

# MARRIAGE CERTIFICATE INDEXING

## Intelligent Data Extraction and Indexing



### SCOPE

The project had an implementation deadline of one week to achieve the following:

- index 30k pages per month (day forward)
- extract 45 fields using ONLY automation and no human validation

### INDUSTRY

Government civil services

### THEMES

OCR  
Indexing  
Data Extraction

### CHALLENGE

Our partner's client in the Government sector faced a significant challenge in efficiently indexing information from state marriage certificates. The client desired a fully automated solution without relying on labor-based validation. Previous attempts with Amazon Textract did not yield reliable results to support direct database feeds, making it challenging to process the substantial volume of data accurately.

### SOLUTION

To address the challenge, our team proposed a comprehensive solution leveraging advanced automation technologies and custom data processing algorithms. The primary components of the solution include:

1. Machine Learning Algorithm: Implementing a machine learning algorithm tailored to the specific characteristics of marriage certificates. This algorithm is trained to recognize and extract information accurately, adapting to variations in document formats and handwriting styles.
2. Custom Data Processing Pipeline: Developing a robust data processing pipeline that integrates seamlessly with the existing infrastructure. This pipeline incorporates the machine learning algorithm for automated extraction and validation of 45 fields from marriage certificates. The system is designed to handle any volume of documents efficiently, ensuring real-time processing of up to 30,000 pages per month.
3. Quality Assurance Mechanisms: Implementing built-in quality assurance mechanisms to ensure data accuracy without human intervention. The system is equipped with error-checking routines and validation steps to minimize false positives and negatives, ensuring a high level of confidence in the extracted information.

### BENEFITS



#### OPERATIONAL EFFICIENCY

The fully automated solution eliminated the need for a labor-intensive validation process, reducing the time required for indexing state marriage certificates from one week to near real-time processing. This resulted in a substantial increase in operational efficiency.



#### SAVINGS

By removing the dependency on manual validation and streamlining the process, the client achieved significant cost savings associated with human resources. The automated solution not only accelerated the workflow but also substantially improved data accuracy compared to earlier approaches.



#### SCALABILITY AND CONSISTENCY

The implemented solution showcased scalability to handle a volume of 30,000 pages per month seamlessly. Moreover, the consistency in data extraction across varied document formats and handwriting styles demonstrated the robustness of the system, ensuring reliable and accurate results consistently.