

**Case Study | AI, NLP, RPA,
DATA ENGINEERING**

**US Based Company engaged in
providing Business Intelligence
(Tenders/Contracts)**



Problem

A US-based business intelligence company encountered inefficiencies in processing tenders, RFPs, & contracts.

Manually reading every email to find relevant correspondence and downloading attachments proved to be a tedious task. Extracting key information from attachments was time-consuming and prone to errors.

Varied email formats and the lack of consistent structure hindered rapid information retrieval.

Solution

Developed a two-part solution to streamline email and tender processing:

1. Email Listener

- Intelligent Table Detection:** Accurately identifying and delineating diverse table types within complex PDFs.
- Customizable Inclusion Criteria:** Users can define criteria based on subject lines or sender's email addresses to filter relevant emails
- Data Extraction and Archiving:** Adeptly extracts and archives essential information from emails, including sender details, date, recipient email address, CC list, email content, and attachments.

2. Intelligent Document Data Extraction & 2B. Contextual Tender Search

- Multi-Format Compatibility:** Ability to extract data from diverse formats including images (JPEG, PNG), digital documents (Word, Excel, PowerPoint), and scanned PDFs or documents.
- OCR Technology Integration:** Utilization of Optical Character Recognition (OCR) technology to accurately extract text from scanned documents and images.
- Data Extraction Accuracy:** High precision in extracting data fields from documents to minimize errors and inaccuracies in the extracted information.
- Document Indexing:** Processes tender documents for efficient search.
- Contextual Understanding:** NLP techniques or similar to grasp the meaning of search queries, going beyond simple keyword matching.



Results

- Significantly faster extraction of vital information from emails and their attachments improving efficiency.
- Increased accuracy in identifying essential email data, minimizing errors.
- Enhanced search capabilities for tender documents, enabling rapid and relevant information retrieval.



Technology Stack

- Mail Listener:** IMAP or POP3 protocol-based Libraries OCR: In-house OCR Engine
- Data Parsing:** Regular expressions combined with rule-based or machine-learning approaches.
- Search Engine:** Custom NLP integrations for contextual understanding.



Software Development

- Methodology:** Iterative development to refine information extraction accuracy and search relevance.
- Focus:** User-friendly interfaces for efficient email processing and tender search.



Before Metrics

Time-consuming manual extraction of email data. Potential for errors impacting the accuracy of business intelligence. Difficulty in quickly locating relevant information within tender documents.



After Metrics

Significant reduction in time spent on email processing. Improved accuracy of extracted email information (e.g., approaching 99%). Faster and more precise retrieval of relevant information from tender documents.