

Anomalo: Finding Data Anomalies

The Company

Anomalo was founded in 2018. It uses AI to automate the detection of issues in data, whether that data resides in structured or unstructured form, and assists with the resolution of the issues that it finds. It has raised over \$80 million in funding from investors, including Block and Discover Financial, as well as Databricks Ventures and Snowflake Ventures.

Anomalo is headquartered in Palo Alto, California, USA.

Visit their website: www.anomalo.com

Anomalo Global Presence & Financial Strength

confidential	confidential	100+	confidential
Annual revenue	Countries with customers	Employees	Corporate customers

The Market

Anomalo operates in the data observability and data quality market. Customers include Nationwide, Lebara, Discover, BuzzFeed and Block.

Overview

Anomalo makes heavy use of machine learning and large language models to allow it to automate the process of detecting anomalies in data. The models study the patterns of real-world data and set thresholds on the natural variations in that data, without having to rely on hard-coded manual rules. Anomalo issues alerts when data deviates from the norm, and allows business users to supplement its rules with business-specific ones via a natural language interface. This interface can be used to suggest the root cause of data issues in plain English.

Key Features:

AI data monitoring	Feature 02	Feature 03	Feature 04
Anomalo uses machine learning to spot patterns and variances in data, not just metadata, so it can set thresholds and generate meaningful alerts when data anomalies are found.	Data quality can be applied to unstructured data, for example setting rules to detect personally identifiable data. It is not restricted to structured data.	Automated rule generation for the bulk of cases can be supplemented by business specific rules set by business users via a natural language interface.	Support for data lineage and data ownership. Can use AI to analyse applications such as Slack to determine likely data ownership.

The Technology

Anomalo can read a wide variety of data sources, both structured and unstructured. It is tightly integrated with the modern data stack, having partnerships with both Snowflake and Databricks. The tool uses machine learning to analyse patterns in data and to set up data quality rules automatically, without needing static limits and thresholds. It supplements these rules, which cover the bulk of cases, with business-specific rules defined by human domain experts. It provides a natural language interface in order to appeal to business users, who can set supplemental data quality rules up in natural language rather than needing to learn a new design tool or programming skills. The

technology comes with its own AI analyst, called AIDA, which can explain the likely root cause of data issues in plain English, based on its monitoring of the patterns in data. This tool is based on a large language model, and there is a verification agent to provide additional reassurance for factual-style queries that users enter through the interface. The product can set data quality rules for unstructured data such as documents too, to detect things like personally identifiable information or to detect abusive language in text sources. The technology is cloud native and runs on major cloud problems as well as within private cloud environments.

Competitors

Anomalo competes with other data observability vendors, such as Monte Carlo, Bigeye, Acceldata and Soda, and with traditional data quality vendors such as Informatica, Ataccama, Precisely and SAS.

Summary

Anomalo is a modern data observability tool with extensive automation capability, able to deal with unstructured as well as structured data, an increasingly important feature in a world of corporate AI projects. It has an intuitive interface and has AI, in different forms, embedded deeply in the architecture, rather than being an afterthought. It should be considered by companies looking to monitor the quality of large-scale data pipelines.

Market Differentiation:

- Use of AI to spot anomalies in data values, not just metadata.
- Schema freshness and lineage monitoring
- No-code rules creation aimed at business users.
- Automated root cause analysis.

"Anomalo was the one bringing statistical analysis to the field"

Mike Randall, Director of Enterprise Data Governance, Nationwide

About The Information Difference

At the Information Difference www.informationdifference.com, we offer in-depth analysis of the data management industry, in-depth analysis of the main vendors, assessments of the marketplace and white papers discussing key issues and best practices.

If you are contemplating a data management project, we can advise you on strategy, vendor selection and best practice.