



# How do Intelligent Document Processing Frameworks work?



Intelligent Document Processing (IDP) is the process of intelligently capturing specific data and streamline document processing activities. Regardless of what kind of document is being processed, electronic, structured, semi-structured or unstructured, the goal of an IDP is to extract information.

*A recent study found that at-least 80% of documents in any organization is unstructured and people spend 60% of their time on average in manually processing those documents.*

## Key challenges in Document Processing

Here below are some of the key challenges that I see, an IDP framework should address. It is certainly a complex situation where the possibilities are infinite and to process them without exceptions is merely impossible. Hence, you will require complex machine learning models to work coherently to achieve the desired output.





## How IDP works?

IDP combines multiple AI technologies like Computer Vision, OCR, Natural Language Processing, Predictive Analytics and Machine Learning into a single highly cohesive solution. The secret is getting them all to work together just right, so you can automate data extraction from even your most complex documents with accuracy.

Intelligent Document Processing and data extraction must go through at-least four phases to process the documents. Typically, as below.



### 1. Pre-Processing

Documents in an organization typically can exist in varying levels of quality, which can impact the results of data extraction. So, you need to do some pre-processing work to improve the quality of the documents. This is done by applying techniques such as noise reduction, binarization, normalizing, and more.

### 2. Document Classification

The first step in the process is to classify what kind of document is being processed & also determine the start and end of the document. Documents shall contain multiple pages with different formats. Intelligent Document Classification shall use AI algorithms to automatically classify and separate multipage documents to pull out the relevant pages of information before extraction.

For example, a loan application typically includes a bunch of pages with documents such as a loan application, payment schedule, bank statements, ID documents, and more. These documents need to be identified and classified accurately.

### 3. Data Extraction

Once the document is classified, the next and most significant step in the process is to extract specific information from the document. This step shall involve OCR to digitize documents along with ML technologies to extract specific data. Typically, IDP solutions include a library of pretrained extraction models, which are pre-populated with the right fields for data extraction. In some cases, Data extraction is powered by pattern matching tools like Regex (Regular Expressions). These tools identify and extract specific data in the document and present them in an easily accessible, digital format.



## 4. Data Validation & Enrichment

Once data is extracted, it goes through a series of validation rules and AI-driven techniques to improve the extraction results. These could be predefined rules, character sets, language dictionaries, validated in an automated fashion and are further enhanced by using RPA techniques. A human in the loop can further validate the data, which allows the process to continuously learn and improve over time.

## 5. Data Export

Now that your data is ready and as a next step of the intelligent data processing, you would be interested to check out the ability to export the information to various business processes needs or workflows automatically. Maybe you wanted to use the output in an RPA process to automate the rest of the process, or you may want to interface it with another system through API integration. The options are endless.

## Summary

That said, I see Intelligent Document Processing as a key catalyst for digitization and transformation in modern organizations. Data plays a critical role in transformation. IDP shall disrupt the way you process data today and shall help you discover valuable insights. If you wonder where people fit in, they sit at the middle of this disruption. IDP augments the modern workforce by providing them a stream of valuable information in an automated fashion.

Alright that's all I had to write for today. I hope to see you next time soon. Bye!

Tags: [Document Processing](#), [AI](#), [ML](#), [Intelligent Automation](#), [IDP](#)