# AI/ML Use-Case: Fraud Detection

## Selection Criteria

**Use-Case Chosen: Insurance Fraud Detection Using Machine Learning**  
Insurance fraud is a significant issue in Kenya, affecting both private insurers and public schemes. Fraudulent activities include false medical claims, staged accidents, and inflated property damage reports.  
**Why This Use-Case:**  
**1. High Financial Impact**: Fraudulent claims significantly affect profitability and increase premiums for customers and financial losses to organizations.  
**2. Regulatory Pressure**: The Insurance Regulatory Authority (IRA) mandates robust fraud detection mechanisms.  
**3. Data Availability**: Insurance companies collect and maintain extensive databases of structured and unstructured data like claim history, customer profiles, and incident reports which are deal for ML application development.   
**4. Feasibility:** ML models can automate the detection of anomalies and patterns that are difficult for human auditors or claim officers to spot.

**5. Scalability**: Once trained, models can be deployed across multiple product lines with little or minimal adjustments.

## Business Value

Anticipated Commercial Impact:  
- **Cost Reduction**: Early detection of fraudulent claims can save millions of Shillings annually by preventing payouts on illegitimate claims.   
- **Operational Efficiency**: Automating fraud detection reduces manual workload for claims investigators allowing them to focus on high-risk cases.  
- **Customer Trust**: Minimizing fraud helps maintain fair pricing and enhances customer satisfaction.  
- **Regulatory Compliance**: Supports adherence to regulators (IRA) guidelines and improves audit readiness and compliance with anti-fraud regulations.   
**Risk Mitigation Benefits:**  
- **Real-Time Alerts**: Machine Learning models can flag suspicious claims instantly, reducing the window for fraudulent activity.  
- **Pattern Recognition**: Detects emerging fraud schemes using historical data.  
- **Reduced False Positives**: Advanced models minimize the risk of incorrectly flagging legitimate claims, preserving customer relationships.

## Technical Considerations

**Data Requirements:**  
- **Structured Data**: Claim details, customer demographics, NHIF records, policy information.  
- **Unstructured Data**: Text from claim descriptions, adjuster notes, scanned documents.  
- **External Data**: Public records, weather data, social media that may be used to validate incident claims.  
**Modeling Approach:**  
- **Preprocessing**: Data cleaning, feature engineering, and text vectorization.  
- **Algorithms**: Logistic Regression, Random Forest, XGBoost for supervised learning; Isolation Forest and Autoencoders for anomaly detection in unlabeled data.  
- **Evaluation Metrics**: Precision, Recall, F1-score, ROC-AUC to balance detection accuracy and false positives.  
**Deployment Strategy:**  
- **Integration**: Embed model into claims processing systems for real-time scoring.  
- **Monitoring**: Continuous model performance tracking and retraining with new data.  
- **Explainability**: Use SHAP or LIME for model transparency and interpretation of model decisions and support of audit trails.