Automating Regulatory Reporting and Data Profiling in Finance sector using GenAI and ML

TRANSFORMING COMPLIANCE THROUGH INTELLIGENT DATA PROFILING
BY
TEAM NAME : MAAYUZZ
TEAM CAPTAIN : SYED MAAIZ SYED SHABBEER BASHA
TEAM MEMBERS : SYED MAAIZ SYED SHABBEER BASHA

Problem Statement:

Regulatory reporting in finance requires compiling vast amounts of data

Data profiling ensures alignment with regulatory requirements

Traditional approach: Manual rule definition based on complex documents

Challenge: Time-consuming, error-prone, and difficult to maintain

Need for an automated solution that can adapt to changing regulations

Solution Overview:

End-to-end pipeline to automate regulatory compliance reporting

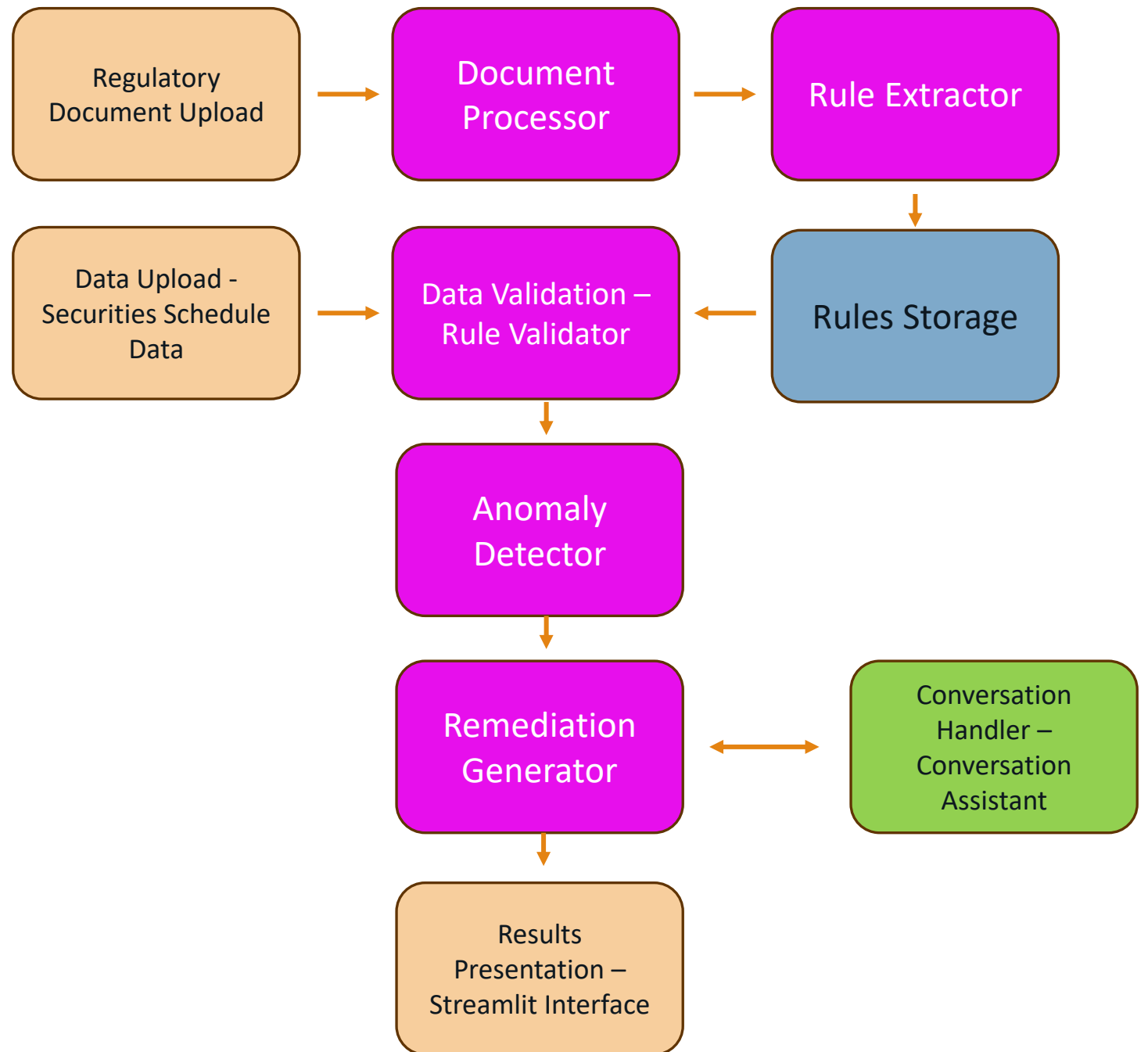Uses GenAI to interpret regulatory documents and extract validation rules

Applies both rule-based validation and ML-based anomaly detection

Provides remediation suggestions for non-compliant data

Features an intuitive interface with conversational assistance

Solution Architecture:

Regulatory Document Upload → Document Processor → Rule Extractor

Rule Extractor → Rules Storage

Data Upload - Securities Schedule Data → Data Validation – Rule Validator ← Rules Storage

Data Validation – Rule Validator → Anomaly Detector → Remediation Generator

Remediation Generator ↔ Conversation Handler – Conversation Assistant

Remediation Generator → Results Presentation – Streamlit Interface

Working Demo

Technologies Used:

LLM Integration: Hugging Face Inference API (Mistral – 7B instruct V0.2)

Machine Learning: scikit-learn for anomaly detection (Isolation Forest)

Data Processing: Pandas for data manipulation

UI Framework: Streamlit for interactive interface

Storage: JSON-based rule persistence

Testing: Python unittest framework for component testing