



Deduplication using text mining for a leading Global Information Solution Company

For a Leading Global Information Solution Company

Based out of United States our client is a consumer credit reporting agency that leverages unique data, analytics, technology and, industry expertise to empower 800 million individuals and 88 million businesses worldwide. They offer credit and demographics related data and services to companies and also sell credit monitoring and fraud prevention services to their customers

Behind the Scenes

The process of transforming knowledge into insights that help make more informed business and personal decisions can be susceptible and tricky. Therefore to power this entire process, Happiest Minds proposed to improve their current search engine accuracy and performance while identifying same person or candidate across multiple channels. This exercise expected to scale up the growth rate concerning data volume and transactions and for better search and match accuracy.

1 Strategy and Objectives

Happiest Minds Technologies proposed to build a new Search Match System based on Data Sciences and Big Data that will improve the efficiency and ensures better performance as compared to the old system. The project was done in four phases:

- Phase 1- Improve Accuracy by around 3 % for 15% of the Offline Retail Data
- Phase 2- Improve Accuracy by around 7% for the offline Retail and Rural data
- Phase 3- Create a Big Data / MPP Platform to improve performance of the Analytics Model and Parallel Run of Data Lever and New Big Data Analytics System
- Phase 4- Provide support for Big Data System if needed. Sunset Data Lever and Plan to replace the existing system.

2 Value Chain

- Build a new data-driven Search and Match system using text mining and analytics that would improve accuracy and increase performance as compared to the legacy system.
- The new Search Engine would ensure that it can scale up horizontally with commodity hardware for future growth with regards to clients, types, and volume of data.
- Applied a customized elastic search indexing over the analytical models for faster identification and retrieval of duplicates
- The new systems investment returns could be measured both qualitatively with regards to performance and accuracy, and quantitatively in terms of the saved cost over licensing and third party maintenance.

3 Enablers

- Hardware infrastructure to host Hadoop cluster and search
- Open Source for the Development of Operating System and Big Data software
- Implementing LSH algorithm in the distributed parallel environment

4 Digital Capital Delivered



7% more accuracy compared to the existing system



Reduction in the number of false positives and true negatives



Highly customizable & self-learning analytics model developed



Plug-in component facilitates easy integration with the existing systems

About Happiest Minds Technologies

Happiest Minds, the Mindful IT Company, applies agile methodologies to enable **digital transformation** for enterprises and technology providers by delivering seamless customer experience, business efficiency and actionable insights. We leverage a spectrum of disruptive technologies such as: Big Data Analytics, AI & Cognitive Computing, Internet of Things, Cloud, Security, **SDN-NFV**, RPA, Blockchain, etc. Positioned as "Born Digital . Born Agile", our capabilities spans across product engineering, digital business solutions, infrastructure management and security services. We deliver these services across industry sectors such as retail, consumer packaged goods, edutech, e-commerce, banking, insurance, hi-tech, engineering R&D, manufacturing, automotive and travel/transportation/hospitality.

Headquartered in Bangalore, India; Happiest Minds has operations in USA, UK, The Netherlands, Australia and Middle East.

Contact Us
business@happiestminds.com