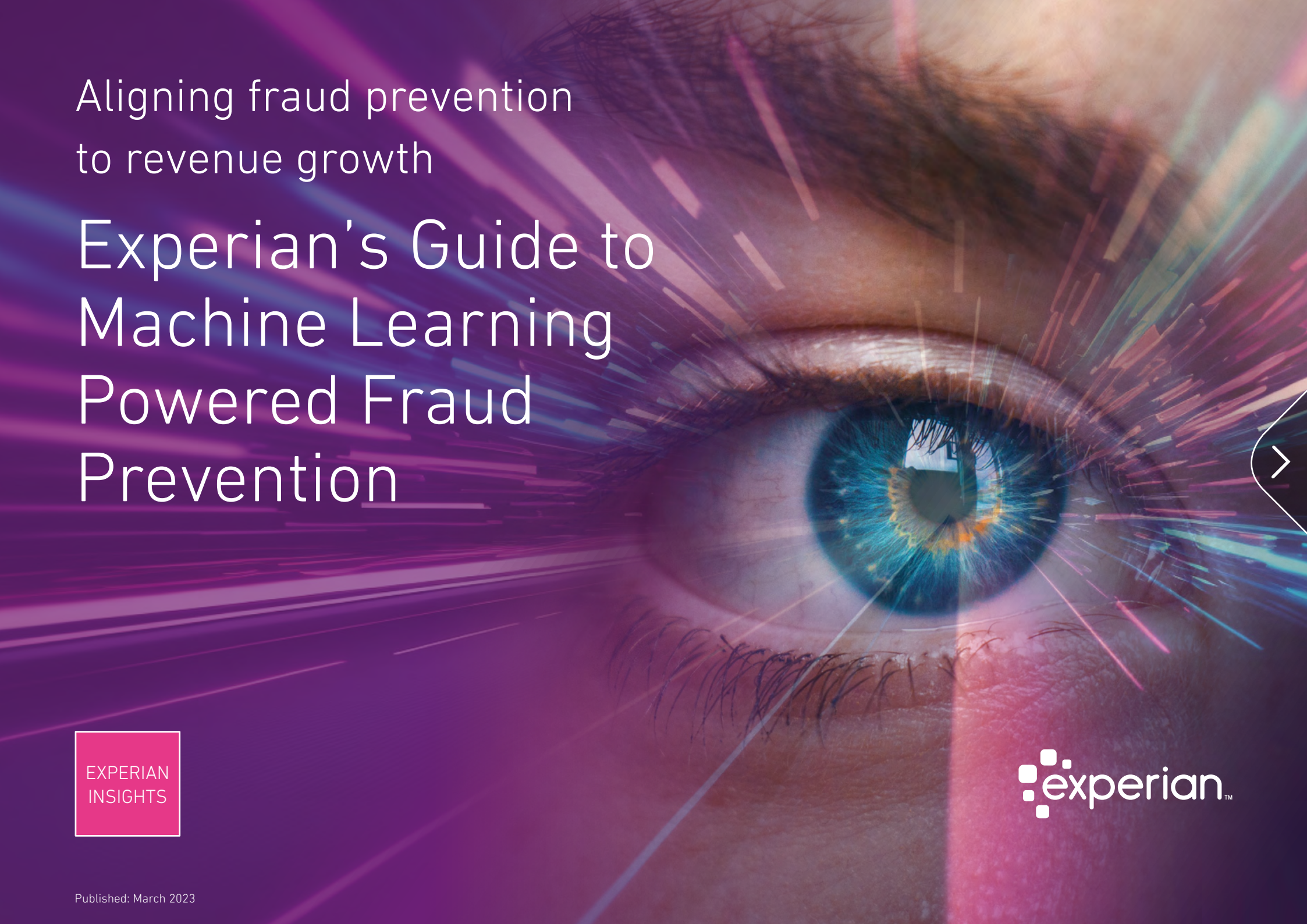
Aligning fraud prevention to revenue growth

Experian's Guide to Machine Learning Powered Fraud Prevention



EXPERIAN INSIGHTS

Published: March 2023

Contents

Fraud inhibits growth. And while digital commerce has flourished due to pandemic-related restrictions so too have fraudsters, taking advantage of the increase in online activity to expand the scale and variety of their attacks. Given the increasing sophistication of these attacks, it is challenging for businesses to stay ahead of the latest threats. A recent study by Juniper Research indicates a 16% growth in ecommerce fraud losses, with $41 billion lost worldwide in 2022.

In a constantly evolving fraud environment, fraud prevention systems that rely on a traditional rules-based approach will struggle to adapt, with rules quickly becoming outdated, causing real customers to be blocked whilst failing to identify the latest fraud patterns.

So how can firms still deliver enhanced fraud prevention without impacting customer experience and conversion? How can they improve fraud detection accuracy when attacks are more sophisticated?

The most powerful solution to meet this challenge is the integration of Machine Learning (ML) into fraud strategies. The complex ML models enable businesses to increase fraud detection whilst more accurately identifying genuine customers.

In this guide, we explore how Artificial Intelligence and Machine Learning are becoming essential attributes in the battle against fraud. We will discuss the whole process in an accessible way to give you a clear understanding of how this technology can help you reduce your fraud rate while growing revenue.

This report references fraud terminology and common fraud types. To aid understanding, we have created a glossary of ML fraud terminology and an explanation of the common types of fraud that online businesses are faced with.

Machine Learning is the cutting edge of fraud detection and prevention; in response to the ongoing surge in cybercrime this technology can provide businesses with unprecedented levels of accuracy in identifying and differentiating between genuine customers and fraudsters.

Luciano Scalise
MD Decision Analytics, Experian
EMEA & APAC

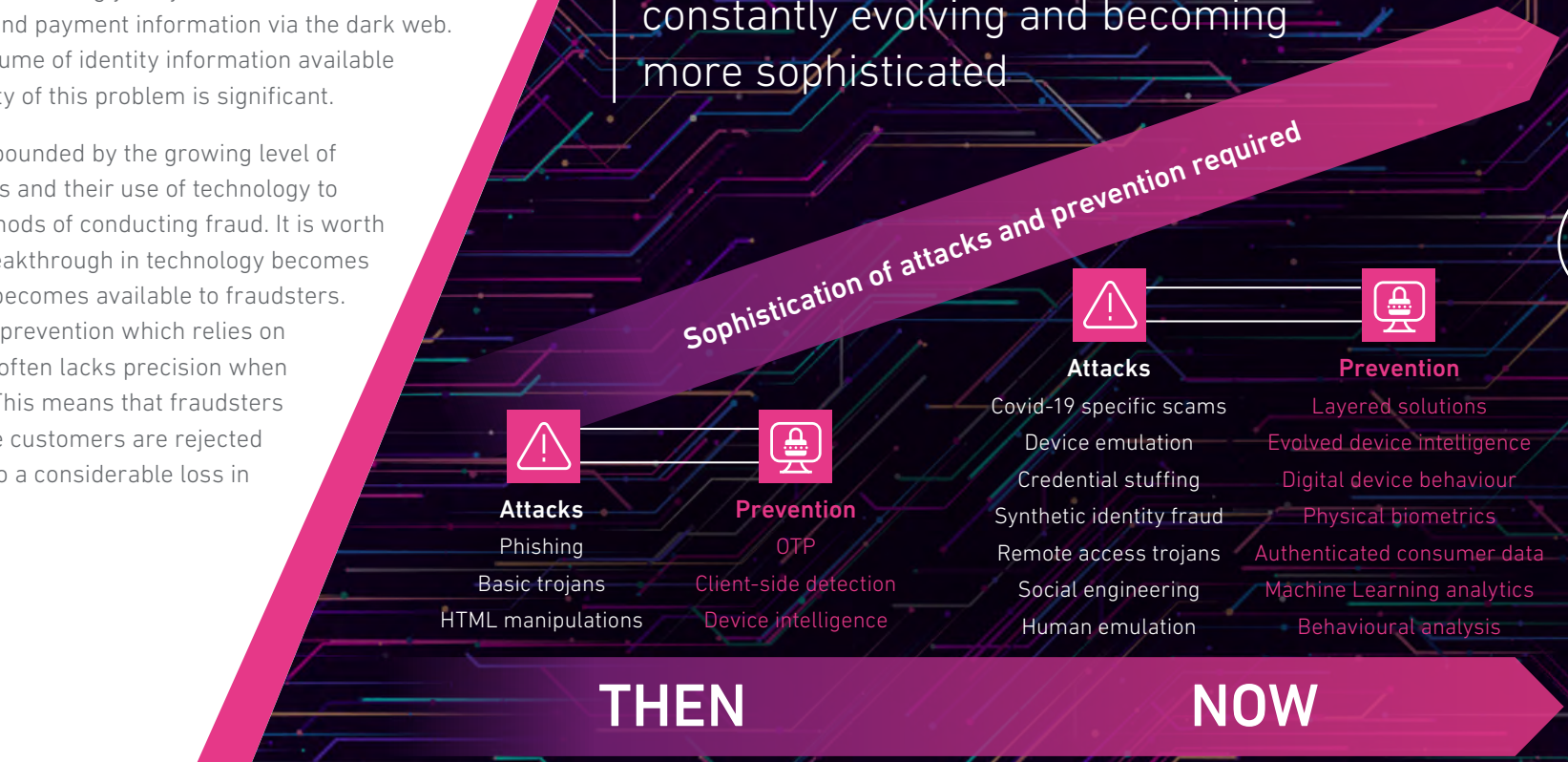
Please note that we use Machine Learning and Artificial Intelligence with the abbreviations of ML and AI interchangeably in this guide.

Sophisticated threats lead to increased losses

There has been considerable growth in global cyber fraud during the last year. The catalyst for this upswing is the seemingly easy access fraudsters have to stolen credit card, bank account, and payment information via the dark web. When combined with the large volume of identity information available through data breaches, the severity of this problem is significant.

The extent of this problem is compounded by the growing level of organisation within fraud networks and their use of technology to create increasingly elaborate methods of conducting fraud. It is worth remembering that whenever a breakthrough in technology becomes available to organisations, it also becomes available to fraudsters. It is therefore apparent that fraud prevention which relies on hard-coded rules-based systems often lacks precision when dealing with sophisticated fraud. This means that fraudsters slip through while many legitimate customers are rejected through false positives – leading to a considerable loss in potential earnings.

Fighting fraud is difficult, with attacks constantly evolving and becoming more sophisticated



Fraud is a problem but so are false rejections

There is no doubt that fraud is a heavy cost for businesses. However, the effort to prevent fraudulent transactions has arguably fuelled an even bigger problem – legitimate customers that are erroneously rejected. In fact, research suggests that the cost of false rejections can often amount to significantly more than the value of fraud losses. This highlights the scale of the problem – whereby firms are not only facing losses due to increased fraud but also considerable loss of revenue due to false positives from inaccurate fraud prevention systems.

One of the biggest pain points for identity and fraud decision-makers is the battle between fraud prevention and revenue generation.

ALMOST 1 IN 3 FIRMS find an inability to align fraud prevention and revenue growth strategies as their top challenge preventing them from successfully managing the costs and risks of fraud*.

Too many fraud prevention barriers mean that a large proportion of good customers are wrongly rejected, which negatively impacts revenue. However, too little focus on prevention, and fraud losses spike.

* Base: 587 EMEA decision makers at Financial Services and Telco firms
Source: A commissioned study conducted by Forrester Consulting on behalf of Experian, August 2022

Customer expectations of low friction vs. increased fraud threats

For high-growth businesses, strong conversion is incredibly important to meet revenue targets. Conversion is essentially when a user completes the desired action online, such as making a purchase or completing an application form. Fraud prevention is therefore a delicate customer experience (CX) balance, with some businesses in the ecommerce sector willing to accept higher fraud rates to keep conversion rates high. But, as the tactics used by fraudsters become more sophisticated, this approach must change – otherwise a greater proportion of revenue will be lost to fraud.

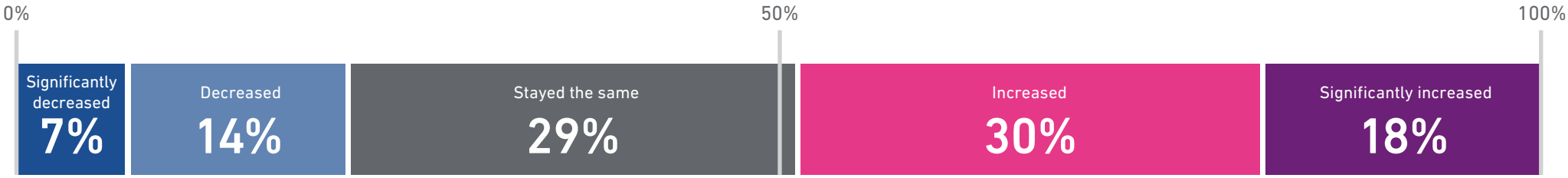
Fraud prevention is the top business priority

Our research indicates a recent shift in business focus. In the Experian EMEA 2022 business and consumer survey, 'investing to improve protection against fraud' was a top priority for 73% of business respondents. Making fraud prevention the number one priority for 2022. Fraud has always been a priority, but the steady increase in losses is driving increased focus from senior leaders. With year-on-year fraud losses rising for 48% of businesses, they recognise that something must be done to combat this problem.

We know how difficult it is to find the right balance between allowing genuine customers to complete their purchase/application and stopping fraudsters.

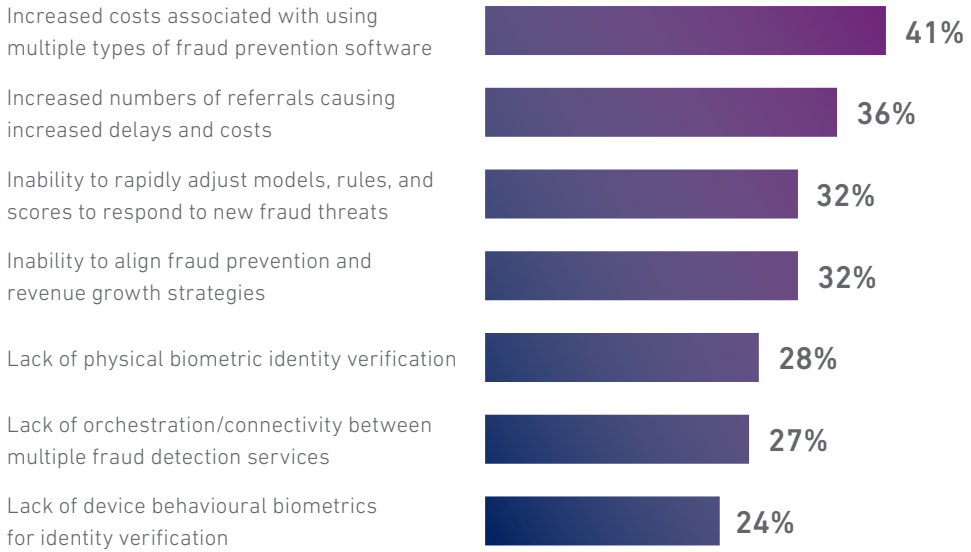
To address this challenge businesses are increasingly prioritising investments in fraud prevention and cybersecurity. But what is the best area to place this investment? This brings us back to advanced analytics, and specifically Machine Learning.

Fraud losses performance over the past 12 months (%)



Base: 587 EMEA decision makers at Financial Services and Telco firms
Source: A commissioned study conducted by Forrester Consulting on behalf of Experian, August 2022.
Note: Percentages may not equal 100% due to rounding.

Top challenges preventing businesses from successfully managing the costs and risks of increased fraud



Base: 587 EMEA decision makers at Financial Services and Telco firms
Source: A commissioned study conducted by Forrester Consulting on behalf of Experian, August 2022.
Note: Percentages may not equal 100% due to rounding.

It's evident that a growing number of businesses are finding it difficult to respond to the increased risk of fraud. Using the right ML fraud prevention system can help businesses deal with all of these issues by accurately detecting fraudsters, reducing both referrals and false positives, and providing a single platform to manage multiple types of fraud prevention software. Another benefit of this approach is that it improves customer experience by reducing the friction involved with a purchase or application.

The role of Machine Learning in identifying fraud

The easiest way to understand how Machine Learning models work is to consider how we learn as humans. Before we undertake any task, we go through a process of gathering information. We use this information and our experience to gain the knowledge we need to complete the task. ML is based on the same principle in that the Artificial Intelligence learns from experience in order to identify the right combination of features that give a certain result.

To identify a suspected fraudster, ML models analyse vast numbers of transactions and learn what combination of features is most likely to result in fraud.

Each of these features is given a weight that indicates its importance when integrating all of these dynamic rules together. Newly collected data is constantly fed into the system to ensure that the ML model can quickly adapt to new fraud threats.

Machine Learning vs. rules-based fraud prevention systems

In the past, many companies relied on a purely rules-based system to prevent fraud. As fraud techniques have become more sophisticated the inadequacies of this type of system have become more apparent. Rules-based systems quickly become convoluted and contradictory, increasing the false positive rate and number of manual reviews. They are also limited by human understanding as they are created manually and can miss subtle correlations in the data, especially in newly emerging fraud patterns.

Another problem associated with rules-based fraud prevention systems occurs during peak events, such as 'Black Friday'. At times like these when transaction volumes increase dramatically there is a corresponding rise in manual reviews, as rule thresholds are exceeded. Consequently, businesses that solely rely on rule-based fraud prevention need to have additional fraud agents available to avoid large backlogs of review cases. In contrast, Machine Learning fraud detection can consistently maintain the same high levels of accuracy with no added manual effort required during sales peaks.

ML models consist of a set of algorithms that combine all its known features to assess each transaction. The value of an ML model is that it is trained to continuously query the features and compares the results in order to make the best possible fraud assessment. This happens in a matter of milliseconds. The power of ML is that it can identify patterns and make inferences from previously unconnected data.

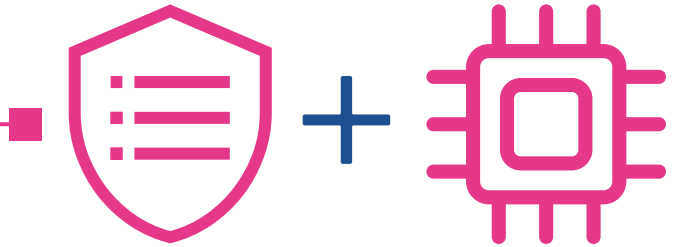
So, for example, if a prospective customer requests a different shipping address to their billing address, this is often very normal. However, in certain circumstances, it could be linked to suspicious and ultimately fraudulent activity. It may depend on the specific geolocations of those two addresses. Machine Learning will learn this and flag those specific address combinations that have shown a pattern of fraud, rather than a more binary rule that may impact 'good' customers too. This ability to understand nuances in the data is the key to how ML detects fraud and what sets it apart from other types of fraud prevention technology, such as hard-coded rulesets.

Machine Learning	Rules-based
ML uses AI to identify subtle patterns and correlations to create features.	Fraud specialist identifies fraud patterns and correlations and then creates rules.
A probability score from 0 to 100 allows the user to set risk appetite.	More rigid thresholds and workflows that lack precision.
Automatic adjustment of features as new fraud techniques emerge.	Manual adjustment of rules as new fraud techniques emerge.
Proactive and agile analysis of real-time fraud trends.	Reactive pre-programmed rules require constant input to stay relevant.
Scalability and cost: Algorithms become more efficient with large data sets.	Difficult to scale and more expensive to maintain as data set grows.

The solution – combining rules and Machine Learning

Although ML has vastly improved fraud prevention systems, there is still a place for unambiguous rules. For example, blacklisted devices should immediately be prevented from making transactions/applications on an online platform. A best-practice approach means combining adaptive ML models with some traditional rules, allowing companies to achieve far greater decision accuracy and flexibility to adapt to changing fraud patterns.

Another fraud prevention technique that can be integrated into an ML-based system to improve detection accuracy is device fingerprinting. This is normally used to identify customers through a range of browser and device data points. However, when combined with ML this information can be used to identify suspicious behaviour by dynamically evaluating the different device attributes.



A step change in fraud detection: the ML process

1

Collect the data

ML models need a large data set of both fraudulent and non-fraudulent labelled transactions to train and test the model.

2

Clean the data

Before the data can be used it must be carefully analysed to remove any inconsistent or biased data.

3

Select an algorithm

There are a variety of different algorithms (e.g. decision tree, logistic regression) that can be used for fraud detection, selecting the appropriate combination depends on the application.

4

Train the model

This involves feeding the clean data into the algorithm and adjusting the model's parameters to minimise the error rate.

5

Evaluate the model

Once the model is trained its performance must be evaluated by using a test dataset which is different from the training dataset. This provides you with a metric to compare the model's predictions with known outcomes.

6

Fine-tune the model

If the model is not delivering the required level of accuracy the parameters are adjusted and additional algorithms included.

7

Deploy the model

At this stage, the model is ready to be used in a live environment.

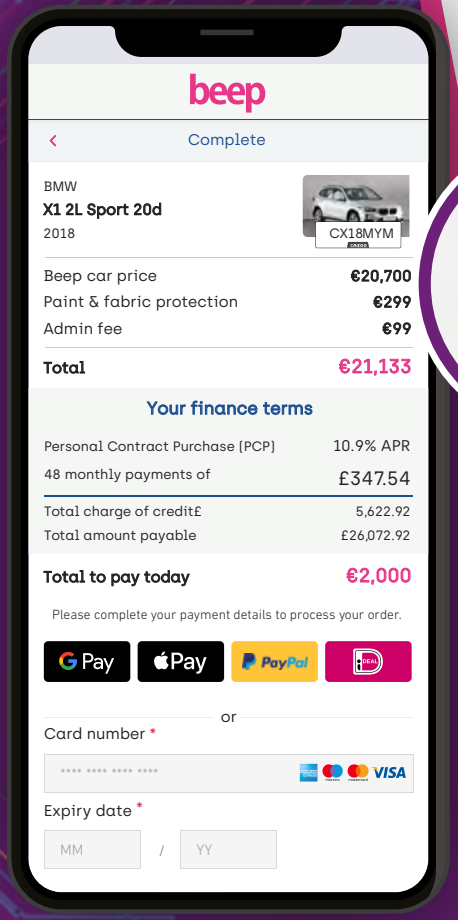
8

Retrain model

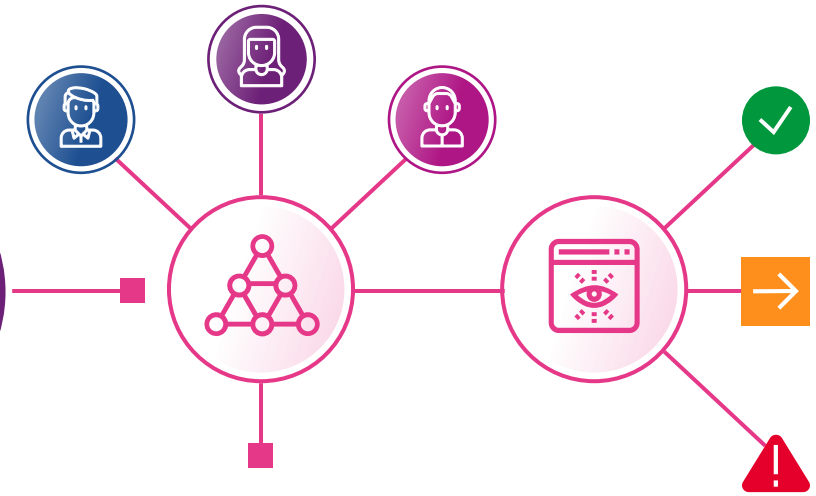
Every transaction that is subsequently fraudulent or flagged for manual review is then used to retrain the model to constantly improve its accuracy.

An ML model in the decision process

Customer begins an online transaction/application.



The model analyses the customer's data in a fraction of a second by comparing it with the previously evaluated data sets.



The model creates a forecast and makes a recommendation based on risk appetite.

Approve

If the transaction/application is approved, it passes through the payment chain or approval process without manual verification.

Review

Review cases are passed into case management where a fraud agent can evaluate them manually.

Reject

Transactions/applications deemed high risk are rejected.

Examples of the type of data used to train an ML model

Confirmed fraud cases

Actual fraudsters that have been detected or rejected.

Credit card chargebacks

Successful transactions subsequently charged back.

Manual reviews

Decisions made by fraud agents.

Device data

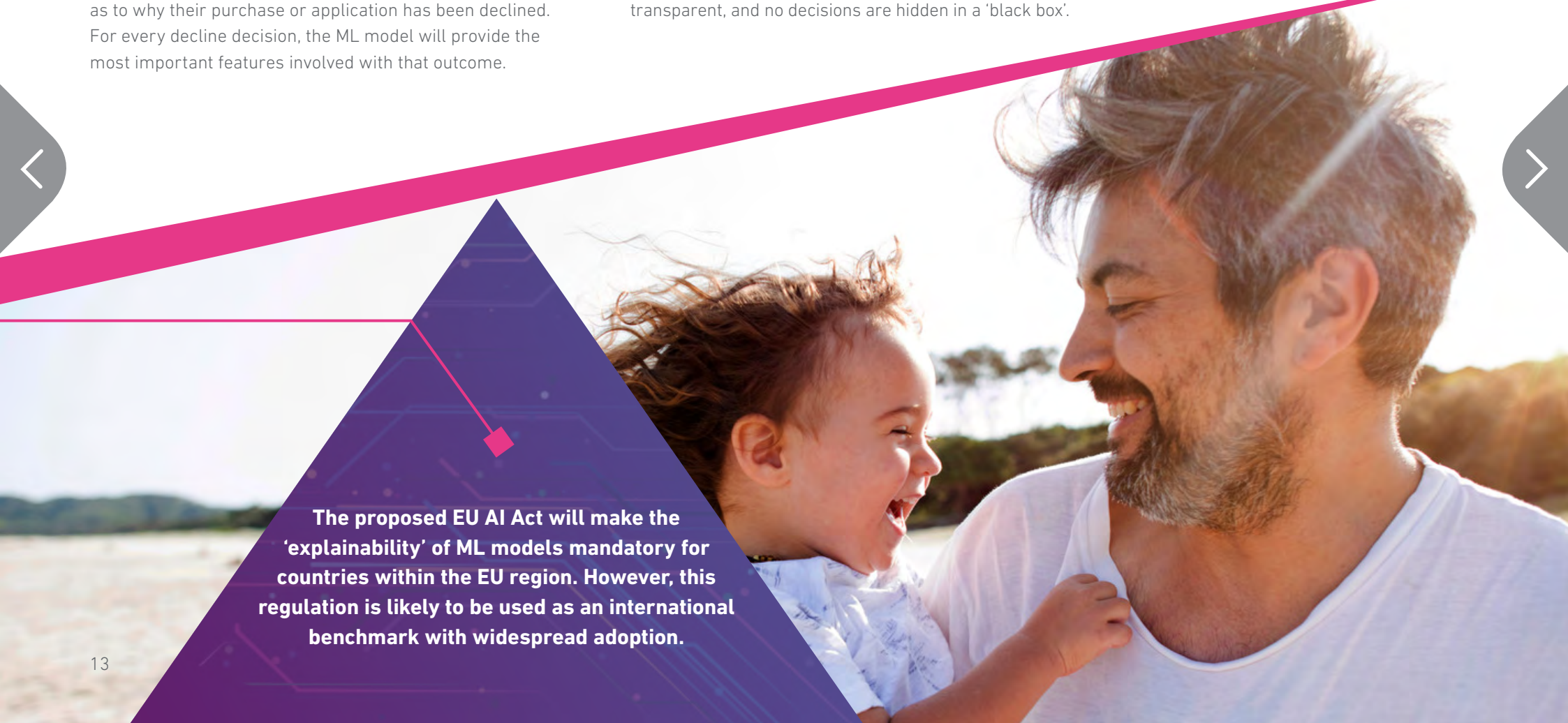
Devices that have been flagged as fraudulent in the past.

The ML model learns from every decision that is made and thus becomes more precise with time.

Explaining Machine Learning outcomes

It's important that any ML model provides users with a summary of the top features that were used to make each decision. This allows businesses to understand the decisions made by the model and provide an explanation to customers as to why their purchase or application has been declined. For every decline decision, the ML model will provide the most important features involved with that outcome.

These features may be issues like a different delivery and billing address or a difference in the browser language and the IP address from the user's device. By stating each feature that has been used to make a decision, the ML model is transparent, and no decisions are hidden in a 'black box'.



The proposed EU AI Act will make the 'explainability' of ML models mandatory for countries within the EU region. However, this regulation is likely to be used as an international benchmark with widespread adoption.

Advantages of Machine Learning based fraud prevention systems

ML models are the most efficient way for companies to predict which of their transactions are fraudulent or are likely to result in chargebacks. This is due to the vast amount of data that can be analysed with Machine Learning.

This efficiency can significantly reduce costs while providing higher levels of accuracy than other approaches to fraud detection. As ML is more effective than humans at identifying fraud patterns and more adaptive than fixed rule sets, it means this approach can drastically lower the false positive rate and allow more genuine customers to complete their purchases/applications.

In addition to this, another major benefit of ML is that the models become more accurate over time by consuming large amounts of additional data. Many online companies generate huge volumes of data that can consistently be used to hone the accuracy of their ML model. The more data that is used to train the model, the more accurate the result will be.

Benefits of Machine Learning fraud prevention

Increased accuracy

1 ML algorithms can analyse enormous quantities of data to detect patterns and trends that are not apparent to human fraud specialists. This leads to a greater level of accuracy than is possible with manual fraud identification methods.

Cost reduction and scalability

2 Scalability is simple to achieve with ML models as an increase in the available data only improves the model. This allows for a wide range of online service providers and merchants to reduce the costs associated with their fraud operations.

Constantly evolving models

3 ML models can continually adapt and improve over time as new fraudulent techniques are identified and blocked. This constant feedback process allows them to stay at the cutting edge of fraud patterns and respond to emerging threats in real-time.

Fewer false positives

4 The accuracy of ML fraud models means that almost all transactions are reliably classified. The benefit is that both genuine customers and fraudsters are identified automatically. Unlike a rigid set of rules, ML models can adapt to changing fraud threats without becoming overly convoluted and misidentifying real customers.

Reduce manual reviews

5 Many online businesses deal with huge numbers of transactions on a daily basis, and this usually requires a large team of fraud specialists to review potentially fraudulent transactions, especially in times of peak demand. The cost and time involved with manual reviews means that 60% of online merchants would prefer to reduce their reliance on reviews or eliminate them entirely.

The improved accuracy of ML models means that far fewer cases require manual review. Once a fraud specialist has evaluated a case the data is added to the model so that in the future similar transactions will not need to be reviewed. This reduces the workload of the fraud team and allows them to spend more time on more complex cases.

Faster decisions with no downtime

6 Even the best fraud specialist can take hours to analyse a complex data set and reach a conclusion. In contrast, ML models can perform this analysis in less than a second and provide a greater level of accuracy. ML systems can maintain this level of precision consistently and continue to deliver accurate decisions non-stop. This is particularly relevant during peak seasons when the volume of transactions is higher than normal.

Automate decisions

7 By using ML in fraud prevention, you automate a greater proportion of transactions, meaning a faster decision and improved customer experience for more customers. ML-based fraud systems can automatically assess huge volumes of transactions while maintaining the same high level of accuracy.

Challenges of implementing Machine Learning in fraud prevention

As with any breakthrough technology, there are a number of key considerations when using ML to prevent fraud. It is vital that businesses are aware of these potential issues to ensure their ML model functions accurately, efficiently and within legal frameworks. We've identified the five most important aspects to consider when starting out with an ML fraud prevention system.

Quality and quantity of data

1 The accuracy of every ML model is dependent on the data that is used to train it. Businesses that are in the process of launching their own ML fraud model need to provide a sufficient volume of clean data to ensure that their algorithms can differentiate between genuine transactions/ applications and fraudulent ones. Without an adequate supply of data, any ML model will be prone to inaccurate predictions.

Explainability

2 The proposed European Union Artificial Intelligence Act will be the first regulation governing the use of AI and is likely to be adopted worldwide as a benchmark for controlling the use of AI in all aspects of our lives. To comply with this legislation, it's essential that businesses use fully transparent ML models that avoid 'black box' algorithms when assessing customers. Businesses that fail to comply face potential prosecution and large fines of up to 30 million euros or 6% of a business's global annual turnover.

Avoiding bias

3 If the data set used to train an ML model includes any bias, then the model will incorporate this bias into its predictive ability. This can result in inaccurate fraud predictions based on the learned bias of the model, such as a specific geographic location or country. The quality of the training data set is key to avoid introducing any potential bias and reducing the accuracy of a model through increased false positives.

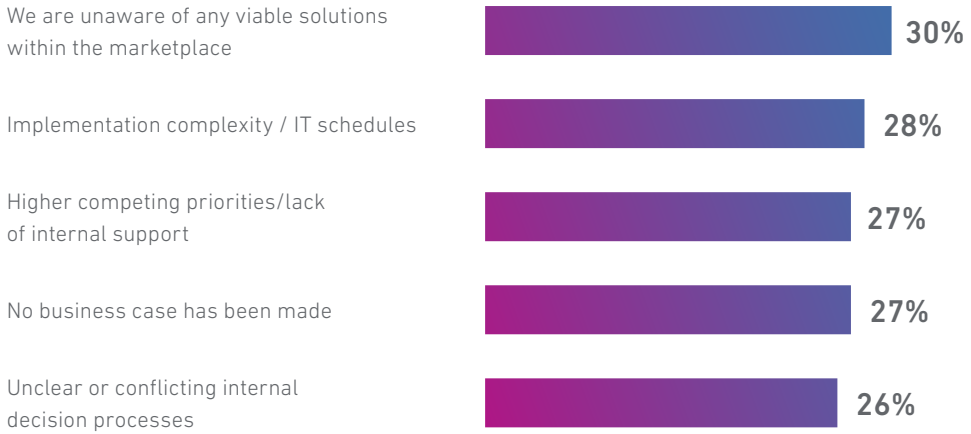
Privacy

4 ML fraud models may involve processing sensitive data such as financial transactions and personal information. It is important that all businesses that intend to use ML for fraud prevention are aware of the applicable laws regarding data privacy and ensure that they have appropriate measures in place to protect the privacy of their potential customers.

Increased IT complexity

5 Our research reveals that the biggest challenge businesses face when adopting AI and ML is the increased IT complexity required to handle this technology. Setting up an ML fraud prevention system is a highly specialised skill set that falls outside of many online businesses' capabilities. The easiest way to overcome this challenge is to partner with experienced experts that understand the operational and legal requirements involved with establishing an ML fraud prevention system.

What obstacles are preventing businesses from implementing Machine Learning models to prevent fraud



Base: 1905 business decision-makers from twenty countries worldwide
Source: A commissioned study conducted by NorthStar Research Partners on behalf of Experian, June 2022

Our research suggests that the biggest obstacle preventing businesses from using ML for fraud detection is a lack of awareness of viable solutions. As the threat of online fraud expands, familiarity with ML will undoubtedly increase as businesses are forced to explore alternatives to rules-based systems.

Once businesses decide to invest in ML they must carefully set out plans for implementation, leveraging external expertise to ensure compliance and successful operational integration. Get these steps right, and the potential benefit in fraud prevention is likely to be worth the investment.

The future of Machine Learning fraud prevention

There is no doubt that ML is going to remain an integral part of fraud prevention in the future. As more sectors mature in their fraud prevention programs there is a possibility that sharing data between organisations and regions – while maintaining privacy – becomes widespread. This global sharing of data is known as collective intelligence and could take our current application of ML fraud prevention to the next level.

Incorporating such vast quantities of data would allow for even greater accuracy in fraud detection. In theory, this could eventually eliminate fraud as all the possible ways to commit fraud were identified and added to the global fraud prevention model. Although this may seem unrealistic the potential for collective intelligence is clear. Consumer sentiment towards data sharing is also improving as the security benefits associated with sharing data with trusted businesses become more widely understood.

According to our global research of over 6,000 consumers worldwide, **56% INDICATED** that they would be willing to allow different companies to share their personal data with each other to ensure greater online security and proactively avoid being the victim of fraud.

As the level of sophistication and the organisation of online fraud syndicates becomes more advanced, the best defence lies in technology like ML fraud identification and collaboration. Collective intelligence sharing between businesses could help them stay ahead of fraudsters and reduce the impact of fraud.

How can Experian help your business detect fraud?

As a leading provider of data, advanced analytics, and fraud prevention software, Experian can help your business develop and implement an ML model specifically designed for your circumstances. We operate across 45 countries and have the experience and expertise to help you through each stage of establishing and maintaining your ML fraud prevention process. We ensure your ML model is fully transparent with explainable outcomes to ensure compliance with local regulations.

Experian has a range of fraud prevention solutions, designed to help clients find the optimal balance between fraud prevention and a seamless customer experience. Our latest fraud solution, Aidrian, is a continually improving fraud solution powered by Machine Learning, with integrated device profiling and a flexible rule engine. It helps clients increase revenue by lowering false positives. By combining our custom ML model with a high-performance Rules Hub, we can classify transactions with greater accuracy than ever before.

CONTACT YOUR LOCAL OFFICE TO FIND OUT MORE

Glossary: Machine Learning fraud terminology

Let's take a look at the terminology associated with cyber fraud and how these apply to ML fraud prevention.

What are Algorithms?

These are a clearly defined sequence of procedures used to solve a specific problem. In fraud prevention, there are a variety of different algorithms used to identify fraudulent behaviour such as decision trees, gradient boosting and random forests. These are used to analyse customer data to identify fraud.

What are APIs?

An Application Programming Interface is a coding tool used between two programs to enable cloud connections and interactions. REST APIs, or representational state transfer APIs, are preferred as they use less bandwidth.

What is a false positive?

When a bonafide customer is wrongly classified as a suspected fraudster.

What is a false positive rate?

This metric allows you to assess the accuracy of your ML model to classify customers and is critical in the evaluation of your fraud prevention process.

What are ML Features?

Every online transaction has a set of individual attributes associated with it, such as monetary value, location, and frequency. Combining one or more attributes gives you a feature, which can be used to train the algorithm. By identifying those features with the most predictive power you can build effective ML models. Feature engineering is a speciality that requires a precise combination of attributes and makes feature-based ML vastly superior to simple yes-no rules-based systems.

What is Machine Learning?

This generic term can be broadly understood as deriving knowledge from experience. ML is a type of AI computer system that learns from examples and identifies patterns with the help of algorithms. By providing enough high-quality data as examples the ML model can deliver an unrivalled 99% level of accuracy. The beauty of ML in fraud prevention is that it delivers a probability score rather than a yes-no decision, this likelihood of fraud score allows businesses to precisely set their risk appetite level depending on the nature of their offering.

What is a set of rules?

Rules are clearly defined descriptions of the criteria and parameters that give you a binary yes or no answer. They are often described as if-then logic and the combination of all the rules that lead to a decision is known as a set of rules. Experian has a standard set of over one hundred rules that it uses to create individual sets depending on the customer.

What is a rule hub?

This is the software that uses decision logic from your set of rules to automatically control your fraud process. It allows you to change rules as needed to constantly adapt to new fraud threats.

What is re-training an ML model?

Once an ML fraud system is set up users can feed new data into the model at regular intervals. This allows you to constantly train your model on the latest techniques that fraudsters are using to try and outsmart fraud prevention systems. This ongoing training ensures that your model is always up-to-date and provides ever better accuracy.

What is a risk score?

This number from 0 to 100 indicates the probability of a potential customer being a fraudster. The score is determined by the algorithms analysing the features. The accuracy of a risk score depends on the quality of the data used to train the algorithms. Understanding the subtleties of your training set data is the key to producing an accurate risk score.

What is a threshold?

A decision threshold is the trigger point for when an action follows, such as the use of a rule. The threshold value or limit is the score that you decide is the tipping point to either accept, decline or review a transaction.

Appendix: Different types of online fraud

There are several types of fraud that are relevant to this guide, let's look at each one to gain a better understanding of how fraudsters operate.

Identity theft

This is one of the most prevalent types of fraud. It involves a fraudster impersonating a real customer by assuming the identity of that person. There are various techniques that fraudsters use to steal an identity such as:

- **Phishing –** an unsuspecting individual willingly provides personal data via an email or website that imitates a legitimate business
- **Pharming –** a virus or trojan from an app or website is used to steal personal data
- **Whaling –** a fraudster impersonates a business leader to target executives and steal company data

Once a fraudster has stolen an identity, they can use it to conduct other types of fraud such as clean fraud, account takeover and loyalty fraud.

Clean fraud

These are fraudulent transactions that appear to be valid as the fraudster has access to the victim's unique payment data. There is nothing suspicious about these transactions apart from the fact that the real owner of the data is not making the order, so they are usually able to bypass the merchant's or service provider's security controls.

Account takeover

This is a form of identity theft in that the fraudster gains unauthorised access to one or more of the victim's online accounts and uses them to make fraudulent transactions. Account takeovers can happen with any online account ranging from banking and other payment methods to email and social media.

Loyalty fraud

Many businesses offer loyalty reward programs and once a fraudster has taken over an account with stolen identity data, they can exploit these benefits for their own gain.

Card testing

To verify if the stolen credit card or payment data is still available fraudsters make a small transaction to reveal which cards have already been cancelled before moving on to make larger transactions. The smaller the transaction amount the less chance there is of it being noticed by the owner of the card.

Affiliate fraud

This type of fraud involves fake activity – often conducted by bots – that generates fraudulent commissions for affiliate marketing. There are a variety of different schemes depending on the affiliate payment model, these include fake leads or impressions and automated clicks.

Re-Shipping

The first stage of this scam is for a fraudster to steal credit card information and order goods online. Instead of having the goods delivered to their own address, they use a middleman to receive and repackage the goods, before sending them to the fraudster.

These intermediaries are often unaware of the crime and are recruited with the promise of a legitimate work-from-home opportunity.

Botnets

These are networks of computers infected with malware that are controlled by fraudsters. By using multiple computers and stolen payment and identity data they can often deceive security controls to make it look as though a transaction originated from the same location as the stolen credit card.

Triangulation fraud

This involves a fraudster advertising goods at low prices on an online auction website. As soon as they have an order from an unsuspecting buyer, they purchase the goods from an online merchant by using stolen identity and credit card data and send them directly to the buyer. As soon as the owner of the credit card reports the fraudulent transaction the original merchant will have to refund the purchase, with a loss of goods at the same time.

Visit our website for more information on how we can help:
www.experianacademy.com

Or contact us:
www.experianplc.com/contact-us/region/#emea

The insight contained within this report is prepared using research performed on both Experian data and external data sources, in addition to market research. All sources, unless referenced, are from Experian insight.