

BIRTH CERTIFICATE INDEXING

Intelligent Data Extraction and Indexing



SCOPE

The project had an implementation deadline of two weeks to achieve the following:

- index a backfile of 2 million pages
- extract 45 fields with high accuracy from birth certificates to feed the State database.

INDUSTRY

Government civil services

THEMES

OCR
Indexing
Data Extraction

CHALLENGE

The client, in its pursuit of digitizing birth certificates, was attempting to use Amazon Textract, but encountered a significant challenge with the inconsistency of the key-value pair output.

The inconsistencies manifested as misalignments, missing data points, and occasional misinterpretations of vital birth certificate details. The complexity of post-processing became apparent as manual interventions were required to align the extracted data accurately. This intricate and time-consuming post-processing not only impeded the efficiency of the digitization process but also raised concerns about the reliability of the extracted information. The challenge was further exacerbated by the sheer volume of birth certificates, with 2 million pages in the backfile, amplifying the need for a more streamlined and automated solution.

SOLUTION

To address the challenge, our approach involved integrating advanced algorithms and pattern recognition techniques to create a dynamic extraction pipeline. This pipeline adapted to varied birth certificate formats, correcting irregularities in real-time and minimizing the need for laborious post-processing.

Machine learning models were employed to enhance accuracy, while cloud-based infrastructure ensured seamless scalability. The result was a streamlined and automated data extraction system, successfully digitizing 2 million birth certificate pages within a two-week timeframe.

BENEFITS



RAPID IMPLEMENTATION WITHIN TWO WEEKS

Despite the complexity of processing 2 million pages of birth certificates, the solution was swiftly deployed within two weeks. This rapid turnaround exceeded client expectations, allowing them to achieve their digitization goals promptly.



HIGH-ACCURACY FIELD EXTRACTION

By leveraging the refined data extraction process, the project ensured a level of precision that obviated the need for any post-processing. This high accuracy not only saved resources but also enhanced the reliability of the birth certificate information provided to citizens through the digital portal.



SCALABLE SOLUTION WITH BROADER IMPLICATIONS

The successful implementation of this project has broader implications for similar data processing tasks within the government sector. The streamlined approach to field extraction showcased here can be scaled and adapted for other document types, offering a template for enhanced efficiency in various governmental processes.