

## Using Data Automation to Minimize Risk

### A CRO's Headache: Success Produces Risk

Due to the sensitive nature of our client's operations, we will not disclose the name of our client or executives within this article. To improve readability, we will instead refer to the following:

- **Global Bank** – A Fortune 50 Commercial Bank
- **Mr. Smith** – The Credit Risk Officer of Global Bank

Our client, Global Bank, was buried in a classification problem that multiplied their credit risk. As one of the world's leading lenders, they had different loans and forms of credit out to millions of legal entities within the U.S. and internationally. While this was great for the business, it was a headache for Mr. Smith, who had to balance a multitude of counterparties, credit obligations, and continuously evolving regulations. To better calculate and manage Global Bank's credit risk, he needed to understand the industry classification, or the line-of-business, for each of Global Bank's counterparties.

To achieve this goal, Mr. Smith tasked a team with developing a unique classification system. The system needed to provide sufficient granularity in areas of heavy lending and higher default risk while affording less specificity in lower-impact industries. This resulted in a classification system with over 120 distinct classifications of varying specificity.

- **Less Specificity** – Agriculture, Retail, Entertainment
- **Medium Specificity** – Investment Advisory Services
- **More Specificity** – Savings Bank, Swap Dealer, City Government



**The Global Bank's previous system required thousands of hours of manual input and still produced low accuracy. The low accuracy undermined confidence, increased costs, and introduced risk.**

## The Credit Risk Officer was forced to manage the intersection of millions of counterparties and an ever-changing regulatory landscape.

Mr. Smith sourced classification data from a third-party vendor and assigned an operations team of several hundred employees with transforming the vendor's classification to a value within Global Bank's classification system for over 120,000 legal entities. This process required thousands of hours of manual input and multiple iterative reviews to achieve the desired accuracy. The low accuracy decreased the classifications' confidence and value, increased the cost of the operation, and introduced considerable operational risk.

To decrease operational risk and increase the process's accuracy, Mr. Smith turned to Kingland for a possible alternative. Through a 2,000-record Proof-of-concept (POC), Kingland worked with Mr. Smith to establish an ongoing process using Kingland's people and the Kingland Data Platform. The POC resulted in the classification of legal entities at over 95% accuracy and the development of a successful, repeatable process for classifying legal entities.

The classification process's goal was to determine the line-of-business of the legal entity and apply an appropriate classification from Global Bank's classification system. Traditionally, manual research such as reading financial statements, looking at the entity's website, and using Google search would determine the line-of-business. However, a manual approach tends to be expensive, and any time humans are introduced, consistency may decrease. Kingland instead explored a hybrid automation-human approach to classifying entities. The successful process included four distinct steps: Source Matching, Opinion Development, Manual Review, and Feedback Incorporation. The following sections will describe each step.

### Extensive manual processes all but guarantee:

1. **Low Quality Data**
2. **Significant Operating Cost**
3. **Elevated Operational Risk**

The Credit Risk Officer, looking for an alternative, partnered with Kingland for a proof-of-concept. He discovered three things:

1. **Vendors Diffuse Operational Risk:** By partnering with a third-party vendor, the bank diversified their operations and reduced operational risk.
2. **Algorithmic Approaches Decrease Costs:** Leveraging Kingland's proprietary data algorithms allowed more classifications to be applied at lower cost.
3. **Human-Automation Hybridization Increases Quality:** Combining an ecosystem of researchers, algorithmic classification, and machine learning resulted in a high-quality classification process.

## Source Matching: Compiling Classifications

**Because classification data cannot come from a single source, Global Bank needed a process that could pull from multiple sources to create a base opinion on entity classification.**

Fortunately, dozens of publicly available data sources provide their own opinion on an entity’s classification. Sources like the GLEIF, datasets from data.gov, and more can be downloaded and accessed for commercial purposes as seen in Figure 1.1. For this process, Kingland used up to a dozen data sources to establish a data lake of legal entity classifications.

Next, Kingland matched the legal entity records to the classifications within the data lake. The more supporting reference data available, such as headquarter addressing, board members, and ownership structure, the more complete and accurate the matching step becomes. A confidence level was applied depending on the match’s confidence and the quality of the data source.

With data sources integrated, there will be either zero, one, or multiple classification opinions aligned with each legal entity. Next, Kingland used an algorithmic approach to formulate a single classification opinion.

Global Bank’s Dataset				Integrated Sources			
Record ID	Legal Entity	Country	Data Source	Legal Entity	Country	Classification	Confidence
0001	Blackrock Inc	USA	GLEIF	Blackrock Inc	USA	Asset Manager	.95
0001	Blackrock Inc	USA	OSHA	Blackrock Inc	USA	Asset Manager	.90
0001	Blackrock Inc	USA	Other	Blackrock Inc		Asset Manager	.75
0002	JPMC	USA	OSHA	JPMC	USA	Investment Bank	.92
0002	JPMC	USA	Other	JP Mogan		Commercial Bank	.80
0003	Midwest Bank	USA	-	-	-	-	-

Figure 1.1

## Opinion Development: Creating Confidence

**Each record required a calculated single classification opinion.**

A proprietary algorithm aggregated multiple different classification opinions into one recommendation with a resulting confidence level. Kingland Data Scientists calculated a confidence baseline, and anything below that baseline had their automated classification removed as seen in Figure 1.2.

Record ID	Legal Entity	Country	Data Source	Legal Entity	Country	Classification	Confidence
0001	Blackrock Inc	USA	GLEIF	Blackrock Inc	USA	Asset Manager	.97
0002	JPMC	USA	OSHA	JPMC	USA	Investment Bank	.35
0003	Midwest Bank						



The Kingland Platform's built-in automation showed below-standard classifications to be removed and replaced in order to set a standard for data quality.

## Manual Review: Limited Human Intervention Augments Automation

The process then saw the banks dataset split into two separate groupings.

**Group 1:** As seen in Figure 1.3, this contained records without any automated classification opinions OR records without a sufficient confidence level. These records required manual research and review by two operational professionals, resulting in a singular classification opinion.

**Group 2:** This contained records with an automated classification opinion AND a sufficient confidence level. These records required manual research by a single operations professional, but no secondary research was necessary if the professional agreed with the automated classification. This step ensured a high level of quality while reducing the costs associated with manual data research.

Record ID	Legal Entity	Classification	Confidence	Grouping
0001	Blackrock Inc	Asset Manager	.97	Group 2
0002	JPMC	Investment Bank	.35	Group 1
0003	Midwest Bank	-	-	Group 1

Figure 1.3

## Feedback Incorporation: Produced 99% Accuracy

Over time, Kingland Data Scientists developed an automated feedback loop and performed additional data analysis. Certain automated classifications had an incredibly high acceptance rate – the manual review consistently agreed with the automated classification. Other data analysis found a trend within names – certain words within a legal entity name would strongly correlate with a specific classification. Eventually, these trends could lead to the application of a classification without the manual review step.

A POC of 2,000 records expanded into a full-scale solution with over 2,000,000 records-under-management through this repeatable automated process. Classifications are applied at above 99% accuracy. Mr. Smith can successfully calculate Global Bank’s credit risk with a significant decrease in operational risk.

Successfully classifying 2,000,000 records empowers Global Bank to manage its credit risk while also containing operational risk.

## Your Data. Our Platform.

The Kingland Platform powers the worlds most integral companies. We transform your existing data into actionable insights to manage risk and generate revenue opportunities.