# Empowering fraud detection in banking

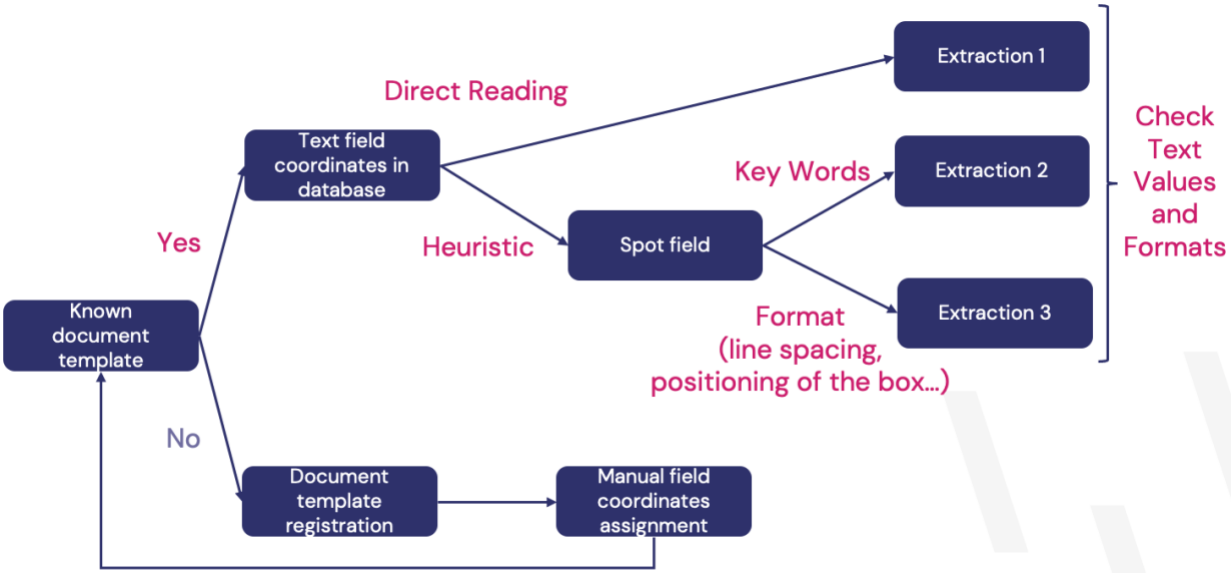
**Business objective**
Process Optimization
**Sector**
Bank and Insurance

Fraud encompasses counterfeit currency and security documents, as well as identity and travel documents, costing the global economy £3.2 trillion (roughly €3.8 trillion) annually. Also considered is document forgery, a common issue in banking, costing a lot of money and resources year after year. We created a solution to help the banking sector tackle this problem.

## Fraud detection technology overview

***Optical Character Recognition (OCR) - the classical technique***

Fraud detection on forged documents relies on Optical Characters Recognition (OCR). OCR checks the text content on the document and the format (for example - size, fonts, the position of the characters). Indeed, it is possible to do it on already registered document templates. The position on the text fields must be pre-defined to ensure the algorithm's proper focus and direct reading of the text. If the document is not registered, you can use a heuristic approach where the algorithm will try to detect the points of interest in the image. Then, it extracts the text and its related information. Of course, the performances are therefore less efficient. Here we illustrate the OCR technique:




Fraud detection with the OCR process

Because it does not need any training data, OCR is very useful. Ideally, you have one template to improve character recognition. However, this technique has limitations - if the document is an image and not a text document and if the forgery is very well done.

***Steganography - the graphical technique to detect non-obvious forgeries***

We can enhance fraud detection of forged documents with steganography. The idea behind steganography is that it hides some information behind what you directly see with your eye, reaching the pixel level of image-based documents to detect modifications. The manipulation for the forgery is creating alterations on the pixels around the forged text. Those manipulations are, for example :

- copy-pasting from inside the document (CPI)
- copy-pasting from outside the document (CPO)
- deleting one or more characters on an image (CUT)
- creating a text box to imitate some text (IMI)

Steganography looks for those background modifications to detect forgeries on documents. Applying specific filters on the image will highlight the modifications around the text. Each filter generates numerical output values, and we use those values to feed the classification algorithm. After the training phase, the algorithm can detect if there is a forgery (or not).

Steganography process

The algorithm's training depends on documents where forgery is well specified and labeled (i.e., wherein a value generated by one filter from an image is related to a forged or genuine document).

OCR cannot work appropriately on image documents (e.g., scanned documents), but steganography perfectly fits this task. We tested this technique on a dataset with payslips that detected 75% of fraud (you can find more information in this scientific article). Moreover, you can apply steganography techniques to other types of documents, such as ID cards or passports. Yet, the quality of the image document and the training documents is a limitation of this technique.

## Agilytic's take on fraud detection

The ability to detect fraud with forged documents significantly improves when combining OCR and steganography. What about other value-adding steps? Both forged and genuine documents are essential to training the algorithm and detecting fraud. Yet, in practice, it isn't easy to identify and collect forged documents.

To deal with this issue, we can subdivide image documents into smaller images, cropping the images. Documents with more than one forgery can bring more images to analyze for the algorithm, thereby increasing the number of forged documents to train the algorithm. Applying the detection algorithm to smaller images improves its performance as people can make forgeries on a tiny and specific part of the image. We then reduce the research area, providing better results. Finally, we perform cross-validation of the information present on some documents. It crosschecks documents like payslips or property titles with data extracted from ID cards and passports.

## Time savings and detection accuracy for banks

Fraud detection of forged documents is a significant challenge for the banking sector. Today, we have a robust solution - combining OCR to detect alterations on text documents, steganography to detect forgeries on image documents, and

cross-validation checking between the documents. This results in time savings and a detection rate improvement for banks.

## About Agilytic

Since 2015, Agilytic helps innovative leaders solve their biggest challenges through the smarter use of data. With over 150 successful projects to date, we have perfected a pragmatic approach to putting data at the service of business goals, be they commercial, operational, financial, or human. Reach out today for a quick introduction, we'd love to hear from you.