



Mortgage Automation

DOCUMENT CLASSIFICATION

LEVERAGING SMART LEARNING



OVERVIEW

WHAT IT IS ALL ABOUT

There are a myriad of factors to consider when it comes to mortgage processing. We all want to be more efficient with faster processing and less errors to book loans faster and increase turns. Being able to turn more loans within the year can dramatically increase profitability, the biggest contributor to success.

Mortgage document classification represents one of the hardest types of tasks for document classification automation. This eBook examines the available technologies and pairs them with those types of solutions that you need. In addition, this eBook provides insights into how best to question prospective vendors on what their offerings are and what they can do in terms of real capabilities.





Mortgage Automation Facing the Challenges



Mortgage document processing is a complex, time-consuming effort.



The problem extends beyond reduced costs and complexity; slower processing means slower capital realization.



A single loan file must be reviewed, document types noted, and individual documents separated from each other.

Even if you use OCR or other rules-based approaches, it is likely that the amount of effort and the associated cost with the identification and organization of key supporting documents create significant bottlenecks in your business processes. This is due to the underlying limitations of rules-based approaches and the typical unreliability of the system output, which requires 100% review of all data. In today's world of AI and digital transformation, organizations need the ability to **reduce manual efforts by greater than 70%**.

When you apply automation, you can sum up these tasks in three main activities. One is to classify or identify documents that are there. This is mandatory from a business process perspective as well as a compliance perspective and risk mitigation perspective. Second, you must go through the process of separating out the documents. Third, whether you use automation or manual labor, you need the information in these tasks to be as reliable and as accurate as possible.

So, when considering technologies to automate these three main activities, it is identifying documents and separating those documents out to make them discreet. Documents should not go into exceptions—which involve manual processes—if it's unnecessary. **Unfortunately, since today's systems require so much attention to configuration and tuning, which require expensive skill sets, most organizations remain stuck in a bad situation.**



Document Classification

Identify specific document types as fast as possible.



Document Separation

If a PDF has more than one document, locate it and create a separate file.



Document Reliability

High reliability of document classification without traditional trade-offs of expensive professional services.

Addressing the Challenges



To address document classification, there are three main approaches: (1) a rules-based classification approach; (2) automated classification focused on content or the text in a document; and (3) automated classification focused primarily on the visual elements. **A large percentage of organizations that are using automated classification are using a rules-based approach.** This is the process of taking a sample set of documents that we want to classify and having a person go through them to identify specific attributes of each document class. That can be encoded into a rule such as if a document has a board appraisal, then assign document class appraisal. Typically, rules-based approaches use text.

If we've got a few document classes, even 20 to 30 document classes, then it's probably not all that difficult for someone to go through and create rules for them, provided that the variants of those document classes isn't high. Variants within classes are different formats or different styles of a particular document class. Where rules-based classifiers fall short is when we introduce a high degree of variance within a specific document class or large number of document classes.

In a rules-based approach, the more classes we have to create, the more rules we have to create and test, which results in a large amount of manual effort. With more rules, we have more complexity, which breeds error. If we create a rule for a specific document class that does not account for all variances, then those variances go into exception.

Automated classifiers may offer a better alternative in cases such as mortgage processing that deal with high complexity and variance. Automated classifiers rely on machine learning. Machine learning algorithms automate the process of analyzing each sample document to discover what we call features or the attributes of each type of document to automatically create rule sets.

Machine learning systems can process enormous volumes of data, much more in less time than any human can do. Also, these systems collect more information and attributes more precisely and accurately. Literally hundreds if not thousands of hours of effort in creating very explicit rules can be translated into five or six hours of compute time. This is why many organizations are deploying automated classification. However, automated classifiers require the curation of samples. You need to have an adequate number of samples that are tagged in order to have the system learn. Most automated classifiers also are typically only text.

In loan processing, there are a lot of text heavy documents, but visual classification can be useful, too. There are at least 800+ naming conventions that equate to approximately 250 or so discreet or specific document types. In the case of document separation, you can implement a rules-based approach or an automated approach just for separation just in the same way you could do that for classification itself. Ultimately, using one approach for everything—rules-based approach, content-based automatic classifiers or visual automatic classifiers—you're not going to get the level of performance that you desire.

RULES-BASED
CLASSIFICATION
APPROACH

CONTENT-BASED
AUTOMATIC
CLASSIFIERS

VISUAL
AUTOMATIC
CLASSIFIERS





Automating Classification, Configuration & Tuning



Configuration and system tuning is a big cost of automation. One element is the upfront configuration or the “time to live” to be fully up and running in your processing environment. Second, tuning the system is necessary so that it achieves the level of performance that you expect over time. And third, processes must be adaptable to handle new or different types of documents or compensate for varying image quality that can impact the system.

Few companies actually plan adequately for this because it is easy to assume that once a system is in place that its accuracy will remain the same over time. Two main ingredients define what accuracy is and what we're trying to achieve: **the total number of documents that can be assigned to a particular class** and **the accuracy of those assignments**.

Humans can apply a class assignment to any document, but it is difficult to be right 100% of the time. Since a loan package often has 100 different documents, as humans, we may mistakenly assign the wrong classes for two of those documents. That's a 98% accuracy rate. **The percentage of all class assignments multiplied by the accuracy rate gets you the whole system performance.** That's why when claims are made of 99% accuracy, typically you have to ask “what do you mean by that?” “How do you calculate that?” This is because we need to consider the class assignments as well as the accuracy rate. It is entirely possible for a system to get 99% accuracy rate of class assignments for only 5% of your total number of documents. That level of performance won't achieve the real accuracy that you want.

At Parascript, our objective is to get higher than human accuracy by applying the right number of techniques and technologies. We combine automated classifiers and other machine learning techniques and let the system decide which one is best so that we're effectively using all the information on the page. Using another layer of machine learning, Parascript software automates the initial configuration that enables organizations to achieve high levels of performance without the traditional cost and complexity. The entire machine learning configuration is stored in a portable file instead of a cumbersome database. Ultimately, we focus on the outcome and often guarantee that performance over the lifetime of the system.

To handle updates to the system, which often impose significant costs to ensure that the system continues to perform well, Parascript software uses a “cascade classifier” that analyzes the output against the original, determines what the class assignment should be and automates optimization. The ultimate result is a much better level of optimization with a 95% to 98% accuracy rate on a significant amount of your documents. This is all without any manual effort. We turn the problems of configuration and optimization into a compute-time problem where you click a button, and everything is done for you. While everyone focuses on features, we focus on outcomes and the challenges required to achieve high levels of performance. We call this **Smart Learning**.



THE PARASCRIPT PARADIGM

Rather than relying on expensive professional services or time-constrained staff to select and implement technologies to create high-performance automation, Parascript has developed a system that emulates the decisions and skills of advanced data scientists.

Using applied machine learning that we call **Smart Learning**, the tasks of configuring, measuring, optimizing and maintaining document classification and separation are eliminated.

Techniques are automatically selected and deployed based upon what will achieve the best performance and Smart Learning includes dozens of different algorithms tuned to different tasks.

Unlike other systems that emphasize features and that require significant investment in time and skills, Parascript software allows you to focus on what really matters: **business transformation.**

Mortgage Automation

Document Classification

Today, software can be made smarter by the practical, pragmatic and intelligent application of the right technology married to the right problem. In practical terms, what this means is that Parascript uses multiple types of classifiers, and the system decides which one is best. Parascript smart learning is machine learning for documents that enables true unattended automation for your complex mortgage data. Parascript software analyzes data and the location of data to generate models to locate and reliably extract it. Parascript software constantly measures the results and adapts to those results. When the system encounters new document types or variances of documents during your production stream, Parascript software runs in the background. It takes that information and uses it to adjust and optimize the systems. The care and upkeep that is typically time consuming and expensive are now automated.

Downtime is expensive. So the thing to keep in mind is Parascript software largely contains its intelligence and configuration internally. This reduces the dependency on highly specialized technical resources and their availability. This also eliminates another point of potential human error. The system learns as a function of processing and validation in a feedback loop. Due to this feedback loop, the system eliminates errors over time even when new ones are introduced.

Ultimately, what matters is whether your system reliably classifies your documents and extracts the right data. Parascript software automates initial configuration for reliable straight through processing and ongoing adaptability of the system once you put it in into production. It's very common to see less than 10% of your overall volume consuming in an inordinate amount, sometimes up to 50% of your workflow resources, handling errors within that process. By reducing the overall processing time per loan, we hit the ledger as an asset must faster. This allows for significant increase in the number of loans processed per year.





CONTACT US TODAY



www.parascript.com



888.225.0169



info@parascript.com

