Automating Certificate Classification, Parsing, Routing

Summary of Google Document AI for Delta AC

Overview

Document AI offers automation for classifying and extracting key information from regulatory certificates received via Gmail, enhancing efficiency in storing data related to our pesticide and fertilizer registration pdfs. The devil is *always* in the details, but below is the thesis.

Key Benefits and Features

I've linked to some timestamps in a 10-minute video that walks through a high-level demo of an example use case: parsing invoices automatically. This should be applied to our certificate pipeline.

1. **Automation of Certificate Processing** 00:00:58
   - Automatically extracts important data such as registration numbers, dates, and compliance details.
   - Speeds up the tracking of regulatory application lifecycles.
   - The video demonstrates (using invoices as an example) how Document AI can automate data extraction, which we can apply to our certificates to speed up the lifecycle tracking of government applications for fertilizer and pesticide products.

2. **Document Classification** 00:00:05
   - Helps Delta AC organize automatically recognize which certificates belong to which customer accounts/products.
   - this section of the video introduces the concept of document classification and data extraction, which is directly applicable to our regulatory certificate processing needs.

3. **Integration with Google Workspace** 00:02:26
   - Seamlessly processes certificates stored in Google Drive, fitting into Delta AC's existing infrastructure.

4. **Customizable Model Training** 00:05:17
   - Models can be trained to accurately handle Delta AC's specific certificate formats, improving accuracy. The more documents that get labeled, the better.

- The video demonstrates what it looks like to train Document AI on our regulatory certificate formats, ensuring it extracts and classifies relevant details, improving accuracy and reducing processing times.

5. **Exportable Data for Workflow Integration** 00:07:42
   - Extracted data can be exported via json (universal api format) and integrated into Delta AC's tracking systems for more efficient compliance monitoring.
   - The video shows how extracted data can be exported into spreadsheets or other systems, allowing us to track certificate statuses and integrate them into our existing workflows, making compliance tracking more efficient.

TLDR

Document AI's ability to extract and classify data from certificates would allow Delta AC to automate and streamline part of our regulatory workflow, reducing manual effort and increasing accuracy in handling pesticide and fertilizer registrations.