pwc

Finding the stolen money using OCR and Language Fuzziness algorithms

One of our clients suspected that certain wire transfers made by the company ended up in employee bank accounts.

The client provided us thousands of scanned bank records from various banks containing information about money transfers from the company to its suppliers, and asked us to identify transactions made to employees.

In order to aggregate the data per bank account, it was first required to run OCR (Optical Character Recognition) on the entire batch of documents.

After running the OCR, we had to match a list of the employee bank accounts against the batch of transactions. However, as some documents were poorly scanned, certain bank account numbers were blurry and distorted - Therefore a perfect match was not possible.

To overcome this, we used Fuzzy Searching techniques that allowed us to find partial matches - For example, we were able to match the following bank account number 5!4564561 to an actual bank account number - 514564561 (As the 1 was blurry, it was identified as an exclamation mark. By using Fuzzy Searching, we found that these two records are actually the same).