# Anomaly Detection Model Documentation

## Introduction

The Anomaly Detection Model is an AI-powered system designed to identify fraudulent or unusual transactions in financial datasets. It leverages an Isolation Forest algorithm for anomaly detection and integrates SHAP (SHapley Additive exPlanations) for interpretability. Additionally, an LLM (Large Language Model) is used to generate business-friendly insights into flagged anomalies.

## Demo URLs

• API Endpoint: http://<IP>/predict  
• Documentation: http://<IP>/docs

## Inspiration

Financial fraud is a growing concern in the digital economy. Traditional rule-based systems struggle to detect sophisticated fraudulent patterns. This model was inspired by the need for an explainable, scalable, and AI-driven anomaly detection system.

## What It Does

• Detects anomalies in financial transactions using machine learning.  
• Provides feature-based explanations for anomalies.  
• Uses LLM-generated insights to help interpret anomalies.  
• Offers a confidence score to quantify anomaly severity.

## How We Built It

1. \*\*Data Collection & Preprocessing\*\*: Gathered financial transaction datasets and cleaned/normalized numerical features.  
2. \*\*Model Training\*\*: Trained an Isolation Forest model and used SHAP for explanations.  
3. \*\*LLM Integration\*\*: Integrated a lightweight transformer model for text-based insights.  
4. \*\*Deployment\*\*: Packaged the model as an API using FastAPI and optimized for scalability.

## Challenges We Faced

⚠️ Data Quality Issues - Missing or imbalanced data required preprocessing.  
⚠️ False Positives/Negatives - Fine-tuning required to improve accuracy.  
⚠️ LLM Response Time - Optimization needed for real-time responses.  
⚠️ Model Interpretability - Integrated SHAP for explainability.

## Model Architecture

### Components:  
• \*\*Data Preprocessing\*\*: Converts raw input data into structured numerical format.  
• \*\*Isolation Forest\*\*: Detects anomalies based on learned patterns.  
• \*\*SHAP Explanation\*\*: Provides feature attribution for anomalies.  
• \*\*LLM Integration\*\*: Generates natural language insights.

Architecture diagram is available separately and can be manually inserted into this document.

## Installation & Setup

### Prerequisites:  
• Python 3.8+  
• pip  
• GPU (Optional, recommended for LLM processing)

### Installation Steps:  
```bash  
# Clone the repository  
git clone https://github.com/your-repo/anomaly-detection  
cd anomaly-detection  
# Create a virtual environment  
python -m venv venv  
source venv/bin/activate # On Windows use `venv\Scripts\activate`  
# Install dependencies  
pip install -r requirements.txt  
# Run the API  
uvicorn main:app --reload  
```

## Tech Stack

• \*\*Programming Language\*\*: Python  
• \*\*Framework\*\*: FastAPI  
• \*\*Machine Learning\*\*: Scikit-learn (Isolation Forest), SHAP  
• \*\*Natural Language Processing\*\*: Hugging Face Transformers (LLM)  
• \*\*Deployment\*\*: Uvicorn, Docker

## Advantages

✅ Fast Processing - Uses optimized Isolation Forest with parallel processing.  
✅ Explainability - SHAP values explain anomaly sources.  
✅ AI-Driven Insights - LLM provides business-friendly anomaly explanations.  
✅ Scalable - Supports large datasets and parallel execution.  
✅ Interpretable Confidence Scores - Easy-to-understand anomaly scoring.

## Future Scope

🔹 \*\*LLM Expansion\*\*: Integrate larger models (e.g., GPT-4, Gemini) for deeper anomaly explanations.  
🔹 \*\*Self-Learning System\*\*: Implement active learning where the model continuously improves based on feedback.  
🔹 \*\*Hybrid Approach\*\*: Combine Isolation Forest with deep learning techniques like autoencoders.  
🔹 \*\*Real-Time Processing\*\*: Adapt to streaming data for instant anomaly detection.  
🔹 \*\*Industry-Specific Customization\*\*: Tailor the model for sectors like banking, cybersecurity, and healthcare.

## Team

Anil Kopparapu

CH MaruthiRam

Gonnabattula, Shanmukha R

Venkateswara Rao Mandava

Naga Raj Cherukupally

## Conclusion

This anomaly detection model is designed for financial transactions but can be extended to other industries. By integrating explainable AI and LLMs, it enhances fraud detection and decision-making. Future improvements will focus on scalability, accuracy, and deeper AI insights.  
  
🚀 \*\*Start detecting anomalies today!\*\*