FullContact

Enhancing Fraud Detection with Big Data Analytics

Fraud attempts are increasing. Fraudsters are getting more sophisticated. Regulations and industry-specific accountability requirements are getting tighter.

It's a perfect storm, and to weather it, businesses need stronger fraud detection and prevention solutions than ever. Unfortunately, traditional solutions aren't cutting it—86% of potential synthetic identities are not flagged by traditional fraud detection models.

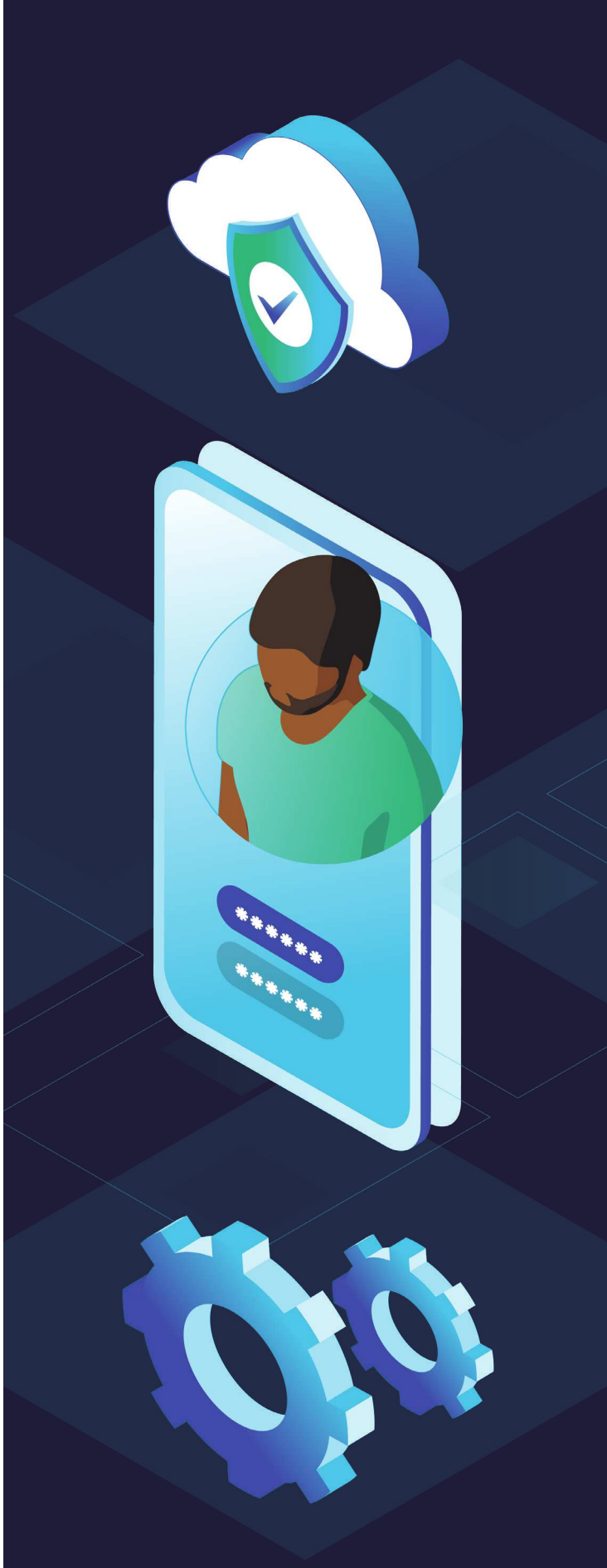
That means significant costs for business.

For example, the most recent LexisNexis® True Cost of Fraud™ Study reports that the cost of fraud for U.S. financial institutions and lending firms has increased between 6.7% and 9.9% compared with before the pandemic. "Every $1 of fraud loss now costs U.S. financial services firms $4.00, compared to $3.25 in 2019 and $3.64 in 2020."

So, what's a business to do?

The answer to more effective fraud detection lies in having the right data for analytics.

Even the best fraud detection models are dependent on the data and signals they process. Using broad signal sourcing and insights can dramatically improve your fraud detection and prevention capabilities.



Understanding the world of big data analytics

Let's start with the basics: what is big data?

Big data refers to the huge amounts of data—structured and unstructured—collected by businesses every day.

Big data analytics refers to the ability to analyze these extremely large data sets to glean insights for improved decision making. Often, these big data analytics are used to help uncover and understand patterns or trends. The larger the data set, the clearer the patterns or trends become, allowing you to make better decisions. This is especially true when it comes to analyzing data on human behavior.

Analyzing big data via data mining, predictive analytics, and other data sciences can provide all kinds of valuable insights that allow businesses to:

- **Identify consumer shopping trends**
- **Personalize marketing**
- **Predict next-best content**
- **Monitor business trends in real time**
- **More effectively manage the supply chain**
- **Detect fraudulent activities**

No industry is seeing the monetary results of big data applications more than consumer ecommerce brands. For example, insights into individual consumers' behaviors and preferences help drive powerful recommendation engines to suggest additional items. Big data analytics also allow consumer-facing brands to develop highly effective lookalike audiences—sets of people that share traits and behaviors with existing customers.

You must balance fraud risk management with customer experience

Demand for fraud detection and prevention will continue to rise as fraud rates increase and attacks become more sophisticated, with the emerging fraud types such as social engineering fraud and synthetic identity fraud.

Figuring out how to effectively detect fraud requires a delicate balance between safety measures that weed out the bad guys—and experience optimizations that let real customers through without too much hassle.

The hard reality is that, as you try to streamline your customer onboarding, improve customer experience, and reduce time and friction to convert, you're probably making it easier for bad actors to break through your defenses.

> "With the accelerated movement to online/mobile transactions and payments, financial services and lending firms must continue to build out and enhance the digital customer experience while protecting against fraud."
>
> **CHRISTOPHER SCHNIEPER,** DIRECTOR OF FRAUD AND IDENTITY, LEXISNEXIS RISK SOLUTIONS

The answer is in big data

If big data analytics can uncover consumer shopping trends, it can also identify patterns that don't fit typical scenarios—without putting undue burden on your legitimate customers. This is the foundation for fraud detection.

But first you need a lot of data to work with—this is where an identity graph comes in handy.

Using an identity graph to inform fraud detection and prevention models

Large and diverse data sets are needed for fraud and risk signals to cross check information, apply machine learning algorithms, and generate trustworthy customer assessments.

A deterministic, vendor-agnostic identity graph can provide these data sets—and distill them into intelligence—to inform effective fraud detection and prevention algorithms.

An actionable identity graph is a database of all identifiers matched with privacy protected customer profiles. You can think of it like billions of contact fragments—names, device IDs, website visits, transactions, etc.—with connections between them. A grouping of fragments and the connections between them represent a person, formed in the graph.

The data intelligence needed to detect and prevent fraud lives within this graph.

For example, identity resolution using the identity graph can determine if a fragmented identifier, like an email address or social handle, is actually associated with a real person. Conversely, the graph can also provide signals as to whether a particular email address may present a fraud risk.

What kind of big data signals are important to fraud detection?

Email addresses offer some of the most important insight to incorporate into fraud scoring algorithms. The right identity graph can give you information like:

- **Are there multiple email addresses associated with a person?**
- **When did our graph first see activity for an email address?**
- **When did our graph last see activity for an email address?**
- **How many data sources say this email address exists?**
- **How long has the email address existed for?**

A brand new email address with no other associated email addresses may be a red flag, for example.

Mobile advertising IDs (MAIDs) can offer similar signals as email addresses regarding a person's mobile behavior.

Social handles also offer insight into potential fraud attempts, specifically Twitter and LinkedIn Social data is valuable because it not only indicates that the person is present online—it also offers insight into how narrow or wide the person's social presence is. For example, a social account with no photo and limited posts may indicate a fraudster.

Location is another important data point. The identity graph can provide information on where a person is likely located. Does this location match with where the person has claimed they are located?

Other data to consider when it comes to fraud detection and prevention include:

- **Does the graph have a first and last name associated with your identifier?**
- **Does the graph have a date of birth associated with your identifier (important in the gaming industry, for example)?**

None of this data on its own means that fraud is definitely afoot. But when the data is combined and built into fraud detection algorithms, it offers a high level of confidence as to whether or not you're dealing with a legit person—without making this person jump through a dozen hoops to prove who they are.

Fraud detection data for growing businesses

Larger organizations have dedicated data science teams to ingest and use these highly-descriptive fraud signals. But not every business has a team of data scientists to build models and complete data analysis.

In this case, simple match functionality can be useful—**do the identifiers you have on a person match what's in the identity graph?**
Do the name, address, email and phone number all point to the same person? If everything matches, your fraud risk is low. If things aren't quite lining up, you might want to gather more touchpoints to verify identity. This approach is particularly useful for synthetic identity fraud.

Pre-built algorithms can also offer a quick thumbs up or thumbs down on fraud risk.
For example, FullContact provides an aggregate email score that indicates risk level based on what we know about that email address. This is very similar to how a credit score works.

We'll be expanding these aggregate measures soon to provide high-level risk scores on:

Employment

Social media

General online activity

Think of it like TSA PreCheck. The PreCheck program lets verified travelers—who present the lowest risk—go through a faster, simpler security checkpoint. Matches and pre-built activity scores give your lowest-risk customers a frictionless verification experience.

Putting the data to work for your fraud detection efforts

Historically, you've had two options:

- Clamp down on fraud so much that everyone has to jump through a lot of hoops.
- Open your doors wide and eat significant costs on the fraud that ensues.

Your ability to find a middle ground has only become more dire as fraud attempts and their costs continue to increase.

Thankfully, fraud detection with big data analytics offers a realistic path forward.

When you can harness big data for your fraud detection programs, you can catch the bad guys without overburdening your legitimate customers. Data signals from a deterministic, vendor-agnostic identity graph offer all kinds of businesses the intelligence they need to achieve this balance.

Learn more about how we use big data to help clients detect and prevent fraud.

Check out FullContact's Verify solution today.

FullContact

Amplify your ability to recognize and reach real people by **3X.**
Or it's on us. **Guaranteed.**

Talk to our Identity Solutions Experts to learn more.
fullcontact.com/contact

Real People

Access and map fragmented physical and digital identities into a **persistent PersonID from a single graph.** Omnichannel input and outputs.

- **248 Million** People
- **50 Billion** Individual Omnichannel Identifiers
- **700+** Ethically Sourced Personal and Professional Attributes

Real Control

Leverage your **FullContact Private Identity Cloud™** to protect & control your first-party data across your enterprise. Enable permission-based partnering without commingling data.

Port the **PersonID** across your ecosystem improving targeting, reach, recognition and measurement.

Manage privacy and permission at a person level at every touchpoint.

- **SOC2 Type II Compliant**

Real Time

Recognize people across platforms and engagement **in the moments that matter.**

Leverage our machine learning, applied graph theory, and distributed computing to **improve resolution.**

High availability, high throughput, and resilient low latency architecture.

- **30+ Million** Updates per Day
- **40 Millisecond** Response Time