Case Study

PROBABILISTIC MATCHING ENGINE ALGORITHM THAT IMPROVED DATA QUALITY AND REDUCED DUPLICATES

OVERVIEW

Mastech InfoTrellis collaborated with a leading consumer bank in North America to configure advanced survivorship rules, which identify duplicates, in order to maintain a golden record and drive data discipline. We configured a Probabilistic Matching Engine (PME) algorithm, which helped compare and match different bulk records with multiple fields, and accelerated identification of suspect duplicates.

AT A GLANCE

CLIENT: A Leading Bank

GEOGRAPHY: North America

INDUSTRY: Banking

TECH STACK: IBM MDM

THE CHALLENGE

▷ Need for a centralized system to manage prospects to reduce re-entry of data and eliminating creation of duplicates across product systems

▷ Comparison report was needed between Deterministic Matching Engine(DME) which is used for suspect duplicate processing and, PME which matches fields in multiple records

▷ A Proof-of-Concept Environment was needed to build PME Algorithms, Survivorship Rules, Enterprise ID notification for the golden data record

THE SOLUTION

▷ We configured a PME solution for indexing data that provides capabilities for grouping persons with common attributes such as address, and telephone etc.

▷ The solution helped in updating data roles to advanced survivorship for contract level information and minimize duplicates

▷ Tuned PME algorithm - anonymous values, equivalent words, weights and helped gather metrics to match PME and DME

HIGHLIGHTS

Improved data quality with reduced duplicates

Advanced survivorship rules helped maintain a golden record of data

Reduced the candidate suspect pool in DME logic **from unlimited to between 100 to 200**

THE OUTCOME

The solution improved the data performance, analysis of existing code and configurations, and reduced suspect duplicate tasks with the advanced survivorship rules. The client was able to present results of the efficient matching between PME and DME, with indexed data that grouped individuals with similar attributes, thereby retaining a single reliable record. With the simplified integration of new sources, data was more readily available through real-time transactions.

Learn More