

# DIPLOMA SCANNING

## Indexing and Intelligent Data Extraction



### SCOPE

The project accomplished the following objectives:

- Digitize and index 2.5K certificates per year (day-forward scanning).
- Digitize and index 30K certificates from the back file.
- Index five key fields per document: student name, national ID, course reference, date of certification, and date of certificate expiry.
- Create a windows-based searchable library for efficient certificate retrieval.
- 2 days to implement and deploy

### INDUSTRY

Education

### THEMES

OCR  
Indexing  
Data Extraction

### CHALLENGE

Our partner was approached by an educational college to address the challenges they faced in managing a large number of certificates granted annually. The existing volumes were too small for on-premises capture or traditional Document Management Systems (DMS). Therefore, they needed a solution that could digitize and index certificates efficiently, allowing for easy retrieval and follow-up on expired certifications.

### SOLUTION

The proposed solution entailed a comprehensive approach to address the challenges in certificate management. It began with the implementation of high-speed document scanning technology to capture both the day-forward certificates (2.5K per year) and the substantial volume of back-file certificates (30K). This digitization process eliminated the need for physical storage and enabled convenient access to certificates.

The solution integrated advanced ML and AI technology to extract essential details from each certificate, including student names, national IDs, course references, certification dates, and certificate expiry dates. Using this extracted information, the system precisely named or renamed the files and organized them chronologically by year.

This approach facilitated seamless and user-friendly certificate retrieval, where a windows-based searchable library was developed. This central repository allowed authorized personnel to quickly locate specific certificates by conducting searches based on student name, national ID, course reference, or certification date. The intuitive interface of the searchable library streamlined the certificate management process, saving time and effort for staff.

By utilizing the indexed data, the system facilitated the monitoring of certificate expiration dates by college personnel, enabling them to generate comprehensive reports or receive timely notifications concerning imminent or expired certifications. Consequently, the college gained the ability to promptly reach out to students, providing them with course refreshers or alternative educational offerings.

### BENEFITS



#### ENHANCED EFFICIENCY AND ACCESSIBILITY

By digitizing and indexing the certificates, the searchable database significantly improved the efficiency of certificate management. College staff no longer needed to manually sift through physical documents, as they could quickly search and retrieve certificates based on student name, national ID, course reference, or certification date. This streamlined process saved time and effort, allowing staff to focus on other essential tasks.



#### TIMELY FOLLOW-UP ALLOWING REVENUE GROWTH

The system's ability to track certificate expiry dates enabled the college to proactively follow up with students whose certifications were approaching or had already expired. This feature ensured that no student's certification went unnoticed or unattended. By promptly addressing expired certifications, the college was able to help students stay updated with their certifications, thereby enhancing their professional opportunities or provided them with course refreshers or alternative educational offerings.



#### SCALABILITY AND FUTURE GROWTH

As the number of certificates grows in the future, the searchable database can easily accommodate the increasing volume. This scalability ensures a sustainable solution that can adapt to the college's evolving needs and growth without compromising performance or data integrity.