



Text Analytics

Match customer information across multiple data sources

The Big Picture

One of the largest providers of credit information and information management services in the world wanted to develop an algorithm to appropriately link names and addresses across multiple datasets. The company was looking to achieve measurable performance improvements by refining match accuracy and reducing misclassification rates.

The company already had a basic framework for matching customers by their names and addresses from different data sources, but were striving to improve the current process. The company wanted an improved framework to work with unstructured address data so that it could provide a single view of customers across different data sources.

Transformative Solution

It was determined that dCrypt's Fuzzy Matching module would provide the solution. It combines several different fuzzy matching techniques such as Levenshtein Distance and token-based distance measures, and applies heuristics to give an overall match score, resulting in higher accuracy and recall. The system offered the following:

- Cleaning and normalizing the data to work within a standardized form.
- Segregating the addresses into logical components such as house number, post code, etc.
- Using the post code as a reference so the request address can be searched within the data of the same post code.
- Matching the addresses by the top 100. The corresponding customer names are also matched and the best outputs are generated based on the overall score.

The Change

Using dCrypt's Fuzzy Matching module for higher accuracy enabled the company to track almost 80 million more candidates for credit rating agencies, which were not captured by the existing system. The overall accuracy of matching improved by 7 percentage points, from 74.2% to 81.2%.

Visit <https://fractalanalytics.com/> or email info@fractalanalytics.com. Follow us:  