalize features for model performance.

- Tools: StandardScaler, MinMaxScaler (Scikit-learn).

8. Data Splitting

- Role: Train/validation/test split.

- Tools: train_test_split, K-Fold, Stratified splits.

9. Handling Class Imbalance

- Role: Improve model fairness and accuracy.

- Tools: SMOTE, Oversampling, Class weights.

10. Model Selection & Hyperparameter Tuning

- Role: Compare models and optimize.

- Tools: Scikit-learn, XGBoost, Optuna, GridSearchCV.

11. Model Evaluation

- Role: Evaluate model accuracy and generalization.

- Tools: Accuracy, F1, Recall, ROC-AUC.

12. Underfitting & Overfitting Handling

- Role: Detect and resolve model misbehavior.

- Tools: Learning curves, Regularization.

13. Final Model Saving & Business Validation

- Role: Save the best model and validate KPIs.

- Tools: Joblib, Pickle, MLflow.

---------- MODEL DEVELOPMENT COMPLETE ----------

Deployment Phase

==========

1. Build Modular ML Pipelines

- Role: Automate loading, preprocessing, and training.

- Tools: Python, ML pipelines, Airflow.

2. FastAPI Endpoints for Model Serving

- Role: Expose API for predictions.

- Tools: FastAPI, Uvicorn, Pydantic.

3. CI/CD/CT Automation

- Role: Automate testing, building, and deployment.

- Tools: GitHub Actions, Jenkins, Kubeflow, GitLab CI.

4. Docker Containerization

- Role: Package everything in isolated environment.

- Tools: Docker, Dockerfile.

5. Deployment Platform Selection

- Role: Deploy model in production.

- Tools: GCP (Cloud Run, GKE, Vertex AI), AWS Sagemaker, Azure ML.

6. Monitoring, Tracking & Retraining

- Role: Monitor data drift and trigger retraining.

- Tools: Prometheus, Grafana, EvidentlyAI, MLflow.