

AUGMENTED DATA STEWARDSHIP

One of the most difficult and time-consuming tasks in master data management (MDM) — or in any aspect of data management — is finding and resolving data issues.

Finding data issues is typically done in two ways:

Top-Down:

The traditional approach is to write data quality rules based on either predicting the issues data professionals expect to see and/or developing data governance requirements that specify what well-governed and managed data should look like.

This approach will find many, but not all data issues

Bottom-Up:

Misspellings, data entry mistakes and other errors may creep into data not be caught by data quality validations either because no-one “expected” that type of error or, as a matter of practicality, it would take too many different rules to trap all those unexpected errors.

These are the data errors that consume so much time and effort for data stewards-until now. Profisee uses a machine learning approach to search for unexpected or unusual values in any set of data and present them to the data steward for resolution.

MACHINE LEARNING

Machine learning (ML) is used in many software systems and can provide significant insights that may otherwise be missed by human-only review. In fact, Profisee MDM has used machine learning for matching for many years.

A common problem with ML, however, is that it can require a lot of setup, including large subject-specific training datasets as well as ML expertise that may not be available. By using an unsupervised ML approach, Profisee avoids the need for complex setup and extensive system training, thus improving the usability and usefulness of the system.

SINGLE, FAMILIAR INTERFACE

Since the objective is to make the data steward’s job easier, it is important that they can interact with the data through a simple, intuitive interface. Both rules-based issues and ML-identified anomalies are presented through the same interface, allowing more data issues to be found and resolved more quickly than ever before.

Rules-based errors

Entity Grid	Enter filter keyword			
HL Bottom Bracket	BB-9108	Components	Bottom Brackets	121.4900
Mountain Bottle Cage	BC-M005	Accessories	Bottles and Cages	9.9900
Road Bottle Cage	BC-R205	Accessories	Bottles and Cages	8.9900
BB Ball Bearing	BE-2349	Components	Parts & Components	0.0000
Headset Ball Bearings	BE-2908	Components	Parts & Components	0.0000
Mountain-500 BLK, 40	BK-M18B-40	Bikes	Mountain Bikes	3,500.0000
Mountain-500 BLK, 42	BK-M18B-42	Bikes	Mountain Bikes	55,539.9...
Mountain-500 BLK, 44	BK-M18B-44	Bikes	Mountain Bikes	539.9900

ML-based anomalies