



Insurance

Match individuals across data sources without a unique key

The Big Picture

A leading credit bureau provided analytics and intelligence support to local credit rating agencies. The company had access to multiple data sources, such as bank data, voter IDs, and tax returns, from where it pulled information for individuals and created their credit score.

However, it was not possible to use information from all these data sources as there was no single unique key. For example, in a bank data set, an individual would have a driver's license number, but in a voter ID data set, the same person would be identified with a voter ID number. The company couldn't know that the individual was the same in both data sets. Therefore, there was a need to create a logic to match these different data sets without having a unique key.

This meant that the company needed an improved framework of working with unstructured addresses data so that it could provide a single view of customers across different data sources. It needed to drive measurable performance improvements by improving match accuracy and reducing false positives.

Transformative Solution

To solve the company's challenges, a solution was deployed to match datasets using names and addresses, since this information was present in all data sources. Since the format of names and addresses were different everywhere, the solution needed to create intelligent and fuzzy logics to standardize names and addresses for mapping purposes.

The approach took raw data and deployed a name and address matching algorithm that was configurable at different levels. The solution incorporated a search capability along with optimization and improvement of matching. Three key steps were:

- **Data standardization:** Data was cleaned and normalized to remove components not adding value to addresses. Addresses were segregated into logical components: house number, locality information, and pin code.
- **Address search:** The approach searched the request address into candidate data using pin code (and derivatives) as a key.
- **Name and address matching:** This step used Fractal's dCrypt to match all the addresses in a key value pair with request address and selected top 100. For top 100 addresses, the corresponding names were also matched and the best output was generated on the basis of name and address matching scores.

The final output provided a list of names and addresses from the candidate data which match the name and addresses from the reference data. Using the algorithm, a matching score was generated between two strings which could be compared with a base matching score already present in the client's sample file. Randomly selected samples were manually checked and a confusion matrix was created for both algorithms.

The Change

As a result of the engagement, the company achieved several benefits:

- Improvement in accuracy by 10% on 11 million household addresses.
- Incorporation of a search capability in the matching algorithm.

- Three different algorithms were used for matching as opposed to a single algorithm for name and address matching, which led to better coverage and efficiency.model.

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

© 2018 Fractal Analytics Inc. All rights reserved | Confidential and proprietary Information of Fractal Analytics Inc. Fractal is a registered trademark of Fractal Analytics Limited.