# How Predictive Analytics Can Be Used in Fraud Detection and Prevention

A person in a sweater

Description automatically generated

**Hrvoje Smolic**

-

21/03/2023

# Fraud Detection and Prevention

Fraud is a global problem that can have far-reaching consequences, affecting individuals, businesses, and society. From identity theft to insurance scams to healthcare fraud, fraudulent activities can take many forms and cost [billions of dollars each year](https://www.sofi.com/blog/fraud-cost-us-consumers-5-8-billion-dollars-last-year/).

Not only are online frauds getting more sophisticated, but they're also continuously increasing in number and show no signs of slowing down. For example, a[report](https://go.recordedfuture.com/hubfs/reports/cta-2023-0117.pdf) published by Recorded Future shows that about[60 million payment](https://www.recordedfuture.com/annual-payment-fraud-intelligence-report-2022?__hstc=156209188.71eea1b72912c2ea1942d6558e3a5c57.1676546108185.1676546108185.1676546108185.1&__hssc=156209188.1.1676546108185&__hsfp=557688205#:~:text=These%20figures%20were%20considerably%20lower%20than%20the%2060%20million%20CNP%20and%2036%20million%20CP%20records%20posted%20for%20sale%20in%202021.) card records were compromised in 2022 and were posted on the dark web for sale.

* [Fraud Detection and Prevention](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#fraud-detection-and-prevention)
  + [What is Predictive Analytics?](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#what-is-predictive-analytics)
  + [Challenges of Using Predictive Analytics in Fraud Detection](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#challenges-of-using-predictive-analytics-in-fraud-detection)
    - [Changing Fraud Patterns Over Time](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#changing-fraud-patterns-over-time)
    - [Ever-Changing IP Address Space](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#ever-changing-ip-address-space)
    - [High Rate of False Positive Alarms](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#high-rate-of-false-positive-alarms)
    - [Lack of Awareness of the Environment](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#lack-of-awareness-of-the-environment)
    - [Diverse and Sophisticated Attacks](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#diverse-and-sophisticated-attacks)
  + [How to Use Predictive Analytics for Fraud Detection and Prevention](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#how-to-use-predictive-analytics-for-fraud-detection-and-prevention)
    - [Conduct a SWOT Analysis](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#conduct-a-swot-analysis)
    - [Build a Dedicated Fraud Management Team](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#build-a-dedicated-fraud-management-team)
    - [Identify the Scope of the Problem and Related data.](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#identify-the-scope-of-the-problem-and-related-data)
    - [Prepare the Data](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#prepare-the-data)
    - [Define the Fraud Detection Model](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#define-the-fraud-detection-model)
    - [Train the Model](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#train-the-model)
    - [Validate the Model](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#validate-the-model)
    - [Implement the Model](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#implement-the-model)
    - [Monitor and Improve the Model](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#monitor-and-improve-the-model)
  + [Predictive Analytics Techniques to Use](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#predictive-analytics-techniques-to-use)
    - [Logistic Regression](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#logistic-regression)
    - [Decision Tree](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#decision-tree)
    - [Neural Networks](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#neural-networks)
    - [Ensembles Methods](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#ensembles-methods)
  + [Best Practices for Using Predictive Analytics to Prevent Fraud](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#best-practices-for-using-predictive-analytics-to-prevent-fraud)
  + [Advantages of Predictive Analytics in Fraud Prevention](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#advantages-of-predictive-analytics-in-fraud-prevention)
    - [Faster Fraud Detection](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#faster-fraud-detection)
    - [More Accurate Fraud Detection](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#more-accurate-fraud-detection)
    - [Reduced Costs](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#reduced-costs)
    - [Fewer Human Interventions](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#fewer-human-interventions)
    - [Proactive Fraud Detection](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#proactive-fraud-detection)
    - [Competitive Advantage](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#competitive-advantage)
  + [Final Words](https://graphite-note.com/predictive-analytics-fraud-detection-and-prevention#final-words)

Traditional fraud detection methods, such as manual audits and investigations, have limitations in speed, accuracy, and scalability. That's where [predictive analytics](https://graphite-note.com/what-is-predictive-analytics-benefit-business) comes into place, as it provides a new avenue for detecting and preventing fraud.

In this article, we'll discuss how predictive analytics can be used in fraud detection and prevention in detail. It'll help you understand how you can use this technique to improve the security posture of your organization and protect it from fraudsters and other malicious online actors.

## What is Predictive Analytics?

Predictive analytics refers to the use of historical and current data to make predictions about the future by employing statistical modeling, data mining, and machine learning algorithms. It helps you identify patterns in large datasets, uncover hidden risks of fraud, and take proactive action to prevent them.

This allows you to identify and act on fraudulent activities quickly, reducing losses, protecting assets, and maintaining the trust of customers.

## Challenges of Using Predictive Analytics in Fraud Detection

Before getting into the details of using predictive analytics for fraud detection and prevention, let's discuss the main challenges associated with this technique. It'll help you understand how to improve your predictive analytics system and safeguard your business against potential financial losses.

### Changing Fraud Patterns Over Time

Fraudsters are constantly evolving and finding new ways to get around the systems to commit fraudulent activities. This makes it difficult for predictive analytics to keep up with the developed patterns and detect fraud.

Therefore, machine learning models need to be continually updated to remain efficient and effective. Failure to update the models can result in a decrease in performance, rendering them useless.

### Ever-Changing IP Address Space

Fraudsters and hackers use various tools, such as TOR (The Onion Router) and a[VPN service for mobile](https://cybernews.com/best-vpn/vpn-for-iphone/), computer, or other devices they use, to change their IP addresses and locations. The use of bots further complicates the prediction process. These techniques make it hard for systems to detect the true location of malicious actors.

### High Rate of False Positive Alarms

Despite the use of good[SIEM (Security Information and Event Management)](https://www.ibm.com/topics/siem) systems, high rates of [false positive alarms](https://csrc.nist.gov/glossary/term/false_positive) still need to be solved. Analysts need to spend a lot of time accumulating and correlating true events from the large data pool.

It's essential to eliminate such events before quantifying the data towards machine learning, which increases workload and distracts security teams from real security threats.

### Lack of Awareness of the Environment

Many organizations are unaware of their environment, especially outside the financial sector. That's because it's no less than a challenge to maintain and keep track of the additions or upgrades made to the hardware or software aspects of products in a large organization.

The negligence and missing components in inventory contribute to the difficulty of forecasting fraud. Therefore, it's critical for organizations to be fully aware of the environment to identify potential vulnerabilities and threats that could lead to fraud.

### Diverse and Sophisticated Attacks

The range of exploits can still be reasonably managed, but the fact that every mind is different adds another layer of complexity. Remember that the use of an exploit differs from person to person, and the way new malware and sophisticated attacks are executed can be challenging to detect.

So, predictive analytics models need to take into account the various ways in which fraudsters can exploit system vulnerabilities and attempt to prevent them.



## How to Use Predictive Analytics for Fraud Detection and Prevention

Predictive analytics has emerged as an effective technique to detect and prevent fraud, and here's how it can be achieved.

### Conduct a SWOT Analysis

Conducting a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis helps you evaluate your current position, identify potential opportunities, and anticipate potential threats.

It'll help you identify the "scope of the problem and related data" (discussed below) to devise a tailor-made system that will provide more accurate predictions of fraudulent activities.

### Build a Dedicated Fraud Management Team

The next step in using predictive analytics for fraud detection and prevention is to build a dedicated fraud management team. It should be equipped with the necessary expertise, tools, and workflows to effectively detect and prevent fraud in the organization.

Keep the following points in mind while creating a fraud management team.

* Identify the roles and responsibilities of each team member carefully based on your organization's specific needs and risk areas.
* Every team member should be well-versed in data analytics, machine learning, and artificial intelligence techniques.
* Train the team so that every member has a clear understanding of the different types of fraud, their modus operandi, and the impact they can have on the organization.

### Identify the Scope of the Problem and Related data.

Once you're done with SWOT analysis and team building, you need to identify the scope of the problem and related data. Start by determining what activities constitute fraud in your organization and what fraud types are most likely to occur. This can be achieved by analyzing past incidents and identifying common patterns and indicators.

Once the problem is defined, the next step is to identify the relevant data sources that can provide insights into fraudulent activities. This includes data from various sources such as financial transactions, user activities, and social media. It's vital to ensure that the data is of high quality and reliable, as poor quality data can lead to inaccurate predictions.

### Prepare the Data

Now you need to prepare the data for analysis by cleaning, transforming, and integrating it to create a comprehensive dataset. This step is critical to ensure the accuracy and effectiveness of the predictive model.

The data preparation process can involve several steps, such as removing missing values, handling outliers, and normalizing the data.

### Define the Fraud Detection Model

The next step is to define the fraud detection model. This involves selecting the appropriate statistical algorithms and machine learning techniques that can be used to detect fraudulent activities. The selection of the model will depend on the nature of the data and the specific problem that needs to be addressed.

### Train the Model

Once the fraud detection model is defined, the next step is to train it. This involves using historical data and creating a baseline that will help you determine when an event or scenario will be considered "fraudulent."

The model needs to be trained on a diverse set of data to ensure that it can detect all possible scenarios. You can improve the accuracy of the model by tuning its parameters and using more data.

### Validate the Model

After the model is trained, you need to test and validate it to ensure that it can accurately detect fraudulent activities. Use a separate dataset that was not used during the training process to determine the effectiveness of the model in real-world scenarios. Then analyze the results of the validation to refine and adjust the model.

### Implement the Model

After the model is validated, you can implement it in your organization's existing security system to perform fraud detection with predictive analytics in real time. Make sure that your model generates alerts or notifications whenever it detects suspicious activities to keep you informed.

### Monitor and Improve the Model

Once the fraud detection model has been implemented, it is crucial to monitor its performance continuously. This allows for the identification of any potential issues, like false alarms or inaccuracies, that could arise from the model's output.

This way, you'll be able to improve the model to increase its accuracy and effectiveness in detecting new and evolving fraud patterns.

**No-code predictive analytics**

Predictive Mastery: No Crystal Ball Required!

[No-code machine learning](https://graphite-note.com/) for everyday business users.

[Try it for free](https://app.graphite-note.com/#/signup)

## Predictive Analytics Techniques to Use

There are several predictive analytics techniques that can be used for fraud detection, including the following.

### Logistic Regression

Logistic regression is a[statistical method](https://www.techtarget.com/searchbusinessanalytics/definition/logistic-regression#:~:text=Logistic%20regression%20is%20a%20statistical,or%20more%20existing%20independent%20variables.) that is used to analyze a dataset with one or more independent variables that determine an outcome. It can be used to predict the probability of a binary outcome, such as fraud or no fraud.

The algorithm calculates the relationship between the independent variables and the outcome and produces a logistic function that estimates the probability of fraud.

### Decision Tree

A[decision tree](https://www.mastersindatascience.org/learning/machine-learning-algorithms/decision-tree/) is a graphical representation of all possible solutions to a problem based on a given set of conditions. It's often used in fraud detection to create a hierarchical model that identifies the conditions that are most likely to lead to fraud.

The algorithm produces a tree-like structure that classifies the data based on a set of "if-then" rules, where each node represents a condition, and each branch represents an outcome.

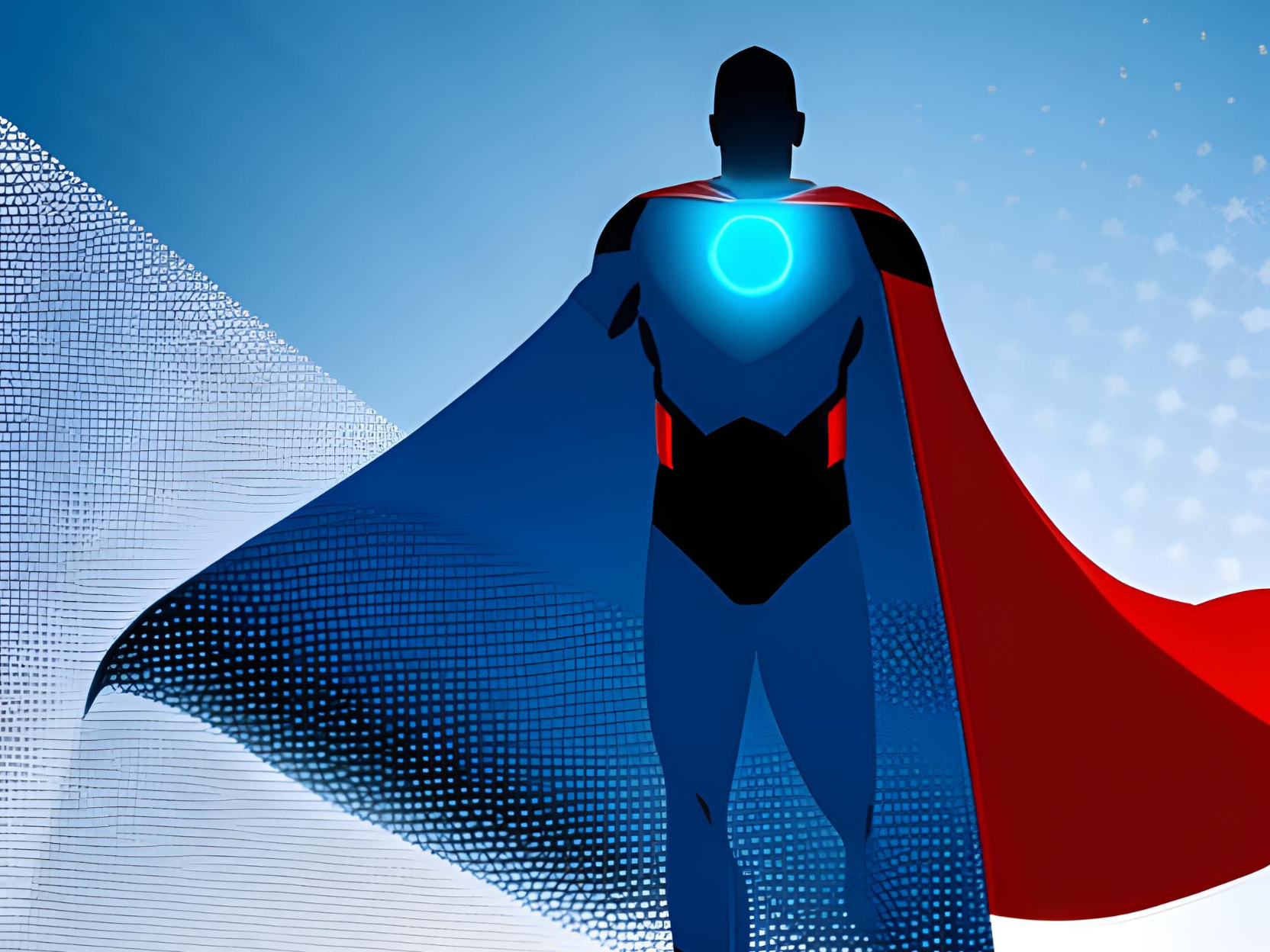
### Neural Networks

[Neural networks](https://www.expressanalytics.com/blog/neural-networks-prediction/) are a type of machine learning algorithm that is inspired by the structure and function of the human brain. You can use them to create a fraud detection model that can learn patterns in the data and make predictions based on them. The algorithm creates a network of interconnected nodes that are trained on a dataset to identify fraudulent patterns.

### Ensembles Methods

Ensemble methods are a group of machine learning algorithms that combine multiple models to improve the accuracy and stability of the predictions.[Bagging and boosting](https://analyticsindiamag.com/primer-ensemble-learning-bagging-boosting/) are two powerful ensemble methods that can be used in fraud detection.

Bagging involves creating multiple models with different subsets of the data and combining their predictions, whereas boosting is about combining various weak models to create a strong model that can make accurate predictions.

[](https://cdn.graphite-note.com/wp-content/uploads/2023/03/How-Predictive-Analytics-Can-Be-Used-in-Fraud-Detection-and-Prevention-superpowers.png?x87235)

## Best Practices for Using Predictive Analytics to Prevent Fraud

The following are some important tips that you can use to start using predictive analytics to detect and prevent fraud.

* **Start with a Clear Definition of Fraud:** Before beginning the analysis, it's essential to define what constitutes fraud in your organization. This will help you identify suitable data sources and select the appropriate models for detecting fraudulent activities.
* **Use High-Quality Data:**Data quality is critical to the accuracy of predictive models. Ensure that your data is complete, accurate, and up-to-date. Cleaning and preprocessing the data is important to ensure that it is suitable for use in predictive models.
* **Choose the Right Algorithms:**Select the right algorithms for the problem at hand. Different algorithms are better suited to different types of fraud and data. Experiment with different algorithms to determine which ones work best for your organization.
* **Combine Multiple Models:** Consider combining multiple models to improve the accuracy of the fraud detection system. Ensemble methods such as Bagging and boosting can be used to connect the results of multiple models and improve the overall accuracy.
* **Monitor and Refine the Models:** The performance of the predictive models should be monitored regularly to ensure that they are detecting new and evolving fraud patterns. The models can be refined and updated as new data becomes available.
* **Involve Subject Matter Experts:** It's crucial to involve subject matter experts such as fraud investigators and analysts in the analysis process. They can provide valuable insights into the types of fraud that are common in the organization and help identify new patterns of fraudulent activity.

## Advantages of Predictive Analytics in Fraud Prevention

The following list of advantages will help you understand how predictive analytics improves fraud detection and prevention.

### Faster Fraud Detection

With the increasing number of transactions and data generated every day, it can be challenging for human analysts to detect fraudulent activities quickly. However, predictive analytics algorithms can analyze large amounts of data in real-time and identify fraudulent patterns almost instantaneously.

It allows you to act quickly and prevent fraudulent activities before they cause any significant financial harm.

### More Accurate Fraud Detection

In addition to being faster, predictive analytics delivers more accurate results than a human agent can do. By processing big data, digital tools have access to more information to make decisions with greater accuracy.

### Reduced Costs

While the cost of creating a predictive analytics system for fraud detection can be very high, it can actually save you money in the long run. That's because it can help you reduce costs associated with fraud by detecting it early, avoiding losses, and reducing the need for expensive investigations.

### Fewer Human Interventions

By maximizing the use of technology and predictive analytics, you can reduce the number of manual interventions required. This reduces turnaround times and frees up employees, allowing them to focus on more valuable, high-impact tasks. This means fewer [human errors](https://www.grcelearning.com/blog/human-error-is-responsible-for-85-of-data-breaches), which can also reduce other security issues.

### Proactive Fraud Detection

A purely reactive approach to fraud prevention is no longer sufficient. Predictive analytics brings in more opportunities to conduct proactive fraud detection initiatives. It can help you identify the root causes of fraudulent activities and combat them proactively.

### Competitive Advantage

By implementing predictive analytics for fraud prevention, you can gain a competitive advantage over your peers. The ability to prevent fraud and ensure a secure environment can enhance your organization's reputation and build trust with customers, partners, and stakeholders.

It can help your business stand out in the marketplace and gain a competitive edge, leading to increased customer acquisition and retention.

## Final Words

Predictive analytics is a powerful tool for detecting and preventing fraud in today's data-driven world. More and more organizations are investing in this technology to stay ahead of fraudsters and protect their businesses.

As technology continues to advance, predictive analytics will continue to play a critical role in fraud prevention, ensuring that businesses and consumers can operate in a secure and reliable environment.

# **Using Predictive Analytics for e-Commerce Fraud Detection**

Predictive analytics may be one of the most effective ways an e-commerce merchant can detect fraud. By analyzing past data, trends, and variables, merchants can build smarter fraud score algorithms and model more accurate fraud prevention strategies.

ClearSale executive vice president Rafael Lourenco walks through a six-step process for using predictive analytics to stop e-commerce fraud in its tracks.

To use predictive analytics for e-commerce fraud detection:

1. Define your historical database of online orders.
2. Define which transactions to include your order database.
3. Analyze the fraud patterns and variables.
4. Model the collected data.
5. Implement the model online.
6. Monitor your results.

Let’s explore these in greater detail.

## 1. Define Your Historical Database

As with any other [**analytical process**](https://blog.clear.sale/clearsales-unique-approach-to-fraud-analytics-and-manual-reviews), Lourenco says, you have to start with a historical database of online order details. When building this database, think carefully about the timeframe for orders. Will your database include six months’ worth of orders? A year of orders? Lourenco encourages using at least one year of data — if available — to build a holistic picture of any sales seasonalities.

On the other hand, relying only on recent orders can be a problem. The reason? Chargebacks and fraud [**can take weeks**](https://blog.clear.sale/what-is-the-time-limit-on-chargebacks) — even months — to show up on a balance sheet. This means trying to predict future sales based on yesterday’s orders can be risky, if not impossible.

Instead, he recommends building in a safe window of at least 90 days. For instance, if it happens to be April when you begin building your database, consider including orders that were made between January to December of the previous year, and eliminate orders from January to March of the current year, as the full picture of chargebacks and fraud from this time period have [**yet to develop**](https://blog.clear.sale/the-chargeback-lag-what-it-is-and-how-it-impacts-your-e-commerce-business).

## 2. Define Which Transactions to Include in Your Order Database

Once you’ve established the timeframe for your analytics, it’s time to define the “what”: What transactions do you include in this historical database?

Chargebacks happen only with certain payment methods, like credit cards. Therefore, don’t include payments made via ACH or any other method that doesn’t create a chargeback.

In fact, adding these payments to your database would likely end up creating bias in your analytics. If ACH is more common in Region A than it is in Region B and you’re including ACH payments in your database, then it will seem like Region A is a safer place in which to do business. But in actuality, Region A’s decreased risk is only because ACH doesn’t result in chargebacks and is at a lesser risk for fraud. So when you’re trying to predict chargebacks and fraud, look only at the payment methods that are contributing most to that risk.

After you define the “what,” you must be able to translate that into a binary variable, such as whether orders made in a certain time period and processed via a certain payment method can ultimately be identified as good or bad orders. In the simplest of terms, every order your system approves and doesn’t come back as a chargeback should be viewed as a good order. Every order that is approved and comes back as a chargeback should be viewed as a bad order.

But here’s the catch: You can’t use chargebacks as the only definition of bad orders. Say your fraud filters avoid every order placed from or delivered to Russia. That means you’ll never experience any chargebacks there, right? Sounds like good news. However, not every order coming from or going to Russia is fraudulent. Therefore, your strategy means you’re declining at least some legitimate orders— and that affects your ability to identify future patterns and variables associated with both good and bad orders.

On the other hand, if you don’t automatically block all orders to and from Russia, you’ll be able to better understand when an order related to Russia is safe and when it’s not. Yes, your all-or-nothing strategy means you’ll never experience chargebacks on these orders, but you’ll also miss out on countless legitimate sales. Even worse, it makes your fraud database look like there’s no fraud in Russia — and we all know that fraud is everywhere.

But that’s not the only reason automatically declining orders is damaging to your business. When you decline orders, you end up with what we call “censored data” — or data (in this case, orders) you’ll never truly be able to objectively evaluate. In the case of the declined Russian orders, we’ll never know which ones were actually fraudulent.

To avoid this, you need to clearly define what a good order is and what a bad one is. Lourenco suggests starting by identifying both chargebacks and confirmed fraud orders that you either avoided, declined or blocked. Then define good orders as the transactions that were approved and never came back as chargebacks.

## 3. Analyze the Fraud Patterns and Variables

Now that you’ve set up your database and defined what good and bad orders are, it’s time to analyze. Through descriptive analysis, try to understand the patterns and variables that help with fraud detection and differentiating the good orders from the bad.

As you analyze orders, create some variables and see if they affect the percentage of bad orders. For instance, you might build a weekday variable, so you can see if Mondays have higher or lower fraud rates than Tuesdays. If you find that all weekdays have the same fraud levels, then you know that’s not a worthwhile variable to analyze. On the other hand, if the weekdays each have very different fraud rates, it’s a great variable to include in your analysis.

While Lourenco encourages creativity while creating variables, he emphasizes creating only the variables that can be analyzed the moment an order is placed. After all, you want to know right away — as soon as someone places an order —what the probability is that the order is fraudulent. For example, you might want to know how many orders a particular phone number has placed before, so you analyze all past transactions and evaluate the likelihood the order is a bad one.

## 4. Model the Collected Data

The fourth step is the modeling itself. Here, you’ll most likely need the help of data scientists, as the techniques for modeling range from old-school styles like logistics regression, neural networks and deep learning to more modern, complex approaches.

Regardless of the technique used, your objective here is to translate all the data you’ve gathered into a predictive probability of the likelihood of fraud.

## 5. Implement the Model Online

When you move into the implementation step of fraud detection, you’ll want to be able to assign a [**referral score**](https://blog.clear.sale/e-commerce-fraud-prevention-best-practices-for-using-fraud-scores) to every credit card order that’s placed on your website to predict the likelihood of that order being a fraudulent one.

Note: It’s often best to use the same resource you used in the modeling step to help you implement the model into your decision flow.

## 6. Monitor Your Results

Last, but certainly not least, is monitoring. Is everything working in the real world like you expect it to? You might think your purchases are equally spaced between weekdays and weekends, so you’ve trained your model in a database in which half the orders were placed on weekdays and the other half are placed on weekends. But when you implement your model and follow up on it in real life, you learn that only 10% of orders are placed during weekends. If this should happen, go back and tweak your model until it’s a good fit with the way your online company actually operates — and even then, you’ll need to continue monitoring so you can adapt as your business evolves.

If you’re looking for a trusted partner in fraud detection to help you monitor the constantly changing fraud landscape, you want a partner that focuses exclusively on CNP fraud prevention. Look no further than the experts at ClearSale. Our analysts use a unique combination of human analysis and robust artificial intelligence to identify emerging fraud patterns and stop chargebacks and false declines from taking a toll on your revenues. [**Contact us**](https://www2.clear.sale/contact) to learn how we can make your fraud prevention strategy more effective.

# **Chapter 4: Predictive Analytics for Fraud Detection**

## Introduction

In predictive analytics, the aim is to build an analytical model predicting a target measure of interest (Baesens 2014; Duda et al. 2001; Flach 2012; Han and Kamber 2001; Hastie et al. 2001; Tan et al. 2006). The target is then typically used to steer the learning process during an optimization procedure. Two types of predictive analytics can be distinguished depending on the measurement level of the target: *regression* and *classification*. In regression, the target variable is continuous and varies along a predefined interval. This interval can be limited (e.g., between 0 and 1) or unlimited (e.g., between 0 and infinity). A typical example in a fraud detection setting is predicting the amount of fraud. In classification, the target is categorical which means that it can only take on a limited set of predefined values. In binary classification, only two classes are considered (e.g., fraud versus no-fraud) whereas in multiclass classification, the target can belong to more than two classes (e.g., severe fraud, medium fraud, no fraud).

In fraud detection, both classification and regression models can be used simultaneously. Consider, for example, an insurance fraud setting. The expected loss due to fraud can be calculated as follows

equation

where PF represents the probability of fraud and LGF the loss given fraud. The latter ...