



---

# From Paperwork Chaos to AI Precision: How teX.ai Unlocked Seamless Text Extraction for a Financial Service Firm

SUCCESS STORY

## Client Overview

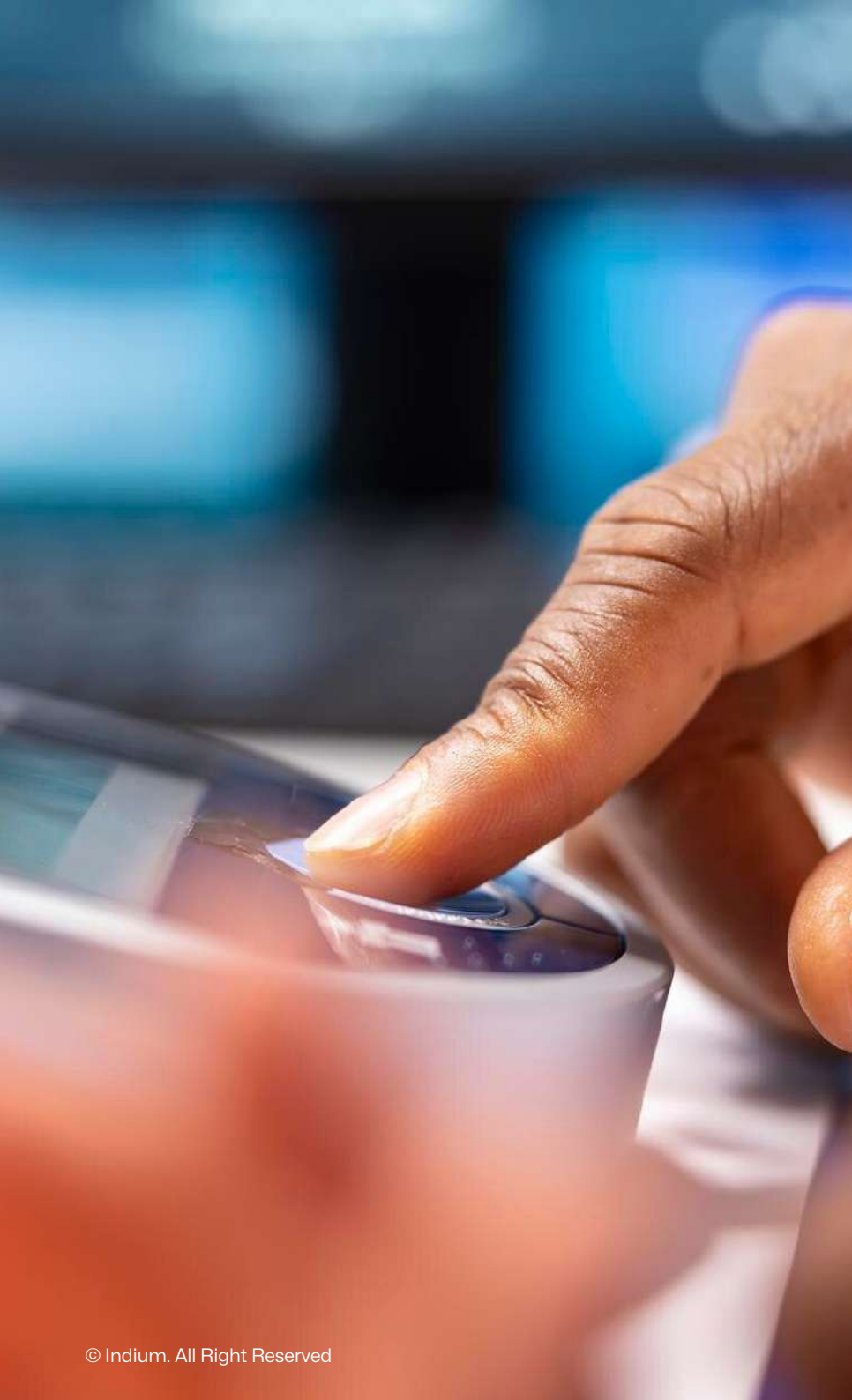
The client is a fast-growing financial firm that has embraced digital transformation to stay ahead in a competitive market. Specializing in rapid and accurate credit score assessments, they help banks evaluate customer creditworthiness with greater efficiency. By leveraging cutting-edge technology and digital platforms, the client aims to revolutionize the credit delivery process in India.

As part of their customer validation workflow, they process thousands of scanned bank statements to meet applicants' KYC (Know Your Customer) requirements. Automating and optimizing this process is crucial to maintaining speed, accuracy, and compliance in their operations.

## Extracting Insights, Not Just Data: The KYC Processing Challenge

To streamline KYC processing, the client needed a robust and scalable text extraction system capable of handling diverse document formats and high data volumes. The goal was to automate the extraction of critical customer details while ensuring accuracy and efficiency.

- ▶ **Two Formats, One Challenge:** The extraction process needed to handle two distinct document formats—scanned images and digital PDFs. Key information was spread across both structured tables (tabular data) and unstructured sections (peripheral data), requiring a precise and adaptable extraction approach.
- ▶ **Finding the Needle in the Haystack:** The extraction process had to identify and extract five key fields from both structured (tabular) and unstructured (peripheral) data sources within the statements. These fields included the **Account Holder's Name, Date, Bank Name, Transaction Details, and Address.**
- **Learning from Data:** The required extraction models were initially trained using a training corpus of 2000 bank statements.
- **Built to Scale, Ready for Growth:** With thousands of new documents arriving daily, the system needed to scale effortlessly, adapting to fluctuating workloads without compromising performance.



## Deciphering the Data: The Roadblocks in Streamlined KYC Processing

Extracting accurate information from bank statements was no simple task. The client's existing tools struggled with precision, consistency, and adaptability, leading to inefficiencies in their KYC processing. To streamline operations, they needed a solution that could overcome these key challenges.

### ▶ **Precision Matters: Accuracy Woes in Text Extraction**

The client was dissatisfied with the accuracy level of the text extraction output generated by their existing tools. They sought a solution that could provide higher accuracy in extracting data from bank statements.

### ▶ **One Field, Many Names: Tackling Inconsistent Terminology**

Different banks used varied terms for the same field. For instance, "Credit Amount" could appear as "Cr," "Credited," or "Amount Credited" across different statements. This inconsistency made it difficult to standardize and extract data effectively.

### ▶ **Lost in Layouts: Adapting to Changing Text Locations**

The location of the text to be extracted also varied across different banks. Some text resided within tables, while others were located outside tables. This variation required a flexible approach to locate and extract the required information accurately.

The client required a scalable, AI-driven approach to overcome these obstacles and ensure smooth, efficient KYC processing.

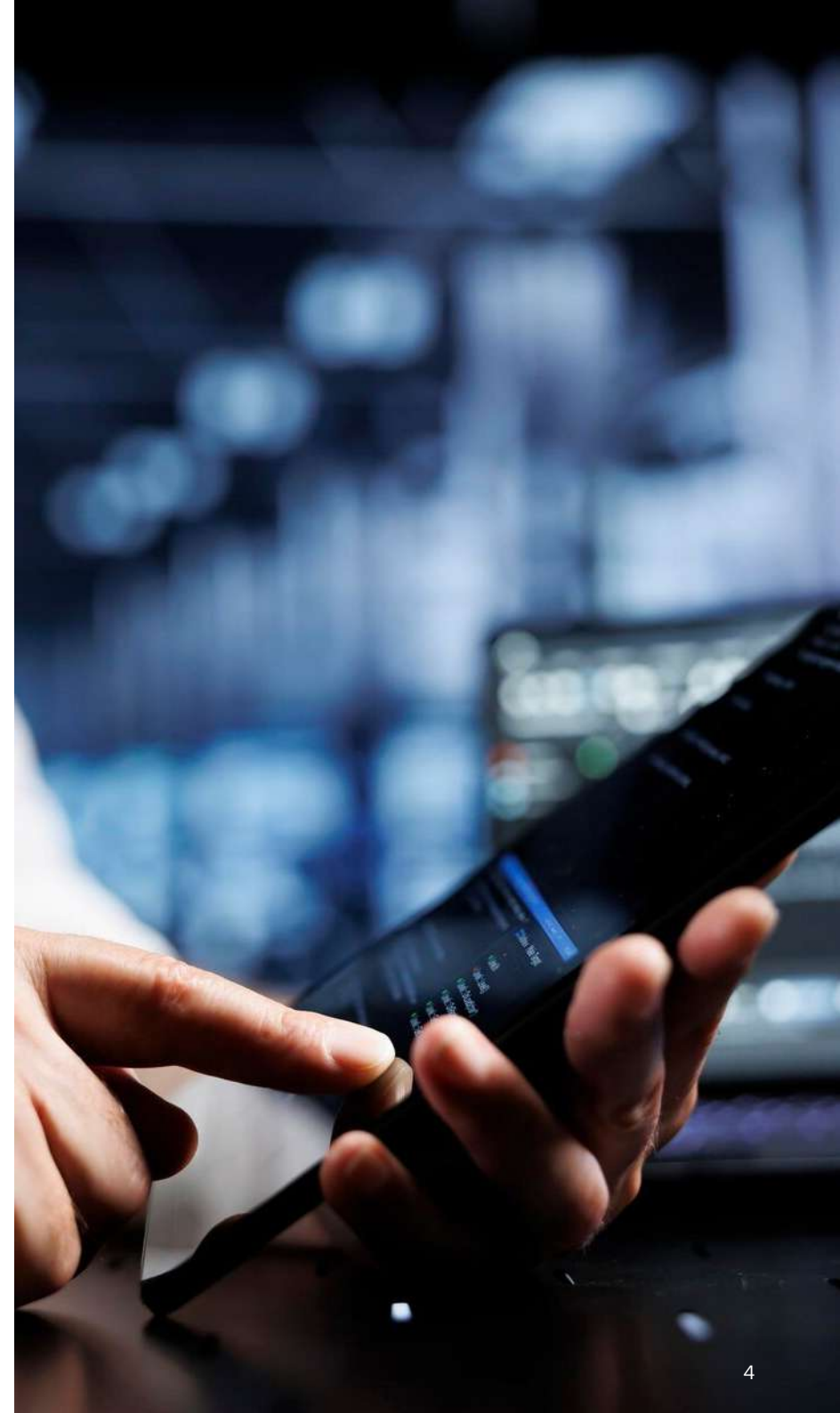
# Smart Extraction, Seamless KYC: How teX.ai Solved the Puzzle

A multi-layered AI-driven approach was implemented to overcome the challenges in KYC text extraction. By intelligently classifying documents, detecting tables, and extracting key data from both structured and unstructured formats, teX.ai ensured precision, efficiency, and scalability.

**Sorting the Chaos: Intelligent Bank Statement Classification:** The Bank statements were systematically classified based on bank type and document quality before extraction. This step streamlined the processing pipeline, applying the right extraction techniques to each statement.

**Finding Structure in the Unstructured: AI-Powered Table Detection:** A bounding box method supported by a Deep Learning Network was deployed to locate and extract text from tables. The teX.ai platform efficiently identified and processed tabular data, reducing manual intervention.

**Beyond Tables: Smarter Text Extraction from Unstructured Data:** Not all critical information resided in tables. A combined RNN model was used for text outside structured sections to intelligently capture and extract key data, ensuring no detail was missed.



# The teX.ai Approach: Implementing Intelligence in KYC Processing

To tackle the complexities of bank statement extraction, Indium implemented a structured, AI-driven approach using our in-house accelerator, teX.ai. By leveraging deep learning, NLP models, and intelligent classification techniques, teX.ai ensured accurate, scalable, and efficient KYC processing.

**Decoding Complexity: Understanding Bank Statement Variability:** The initial analysis uncovered 200 different types of bank statements within the provided dataset of 2,000 files, highlighting the need for a flexible and adaptable extraction solution.

**Quality First: Training the AI on High-Fidelity Data:** Out of the 2,000 bank statements, 830 high-quality files were selected to train the initial NLP model, ensuring superior text recognition and extraction accuracy.

**Cracking the Code: AI-Powered Table Detection:** A deep learning object detection model was used to locate tables in scanned bank statements. Bounding boxes were drawn to identify table structures, enabling precise extraction.

**Extracting the Essentials: Turning Tables into Usable Data:** Tabula and Camelot, integrated within teX.ai, were deployed to extract text from both digital and scanned PDFs. These tools ensured high-accuracy table content extraction.

## Extraction of Data Outside Tables (Peripheral Data)

Beyond the Grid: Extracting Peripheral Data with AI: Not all critical information was inside tables. A combined RNN model (LSTM-CRF) efficiently captured the required fields across different document formats for scattered data outside structured sections. This model effectively extracted the peripheral data regardless of the PDF document type.

### Output

**Ready-to-Use Insights: Structured JSON Output:** The extracted data was delivered as a JSON file, seamlessly integrating into the client's in-house systems for further processing and analysis.

### Set-Up

**Secure & Scalable: On-Premises Deployment:** teX.ai was deployed on-premises within the client's data center, ensuring full control and security over sensitive financial data. The system was built using Flask and Requests, facilitating a smooth and scalable data processing pipeline. Training was also provided to the in-house team for seamless adoption.

## teX.ai in Action: Revolutionizing KYC Processing with Measurable Impact

The implementation of teX.ai revolutionized the client's KYC processing, delivering speed, accuracy, and scalability like never before. Here's how it made a measurable difference:

- ▶ **80% Boost in Throughput: Faster Processing, Higher Efficiency**

With teX.ai, the processing time for a single bank statement dropped to under a minute, allowing the client to handle a significantly higher volume of applications per day. This resulted in an increase in the overall throughput, accelerating loan approvals and customer onboarding.

- ▶ **90% Accuracy: Reliable Data for Smarter Decisions**

The AI-driven text extraction achieved pronounced accuracy, ensuring clean, structured, and reliable data for further processing. As the system processed larger datasets, its accuracy improved, delivering even more precise and high-quality outputs over time.



## About Indium

Indium is an AI-driven digital engineering company that helps enterprises build, scale, and innovate with cutting-edge technology. We specialize in custom solutions, ensuring every engagement is tailored to business needs with a relentless customer-first approach.

Our expertise spans Generative AI, Product Engineering, Intelligent Automation, Data & AI, Quality Engineering, and Gaming, delivering high-impact solutions that drive real business impact.

With 5,000+ associates globally, we partner with Fortune 500, Global 2000, and leading technology firms across Financial Services, Healthcare, Manufacturing, Retail, and Technology—driving impact in North America, India, the UK, Singapore, Australia, and Japan to keep businesses ahead in an AI-first world.

### USA

Cupertino | Princeton  
Toll-free: +1-888-207-5969

### INDIA

Chennai | Bengaluru | Mumbai | Hyderabad | Pune  
Toll-free: 1800-123-1191

### UK

London  
Ph: +44 1420 300014

### SINGAPORE

Singapore  
Ph: +65 6812 7888

[www.indium.tech](http://www.indium.tech)



For Sales Inquiries  
[sales@indium.tech](mailto:sales@indium.tech)



For General Inquiries  
[info@indium.tech](mailto:info@indium.tech)

