**Monitoring Plan for Acuant Fraud Detection System**

**1. What is Acuant?**

**Acuant** is a third-party, cloud-based, black-box solution integrated within the bank’s digital onboarding platform, specifically for **deposit product applicants**.  
It performs **real-time document verification** and **tampering detection**:

* **Document Authentication:**  
  Validates official documents (Driver’s License, Passport, Birth Certificate, Digital Identity Token) uploaded by applicants.
* **Tampering Detection:**
  + **Photo Tampering Detection:** Checks if the photo section of a document has been fraudulently substituted (e.g., cut-paste manipulation).
  + **Physical Document Presence Detection:** Ensures the ID shown is a real, physically-present document — not a screen capture or replayed video.
  + **Text Tampering Detection:** Detects whether textual fields like Name or Date of Birth have been altered on the document.

✅ The Acuant service is **embedded** into the digital onboarding flow.  
✅ It is **triggered automatically** — the system **sends applicant data to Acuant**, and Acuant **returns an instant decision** (Accept / Reject) along with tampering flags.

✅ **Important:**  
We **do not have access to Acuant’s internal model** (proprietary logic). Monitoring must rely purely on **input-output observations**.

**🔹 2. What We Intend to Do (Our Monitoring Strategy)**

Since we cannot validate the model internally, the monitoring plan will focus on:

**a) Operational Monitoring**

* Was Acuant triggered for every applicant as expected?
* Was the Acuant API response successful (no technical failure)?
* Was the response time acceptable?

**b) Output Monitoring**

* What % of applicants are flagged for tampering (Photo, Text, Physical Presence)?
* Are there sudden spikes or drops in tampering rates (model drift or technical issue)?

**c) Proxy Performance Monitoring**

* How many applicants cleared by Acuant later turn out to be fraud?
* Are false negatives increasing?

**d) Input Data Quality Monitoring**

* Are the documents uploaded by applicants in good quality (clear, not blurred)?
* Are there failures during document upload?

**e) Thresholds and Alerts**

* Set normal baseline ranges for each metric.
* Trigger **Amber/Red alerts** when thresholds are breached.

**🔹 3. Data Points to Request from Data Warehouse**

| **Column Name** | **Purpose** |
| --- | --- |
| application\_id | Unique ID for each applicant |
| application\_date | For daily/weekly volume tracking |
| product\_type | Helps to segment monitoring if needed |
| acuant\_api\_triggered\_flag | 1 if Acuant was triggered for applicant |
| acuant\_response\_status | Success or Failure of Acuant response |
| acuant\_response\_time\_sec | Time from request to response in seconds |
| acuant\_final\_decision | Accept / Reject decision by Acuant |
| photo\_tampering\_flag | 1 if photo tampering detected |
| text\_tampering\_flag | 1 if text tampering detected |
| physical\_presence\_flag | 1 if physical document spoof detected |
| manual\_review\_required\_flag | 1 if applicant was manually reviewed post-Acuant |
| final\_kyc\_decision | Approved / Rejected by the bank (final onboarding decision) |
| document\_upload\_quality\_flag | 1 if document upload was blurry or unreadable |
| fraud\_case\_reported\_flag | 1 if applicant later found to be fraud |

**🔹 4. Monitoring Metrics and Calculations**

| **Metric** | **Calculation** | **Industry Benchmark Example** |
| --- | --- | --- |
| **Acuant Utilization Rate** | Successful Acuant Responses / Applicants | > 98% |
| **Average Response Time** | Avg(acuant\_response\_time\_sec) | < 2 seconds |
| **Photo Tampering Rate** | Sum(photo\_tampering\_flag) / Successful Applicants | ~1% typical |
| **Text Tampering Rate** | Sum(text\_tampering\_flag) / Successful Applicants | ~0.5% typical |
| **Physical Presence Issue Rate** | Sum(physical\_presence\_flag) / Successful Applicants | ~0.2%–0.5% |
| **Manual Review Escalation Rate** | Sum(manual\_review\_required\_flag) / Total Applicants | <2% |
| **False Negative Proxy Rate** | Fraud Cases Missed / Applicants Cleared by Acuant | <0.1% |
| **Poor Document Upload Rate** | Poor Uploads / Total Applicants | <1.5% |

**🔹 5. Example with Hypothetical Numbers**

Suppose in 1 day:

* 10,000 applicants
* 9,950 successfully processed by Acuant
* 100 flagged for photo tampering
* 50 flagged for text tampering
* 25 flagged for physical presence issues
* 5 fraud cases later reported from applicants Acuant accepted

**Metrics would be:**

* Acuant Utilization Rate = 9,950/10,000 = **99.5%** ✅
* Photo Tampering Rate = 100/9,950 ≈ **1%** ✅
* Text Tampering Rate = 50/9,950 ≈ **0.5%** ✅
* Physical Presence Issue Rate = 25/9,950 ≈ **0.25%** ✅
* False Negative Proxy Rate = 5/9,950 ≈ **0.05%** ✅

*All numbers within healthy, expected bands.*

**🔹 6. Thresholds and Alert Logic**

| **Metric** | **Trigger (Amber)** | **Trigger (Red)** |
| --- | --- | --- |
| Acuant Utilization Rate | < 97% for 1 day | < 95% |
| Photo Tampering Rate | Drop/spike by ±50% from baseline for 2 days | Drop/spike by ±70% |
| Response Time | >3 seconds for 1 day | >5 seconds |
| Document Upload Failures | >2% for 2 consecutive days | >3% |