Artificial Intelligence and Data Quality

Understanding the role of AI and machine learning in data management

As large financial institutions adapt new strategies, reach new customers, and leverage existing relationships through data-driven digital transformation, they face the ever-growing problem of data quality management. Banks utilize a wide variety of structured and unstructured data sources in risk-based workflows: credit risk, AML and fraud, operations, and compliance. To improve data quality, standardization, and extraction while preserving dependencies and relationships across data systems is perhaps the most important challenge that artificial intelligence can help solve today.

Daric offers a variety of tools proven to improve data quality across disparate data sets and documents and enable graphical representation of relationships for a given workflow. This document describes a subset of commonly encountered data quality issues and approaches to mitigating their downstream effects on risk-based workflows.

Confronting Data Quality Issues

Banks experience the greatest business impact of data quality issues on regulatory compliance and risk management. Unfortunately, in many cases, the data quality management tools currently in place are geared toward systems analysts rather than workflow managers and data stewards, necessitating a transition to manual processes and neglecting the return-on-investment (ROI) that AI-based tools can offer in a variety of data environments. In Daric's extraction and cleansing workflows, which prepare financial data from a variety of different systems, the typical focus areas are the improvement of data standardization, accuracy (including de-duplication), validity, and integrity. The flexibility of the data routines enables implementation at the source level, ETL, data warehouse, and any other siloed environment.

Data Standardization and Uniqueness

The use of machine-learning (ML) based matching algorithms enables the Daric platform to ingest data for standardization at scale. Typical use cases include matching specific records or data sets to a common standard and transforming data to this standard, allowing for the creation of relationships and accurate links between base data and derived data (e.g. replacing instances of 'DRep' with 'Dominican Republic' in CDD data). This workflow is particularly important in AML and fraud detection scenarios, where a high volume of customer due diligence and transaction data from disparate systems necessitates extensive standardization to set flags and generate meaningful derived data.

Standardization simplifies de-duplication (uniqueness) issues and accuracy-related data quality problems. Conformance to schema and to business rule constraints are typical metrics that banks seek to optimize during the standardization process. The flexibility of ML means that changes to these metrics can be applied across the entire data set in a cost-effective way, reducing the overhead of moving to a new standard.

⊙ DARIC

Data Accuracy and Completeness

The AI-driven features of Daric's data accuracy solutions recognize the limitations of pattern matching and predictive data completion in incomplete records. However, the use of best-in-breed technology to create a multi-model approach that "learns" from ingestion of additional data establishes analytic routines for error detection, flagging, and suggested corrections. These routines, which incorporate proprietary NLP technology along with a user-friendly data interface, enable analysts to augment manual data accuracy improvement efforts with machine intelligence.

The combination of machine and human data analyst intelligence is powerful. Driving down data accuracy errors and the time to integrate data from disparate sources translates to operational gains at all points in a workflow. For example, self-correction routines that adjusted frequent mistakes on customer input databases improved the processing time on home equity loan applications for a representative bank by nearly 160 minutes on average.
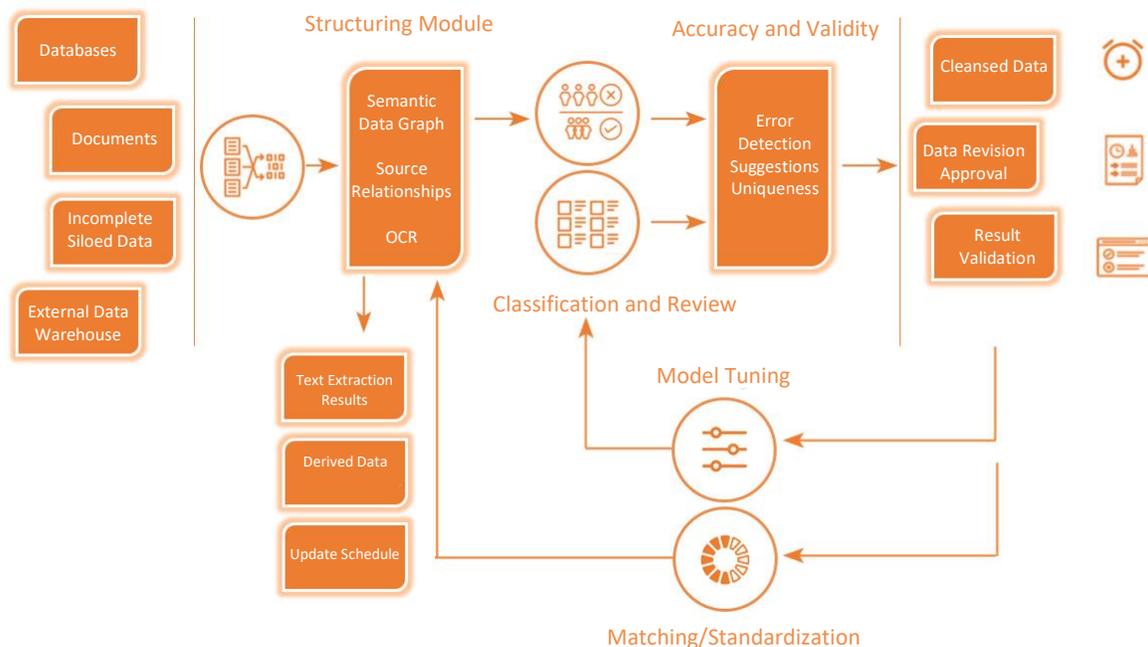
Data Extraction, Timeliness, and Consistency

Scalability is a key concern with a data extraction process that requires adaptation to a variety of different formats, structuring levels, and data environments. Often overlooked in the data integration process, the use of artificial intelligence to time extraction routines to reflect data interdependencies and to optimize resource allocation enables the creation of consistent, timely, and structured data which can be transformed into a semantic graphical representation.

Timeliness requires a consistent, repeatable update process that runs when necessary but does not consume resources when it is not required. Additionally, because many data sources in unstructured form must be converted with OCR (optical character recognition) routines to structured data, preserving the order of data source loads and dependency-based relationships in graphical (node-based) form is a key differentiator of Daric's approach to data integration. For example, to generate AML risk flags from base data sets on transactions with correspondent banks, it may be necessary to load KYCC documentation on correspondent banks, turn it into structured data, integrate the results with information from variety of other customer and transaction databases, and finally generate the set of flags.

Semantic technology and data graph representations preserve consistency without sacrificing timeliness, accuracy, and cost-effective standardization priorities. A sample overview diagram is provided above on the use of the AI-based data cleansing layer from a workflow-agnostic perspective.

DARIC

Learn More

To request more information on our data quality management technology, please contact us at +1 (650)-300-7740. You can also reach us at:

O DARIC

4677 Old Ironsides Drive
Santa Clara, CA 95054, USA
solutions@daric.com

About Daric Inc.

Daric, Inc. provides credit risk management, digital lending, and AML platform solutions to banks, credit unions, and financial institutions worldwide. Backed by leading figures in the financial and technology industries including former Wells Fargo CEO Richard Kovacevich, Daric is based in Silicon Valley, CA, providing solution services and worldwide support. Our team includes veterans of the world's leading financial and financial technology companies such as Palantir Technologies and Teradata.