WHITEPAPER

Unraveling Fraud Networks

fractal

Harnessing graph-based techniques for robust and real-time fraud detection

Fraud is nothing new, but today the urgency for a fail-safe fraud detection system is more critical than ever. This necessity springs from the paradox that technological advancement, while a boon for users, also empowers fraudsters.

Fraud detection has traditionally been anchored on data mining and statistical analysis — tools sufficient for detecting comparatively simple fraud. But as fraudsters begin to weave more complex webs of deceit, these traditional methods are fast becoming outpaced and outdated.

As a result, cutting-edge detection mechanisms and architectures have surfaced, bolstering companies' capabilities to spot fraud. This whitepaper delves into the innovative role of graph-based technologies, demonstrating their potential in real-time detection and accurate prediction of complex fraudulent activities, such as money laundering and other elaborate schemes.

Breaking away from tradition

As fraudsters become more sophisticated, traditional approaches become less effective. Current methods for fraud detection include:

1. Statistical analysis

A traditional approach where statistical models scrutinize data, looking for unusual patterns or anomalies that could hint at fraudulent activity.

2. Data mining

This method involves an exhaustive analysis of vast data sets to unearth patterns and connections that could signal fraud.

3. Rule-based systems

This approach operates by devising a set of predetermined rules or criteria, serving as a beacon to pinpoint potential fraudulent transactions.

4. Pattern recognition

By leveraging machine learning algorithms, this method identifies recurring patterns or data anomalies that might suggest fraud.

Innovative strategies like machine learning and graph-based technologies that combine traditional and advanced methods are needed to deter and prevent fraud effectively.

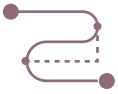
A fresh perspective on fraud detection

Existing techniques struggle to discern the intricate relationships between entities, often the key to spotlighting suspicious behavioral patterns. Graph-based algorithms have emerged as a compelling answer to this challenge. In this approach, transactions and customers are transformed into nodes and edges, enabling fraud detection algorithms to tap into the strength of relationship mapping to identify fraudulent activities.

> Graphs underscore the relationships between entities, making it convenient for investigators to discover patterns that would remain camouflaged within conventional tables and help reduce the false positives that often plague traditional methods by offering an encompassing visualization of the network of connections. This method proves invaluable in unmasking fraud networks, where behaviors are interwoven rather than standalone.

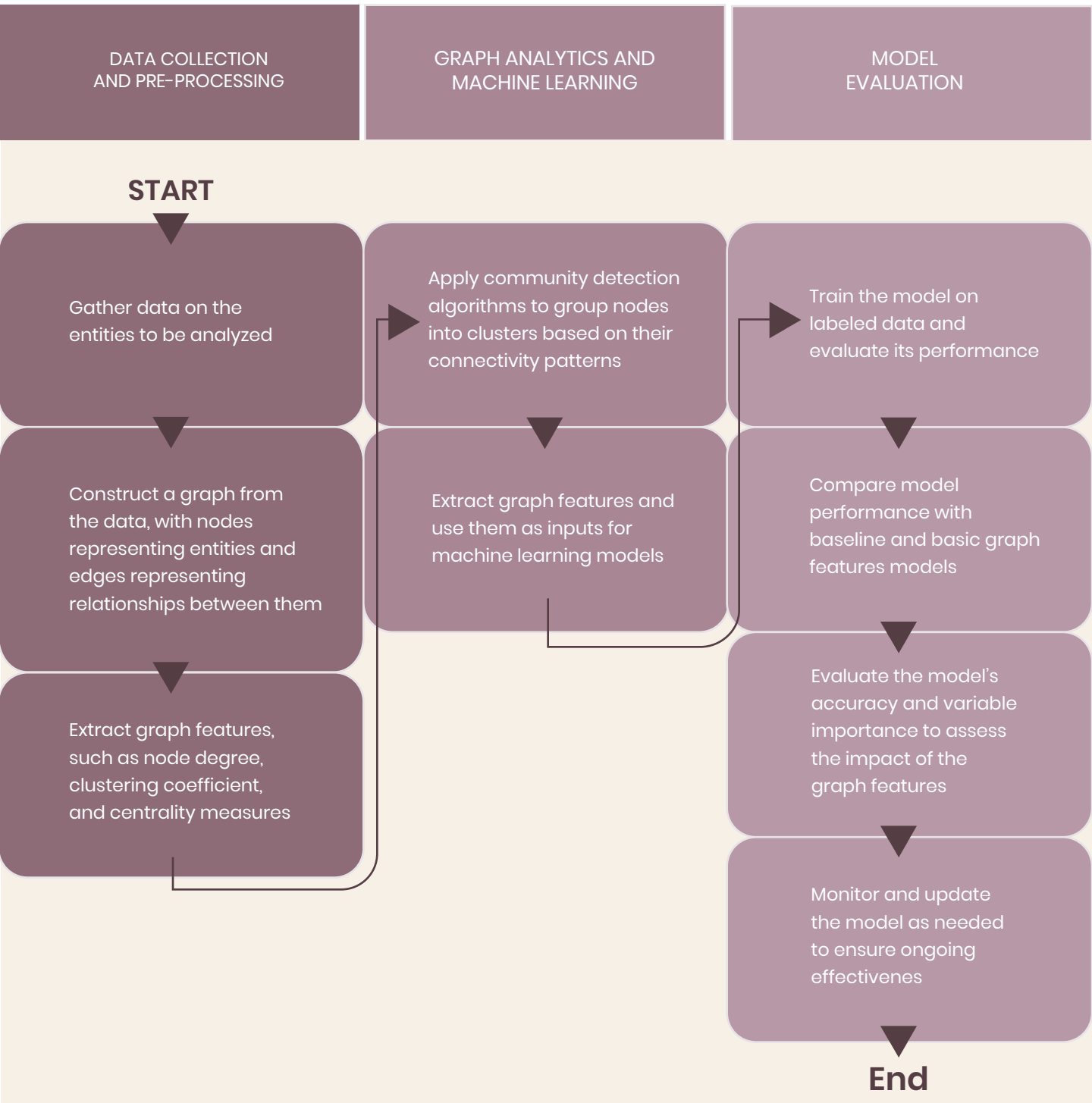
Key metrics: The facets of graph algorithms

KEY METRICS	USING KEY METRICS TO DETECT FRAUD	WHY KEY METRICS ARE CRUCIAL
Community Detection 	Community detection clusters nodes in a graph using modularity or spectral clustering methods based on attribute or connection similarities. This paves the way for detailed analysis of these clusters to spot potential fraudulent actors or activities.	Community detection plays a crucial role by pinpointing groups of nodes exhibiting similar properties or behaviors, potentially signaling fraudulent activity. Given that fraudsters often operate in cohorts or employ similar strategies, identifying these communities is instrumental in fraud prevention.
Centrality Analysis 	Centrality analysis utilizes measures like PageRank or eigenvector centrality to pinpoint influential nodes within a graph. By harnessing these metrics, we can enhance our ability to identify potential perpetrators of fraudulent activity.	Centrality analysis highlights influential nodes in a graph that could potentially signal fraudulent activities. If each node represents a criminal act, this analysis highlights the crime with the most involvement, offering a glimpse into its popularity.

KEY METRICS	USING KEY METRICS TO DETECT FRAUD	WHY KEY METRICS ARE CRUCIAL
Page Rank 	PageRank gauges the significance of each node in a graph, assigning a score based on the quantity and quality of its interconnected links. It thoroughly evaluates incoming and outgoing connections to create a comprehensive link structure analysis.	PageRank scores nodes in a network, identifying anomalies based on their prominence. Nodes with high scores often have numerous inbound links from dubious sources, indicating potential fraudulent involvement. A thorough investigation of these nodes could significantly reduce fraud network risks.
Clustering Coefficient 	Clustering coefficient analysis groups graph nodes based on attribute or connection similarities using hierarchical or k-means clustering techniques. The resulting clusters are cross-checked against a maintained list of fraudulent transactions for potential matches.	Graph-based clustering analysis groups similar or proximate nodes, which could signal fraudulent activity. As fraudsters typically operate in clusters or employ similar methods, detecting these groupings can prove beneficial in identifying fraud.
Shortest Path 	The shortest path algorithm traces the quickest route between two graph nodes, highlighting the minimum number of connecting edges. This tool proves valuable in fraud detection, unveiling suspicious transactions across multiple nodes, and potentially exposing indirect connections or intermediary involvement.	The shortest path analysis uncovers hidden node relationships within the network. Fraudsters often employ indirect connections to elude detection. The shortest path algorithm can expose these hidden links, assisting investigators in identifying suspicious transactions.
Classification 	Classification employs evidence from past cases to predict an entity's category, serving as a robust fraud prevention tool. The model harnesses graph-extracted features like node attributes, transaction specifics, and inter-node relationships. After training, it can classify new transactions or nodes as legitimate or suspect.	Classification aids in the real-time identification of potential fraudsters and their activities. Automated fraud detection allows swift identification and flagging of dubious transactions or customers, mitigating financial risks and preserving an institution's reputation.

How to implement a graph-based algorithm

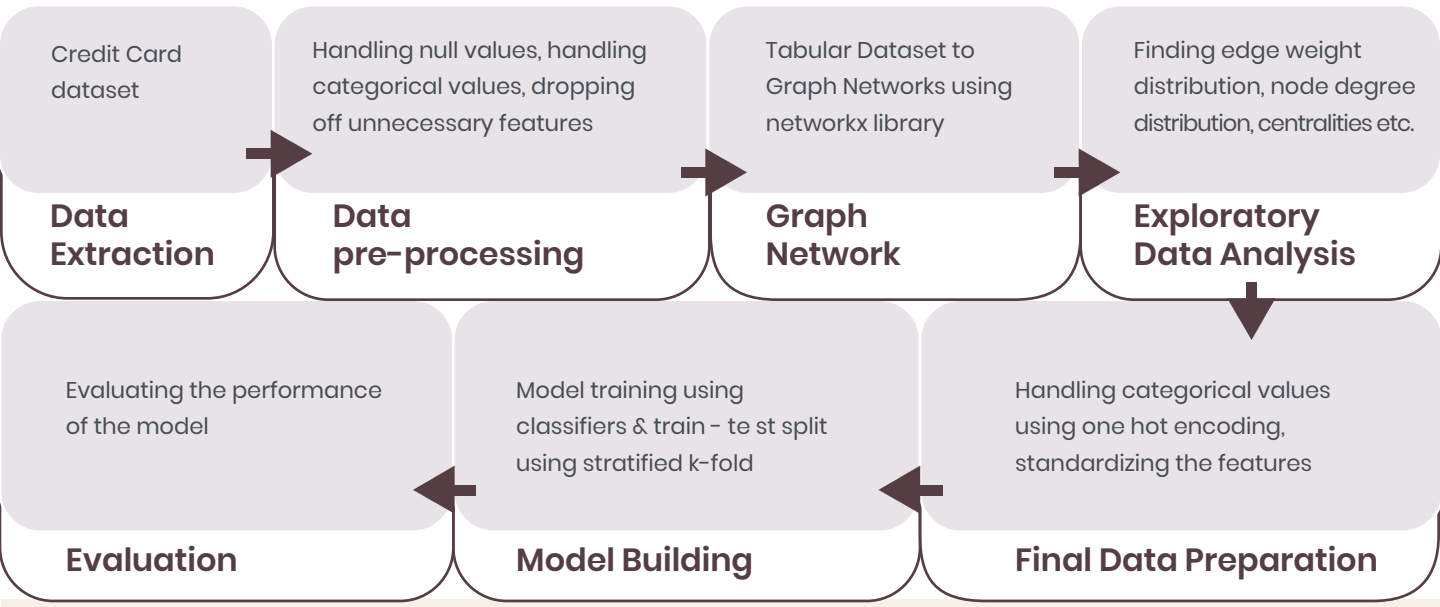
There are three key steps to implementing graph-based algorithms to detect fraudulent activities.



Case Study: Graph-based Techniques in Action

The Client	A financial institution or credit card company.
The Challenge	The client is grappling with detecting and preventing fraudulent transactions within their credit card platform. Their goals are twofold — to curtail financial losses and to shield their customers from unauthorized charges.
The Proposed Solution	The deployment of a real-time fraud detection system capable of accurately identifying fraudulent transactions. This proactive approach enables the financial institution to initiate timely countermeasures, mitigating risks.
Fractal's Role	To illuminate complex transactional relationships, which are instrumental in detecting potential fraudulent behavior.

Our approach follows a logical progression from extracting the relevant data to evaluating the results.



To accurately reflect the connections between customers and their transactions, Fractal creates graphs that are structured as follows:

Nodes: These represent the credit card number and merchant.
Edges denote transactions between the credit card number and the merchant.
Edge Weight: This signifies the transaction's magnitude or amount.

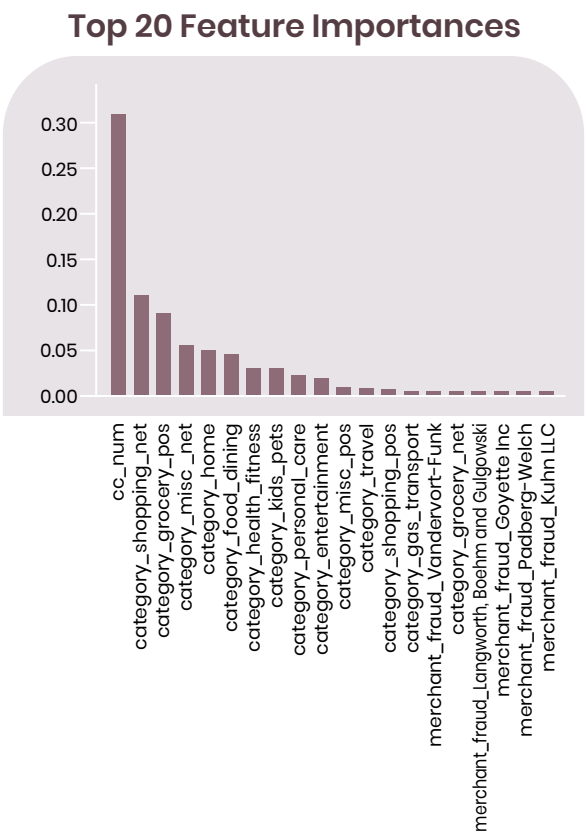
When graph features are incorporated into the model, they emerge as the most influential factors. In our case study, an increase in accuracy was noted, with the Area Under the Curve (AUC) metric rising from 0.72 to 0.76, reflecting an increase of nearly 6%.

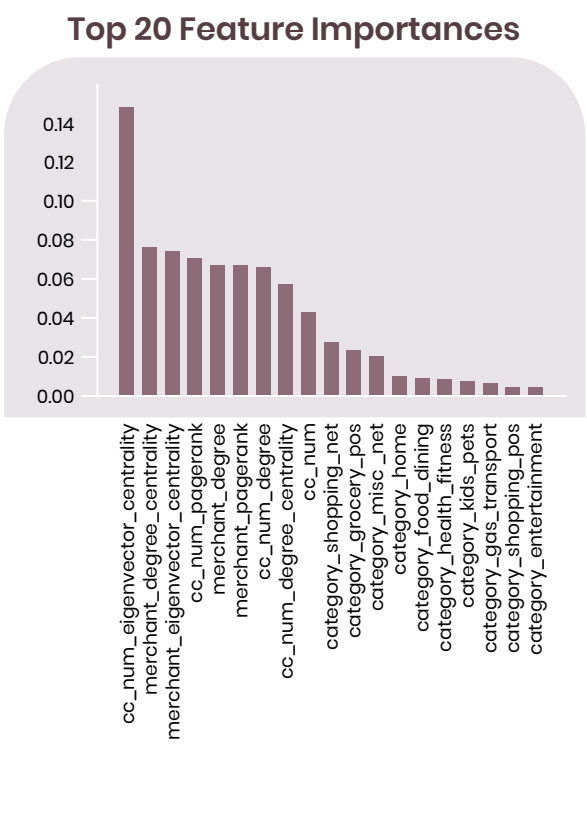
Using intrinsic features

	precision	recall	f1-score	support
0.0	0.66	0.76	0.71	1307
1.0	0.71	0.60	0.65	1258
accuracy			0.68	2565
macro avg	0.69	0.68	0.68	2565
weighted avg	0.69	0.68	0.68	2565

Using graph features

	precision	recall	f1-score	support
0.0	0.72	0.72	0.72	1307
1.0	0.71	0.70	0.71	1258
accuracy			0.71	2565
macro avg	0.71	0.71	0.71	2565
weighted avg	0.71	0.71	0.71	2565

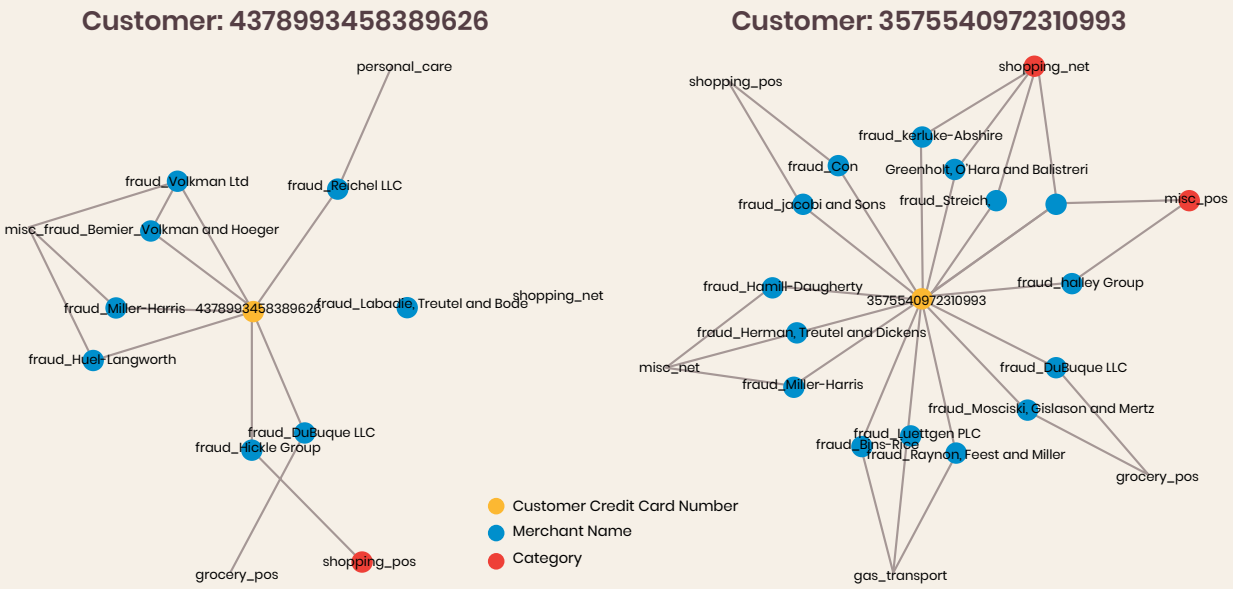




The key features mimic those of the model that lacks graph features; however, it's been enriched with graph metrics. Additionally, the relevance of each feature sees a significant boost compared to their counterparts in the base model. This suggests that including fewer graph features can augment the model's accuracy and insight, offering a greater yield than a wide range of features used in a model without graph elements

The key features mimic those of the model that lacks graph features; however, it's been enriched with graph metrics. Additionally, the relevance of each feature sees a significant boost compared to their counterparts in the base model. This suggests that including fewer graph features can augment the model's accuracy and insight, offering a greater yield than a wide range of features used in a model without graph elements

Three-Level Network Graph for Fraud Transactions



The tri-level network graphs mapping fraudulent transactions (illustrated above) provide a striking visualization of relationships between customers, fraudulent merchants, and corresponding categories for two customers. This graphical representation is instrumental in pinpointing patterns, deciphering connections, and distinguishing clusters within fraudulent transactions.

The inclusion of graph features also enables us to analyze the correlation between the target and the graphical attributes. Integrating these graph features has enhanced precision and accuracy, outperforming the model that lacks graph features. The model's performance could be further amplified by leveraging the power of community analysis, suggesting promising avenues for future optimization.

The final hurdles

As with any emerging technology, several challenges must be addressed before graph-based fraud detection methods can be widely adopted. These include:

- **High computation time:** The process of computing graph features can be time-consuming, mainly if the data set is large.
- **Data quality:** Sparse data and missing information can introduce complexities in creating graphs or network features.
- **Graph network visualization:** Plotting a graph network can be extremely challenging when dealing with a dense network or a large data set.
- **Domain expertise:** A strong foundation in the subject matter is a crucial prerequisite for identifying network structure and determining relationships.

Conclusion

Harnessing the potential of graph techniques can highlight underlying data and its relationships, providing critical insights into seemingly unconnected events in a given use case.

In the banking industry, where fraud incurs high costs, financial services firms using graph database techniques have reported millions of dollars in savings due to the increased accuracy when using graph techniques. The strength of this network approach enables stakeholders to pinpoint and address critical areas in the network, broadening the possibilities for graph analytics and other computational applications.

To build this capability, substantial investment in infrastructure is required, alongside the development of unique customer identifiers that can be used across various systems. Multiple tools are available today for creating graph databases and graph features, which can be subsequently integrated into machine learning models to increase prediction accuracy.

Authors

Ashna Taneja

Consultant,
Fractal Dimension

Sray Agarwal

Principal Consultant,
Fractal Dimension

Supriya Panigrahi

Consultant,
Fractal Dimension

About Fractal

Fractal is one of the most prominent providers of Artificial Intelligence to Fortune 500® companies. Fractal's vision is to power every human decision in the enterprise, and bring AI, engineering, and design to help the world's most admired companies.

Fractal's businesses include Crux Intelligence (AI driven business intelligence), Eugenie.ai (AI for sustainability), Asper.ai (AI for revenue growth management) and Senseforth.ai (conversational AI for sales and customer service). Fractal incubated Qure.ai, a leading player in healthcare AI for detecting Tuberculosis and Lung cancer.

Fractal currently has 4000+ employees across 16 global locations, including the United States, UK, Ukraine, India, Singapore, and Australia. Fractal has been recognized as 'Great Workplace' and 'India's Best Workplaces for Women' in the top 100 (large) category by The Great Place to Work® Institute; featured as a leader in Customer Analytics Service Providers Wave™ 2021, Computer Vision Consultancies Wave™ 2020 & Specialized Insights Service Providers Wave™ 2020 by Forrester Research Inc., a leader in Analytics & AI Services Specialists Peak Matrix 2022 by Everest Group and recognized as an 'Honorable Vendor' in 2022 Magic Quadrant™ for data & analytics by Gartner Inc. For more information, visit fractal.ai

Corporate Headquarters
Suite 76J,
One World Trade Center, New York,
NY 10007

Get in touch