🔁 Analytical Fraud Model Life Cycle

Here's a standard life cycle, broken down into phases:

---

1 Problem Definition

- **Goal:** Define what type of fraud you're targeting (e.g., credit card fraud, insurance fraud, identity theft).
- **Questions:**
  - What does fraud look like in this domain?
  - What business rules already exist?
  - What is the cost of false positives vs. false negatives?

2 Data Collection

- **Sources:**
  - Transactional data
  - Customer profiles
  - System logs (IP, device ID, geo)
  - Historical fraud labels
- **Important:** Ensure data privacy and security compliance.

---

3 Data Preparation & Feature Engineering

- **Clean & preprocess** data (handle missing values, normalize formats).
- **Create features** such as:
  - Time since last transaction
  - Transaction velocity
  - Distance between IP and billing address
  - Historical fraud frequency per user

4 Model Selection & Training

- **Algorithms Used:**
  - Logistic Regression
  - Decision Trees / Random Forest
  - Gradient Boosting (XGBoost)
  - Neural Networks
  - Isolation Forest / Autoencoders (for anomaly detection)
- **Training:** Use historical fraud-labeled data to train the model.

Model Evaluation in fraud detection focuses on metrics suited for imbalanced data.

- **Precision** measures how many predicted fraud cases were actually fraud.
- **Recall** shows how many actual fraud cases were correctly identified.
- **F1-score** balances precision and recall, useful when both false positives and false negatives matter.
- **AUC-ROC** evaluates the model's ability to distinguish between fraud and non-fraud across thresholds.
- **Confusion matrix** gives a snapshot of correct and incorrect predictions.
- **Cross-validation** helps ensure the model generalizes well, while **tuning** optimizes its performance without overfitting.

6 Model Deployment

- **Integration:** Embed into fraud monitoring system or API layer.
- **Real-time or batch scoring** depending on the use case.
- **Assign traffic light labels** (🔴 High Risk, 🟡 Medium, 🟢 Low) based on fraud score thresholds.

---

7 Monitoring & Feedback Loop

- **Drift Detection:** Fraud patterns evolve—monitor for performance decay.
- **Feedback Loop:** Use confirmed fraud cases to retrain the model.
- **Threshold Tuning:** Adjust decision thresholds to balance alert volume and accuracy.

---

8 Model Governance & Audit

- **Documentation:** Ensure all stages are documented.
- **Compliance Checks:** Make sure the model adheres to legal and ethical standards.
- **Version Control:** Track updates, training data, and parameters used.