

**CASE STUDY**

# AI-Powered Document Extraction



# About the Client

The Client is one of the world's largest providers of high-technology interconnect, sensor and antenna solutions. Their products Enable the Electronics Revolution across virtually every end market including Automotive, Broadband Communications, Commercial Aerospace, Defense, Industrial, Information Technology and Data Communications, Mobile Devices and Mobile Networks.

## Employees

95,000 – 100,000

## Industry

Electronics Manufacturing

## Region

USA

# Operational Hurdles



### Manual and Time-Consuming Data Extraction

Technical teams manually reviewed over 1,000 PDF datasheets to extract specifications, dimensions, and ordering information — a process that was slow, repetitive, and resource-intensive.



### Inconsistent Document Structure

Datasheets lacked uniform formatting or standardized field labels, making it difficult to apply a single extraction logic. This variability increased the dependency on human interpretation and manual corrections.



### Complex Formatting and Technical Diagrams

Datasheets included nested tables, multilayered technical specifications, and complex diagrams that traditional extraction tools struggled to interpret accurately.

### High Error Risk and Scalability Challenges

Manual Excel population often resulted in inaccuracies in part numbers and dimensions. With 35+ new datasheets added annually, scaling the process became increasingly inefficient.

# Solution



## Intelligent Data Extraction with OCR

An automated data extraction engine built on PyTesseract OCR technology processes PDF datasheets, converting them into machine-readable text that accurately captures technical specifications, measurements, and order details.



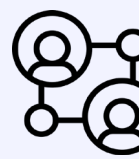
## AI-Driven Comprehension with LLMs

The system uses a Large Language Model (Google Gemini) to interpret complex document layouts, understand technical context, interpret nested tables and accurately identify key data points despite structural variations.



## Automated Permutation Generation

A dynamic logic engine generates all valid ordering number permutations, identifies variable placeholders, and applies business rules, automating a once manual and error-prone process.



## Structured Output and Validation

Using Python libraries like Pandas and OpenPyXL, the system maps data into a standardized Excel template. Automated validation flags anomalies for human review to ensure data accuracy and reliability.



## Scalable and Integrable System Architecture

Built on a scalable technology stack, the platform is designed to efficiently handle growing document volumes and integrates seamlessly with downstream systems like quotation engines and ERP modules.

# Impact

## Accelerated Processing

**Reduced manual extraction and Excel population time from hours per datasheet to minutes, dramatically improving turnaround time.**

## Enhanced Scalability

**Enabled continuous and consistent processing of new datasheets without additional manpower or process redesign.**

## Risk Mitigation Achieved

**Automated exception handling, robust permutation validation, and adaptive AI logic significantly reduced the risks of incorrect table mapping or part number generation.**

## Improved Accuracy and Reliability

**AI-based validation reduced extraction errors and eliminated manual entry mistakes, lowering error rates to near zero mistakes.**

## Greater Data Consistency

**Standardized data formatting across all templates improved reliability for reporting and analytics.**

## Future-Ready Foundation

**The structured outputs provide a foundation for seamless integration with downstream systems like quotation engines or ERP modules.**

# Decoding Difinity: Our World and What We Do

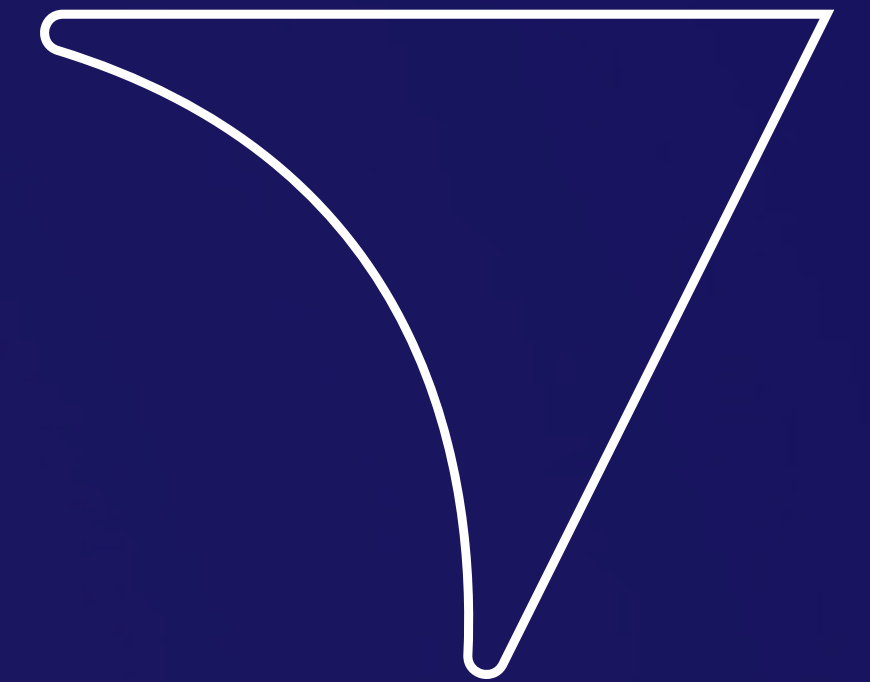
At Difinity Digital, we're not just a tech consulting firm; we're your innovation partners. We specialize in crafting transformative solutions that empower enterprises worldwide to redefine their technology engagement. Our mission is to drive sustainable growth by harnessing the power of intelligent automation, data analytics, and custom software development.

Our dedicated team of experts leverages cutting-edge technologies to streamline operations and boost efficiency across various sectors, including finance, accounting, banking, insurance, logistics, and software and digital platforms. From automating business processes with AI to enhancing KYC compliance and optimizing accounts receivable/payable workflows, our solutions are designed to elevate your business.

At Difinity Digital, we believe in revolutionizing industries and creating lasting impact. We're not just providing services; we're shaping the future of digital transformation.

+91 9995 099 503  
info@difinitydigital.com

**difinitydigital.com**



FOLLOW US ON

