How HabileLabs Helped Utkarsh Small Finance Bank Mask 2.8 Lakh Aadhaar Documents in One Week



Client: Utkarsh Small Finance Bank
Industry: Banking & Financial Services

SOLUTIONS
AI-Powered Document Processing, Aadhaar Masking, OCR, Dockerized Offline Deployment

Introduction

Utkarsh Small Finance Bank is an RBI-licensed small finance bank serving a large and growing customer base across the country. As part of its regulatory obligations, the bank was required to comply with the Reserve Bank of India's data masking regulations for Aadhaar documents, a mandate that applied to its repository of over 2,80,000 Aadhaar documents stored within its Document Management System.

Before masking could begin at scale, the bank needed clarity on the current state of its repository: specifically, which documents had already undergone masking and which still required processing. Utkarsh Small Finance Bank engaged HabileLabs to design and deploy a solution that could deliver this clarity, and then complete the full masking workflow, accurately, securely, and within the compliance window.

The Challenges

Utkarsh Small Finance Bank's path to RBI compliance involved several interconnected challenges, each of which shaped the requirements for any viable solution.

- 1. Identifying Pre-Masked Files — The bank's document repository included files that had already been partially or fully masked through earlier processes, alongside those that still required masking. Without a reliable method to distinguish between the two, the bank risked either redundant processing of compliant documents or gaps in coverage. Establishing an accurate inventory of masking status across 2,80,000 files was a necessary first step before any masking activity could begin.
- 2. Data Security and Compliance in an Offline Environment — The bank's security policy was unambiguous: no data could leave the building. Any processing solution was required to operate within a completely offline environment, with no internet access permitted at any point. This ruled out cloud-based or externally hosted processing options and placed strict constraints on the solution architecture from the outset.
- 3. Limited Training Data for the AI Model — Building an AI model capable of accurate Aadhaar detection and masking typically requires substantial training data. Utkarsh was able to provide 200 sample files, a dataset that demanded a more resourceful approach to model development if the solution was to achieve the accuracy required for regulatory compliance.

- 4. Processing Over 2,80,000 Documents Within a Tight Timeframe — The volume of documents, combined with the strict timeline for RBI compliance, required a processing solution optimised for speed. Completing the masking operation without disrupting core banking operations added a further constraint: throughput had to be high, and the process had to run reliably from start to finish.

The Solution

HabileLabs built an end-to-end AI masking pipeline designed around Utkarsh's specific constraints, offline operation, limited training data, large document volume, and a defined compliance deadline. Each component of the solution addressed a specific challenge in the bank's situation.

1. **Intelligent Identification of Pre-Masked Files**
   HabileLabs used AI and OCR to sequentially analyse the full document repository, accurately identifying and reporting files that were already masked. This step ensured that only documents genuinely requiring processing entered the active masking pipeline, avoiding redundant work and improving overall efficiency.

2. **Robust AI Model Development**
   Using advanced data augmentation techniques, HabileLabs built highly accurate models for Aadhaar number detection and QR code identification from the 200 sample files available. The augmentation approach expanded the effective training dataset, enabling reliable model performance despite the limited starting point.

3. **Comprehensive Compliance Management**
   The solution covered the complete compliance workflow, from identifying pre-masked files through to the full implementation of required masking protocols. Audit logging and detailed reporting were included as part of the delivery, giving the bank a documentable record of processing outcomes across the entire repository.

4. **Secure and Compliant Dockerized Deployment**
   The AI masking pipeline was containerised with Docker and deployed within the bank's internal infrastructure. No data left the bank's offline environment at any stage. The deployment architecture was built to meet the bank's security requirements without compromise.

5. **High-Speed Parallel Processing on a 16-Core Server**
   To handle the volume of 2,80,000 documents within the compliance window, HabileLabs deployed the pipeline on a 16-core server

Business Impact

- **99% masking accuracy**
  The solution delivered the accuracy required for RBI compliance across the full 2.8 lakh document repository, significantly reducing the risk of regulatory penalties.

- **2.8 lakh documents processed in one week**
  High-speed parallel processing enabled completion of the full masking operation within the compliance window, with no disruption to core banking operations.

- **Reduced operational costs and resource allocation**
  By accurately identifying pre-masked files upfront, the solution avoided redundant processing across a portion of the repository, delivering direct savings in computational resources and time.

- **Strengthened data security and customer trust**
  Sensitive Aadhaar data for over 2,80,000 customers was processed entirely within the bank's offline environment. No data left the bank's infrastructure at any point.

- **Enhanced compliance and risk mitigation**
  Full masking coverage across the repository, combined with 99% accuracy, addressed the bank's RBI compliance obligations and reduced exposure to regulatory risk.

configured for parallel processing. This approach delivered the throughput required to complete the masking operation within one week, without disrupting live banking operations.

Audit-ready reporting across 2.8 lakh documents
Detailed processing reports gave the bank's compliance and governance teams a clear, documentable view of masking outcomes across the entire repository.

Why Utkarsh Small Finance Bank Chose HabileLabs

Proven AI capability under real-world data constraints

For a compliance project of this nature, model accuracy is non-negotiable, and most vendors need thousands of labelled samples to get there. Utkarsh needed a partner who could work effectively with what was available: 200 sample files. HabileLabs' use of data augmentation to build compliance-grade models from a constrained dataset was the deciding technical differentiator.

Domain fluency in Indian banking regulation

Utkarsh required a partner who already understood RBI data protection requirements and the security standards expected of regulated financial institutions, not one who would learn them during delivery. HabileLabs' existing knowledge of the Indian banking regulatory landscape meant the solution was designed for compliance from the first conversation.

A solution scope that matched the actual problem

The compliance requirement extended beyond masking itself, it began with understanding what had already been masked. Utkarsh chose HabileLabs because the engagement addressed the complete workflow: pre-mask identification, active masking, audit logging, and reporting. Nothing was left for the bank's team to stitch together separately.

The Conclusion

Utkarsh Small Finance Bank came to HabileLabs with a clear mandate: mask 2,80,000 Aadhaar documents to RBI standards, within a defined compliance window, inside a completely offline environment. The constraints were real, limited training data, strict security requirements, and a large document repository that needed to be processed without disrupting active banking operations.

The outcome was a 99% accurate Aadhaar masking operation completed in one week, with full audit reporting, no data leaving the bank's infrastructure, and core banking functions unaffected throughout.

For any regulated financial institution facing the same challenge, this engagement demonstrates what the right technology partner, with the right domain knowledge and engineering approach, can deliver when the stakes are real and the timeline is fixed.