AT A GLANCE

Company

Whitepages Pro provides businesses with global identity verification solutions via enterprise-scale APIs and web tools to help companies identify legitimate customers, prevent fraudulent transactions, and smooth new customer account creation. We developed our own fully-integrated, high-availability Identity Graph database that houses more than 5 billion global identity records. This real-time data integrates into existing platforms, authentication workflows, and data models to help businesses confidently assess and verify consumer identities worldwide.

Problem

Our consumer product, **Whitepages Premium,** which provides access to U.S. public records to verify identities, was experiencing a high number of chargebacks due to new account fraud. Chargebacks needed to be at a manageable level. Otherwise, the company risked exceeding the 1% fraud rate which could result in the possibility of increased fines, or worst case, getting shut down by Visa or Mastercard.

Solution

Build a machine learning model for Premium that scores transactions and identifies fraud. Leverage Pro's Identity Check Confidence Score to improve the effectiveness of the model.

Results

- Model significantly reduced fraud and the number of chargebacks.

- Estimated that adding Identity Check Confidence Score to the model would save 12% per order.

Reducing Fraud with a Machine Learning Model and Whitepages Pro Identity Check

All businesses are at risk for chargebacks due to fraud, especially online businesses. The FTC **reported** that 3.1 million complaints were collected by the Consumer Sentinel Network in 2016. In fact, 13% of those complaints were related to identity theft, and credit card fraud complaints rose to more than 32%. Card-not-present (CNP) fraud in the U.S. increased 40% in 2016 according to a Javelin Strategy & Research **study**, and the rising adoption of EMV cards and terminals are driving many fraudsters to open new, fraudulent accounts.

E-commerce companies know that it's crucial to keep their fraud rate below 1% otherwise they risk losing the ability to accept credit cards, not to mention the lost revenue from having to pay for numerous chargebacks. At Whitepages, we found ourselves being subjected to an increasing amount of fraud and chargebacks after launching our consumer product, **Whitepages Premium**, which provides subscribers access to US public records to verify contact details, mobile numbers, bankruptcy history, criminal

---

Whitepages Pro provides businesses with global identity verification solutions via enterprise-scale APIs and web tools.

pro.whitepages.com | +1.877.767.8052

records, and more to facilitate trusting interactions in today's sharing economy.

Our team created a machine learning model to help solve our fraud problem, particularly new account fraud. With the introduction of Identity Check Confidence Score, a feature of the Whitepages Pro Identity Check product, it further improved the results of the model.

THE PROBLEM

Fraudsters were opening new Whitepages Consumer Premium accounts with stolen identities in order to steal our Premium data. Eventually, the company would be alerted by consumers whose identities had been stolen, and we had to pay those consumers back for the fraudulently purchased Premium subscription product. Because of new account fraud, we were experiencing a high number of chargebacks and needed to return the number of chargebacks to a manageable level. Otherwise, we ran the risk of exceeding the 1% fraud rate which could result in possibility of increased fines, or worst case, getting shut down by Visa or Mastercard.

We needed a machine learning solution that could detect new fraudulent accounts and could also help with alleviating credit card fraud.

THE SOLUTION

The team created a logistic regression model for Whitepages Premium that takes in a number of inputs such as payment behavior, information about the customer, and user behavior on the website, and then combines that data into a score. A higher score is usually indicative of fraud, and a lower score is typically indicative of a legitimate customer. While the model did help mitigate the fraudulent use of our Premium product and reduce the number of chargebacks, there was still room for improvement.

We decided to leverage Whitepages Pro's **Confidence Score**, which at the time was a newly built feature of the

Identity Check product. Identity Check helps businesses fight fraud and identify good customers by leveraging real-time global data, machine learning, and network insights across the five core consumer data attributes of email, phone, person, address, and IP. For example, Identity Check could verify that the email matches the name, or tell you information about the email – e.g. email first seen date, disposable status, etc. Identity Check returns 70+ data signals, leverages proprietary Identity Network insights, and provides a Confidence Score in less than a second.

The Confidence Score provides a comprehensive assessment of each transaction by leveraging the millions of patterns across our Identity Network and Identity Check's 70+ data signals. This assessment is delivered as a single score on a 0-500 range. The Confidence Score can be accessed in three ways: from a fraud platform that has Identity Check directly integrated (**see platform integrations**), as an attribute in the **Identity Check API**, and in the Identity Check Web interface.

The Confidence Score is powered by a machine learning model that is trained based on outcome data - historical transactions where our customers tell us if transactions are good or bad. **» Outcome data from our customers helps us continuously improve the accuracy of our Confidence Score.**

Looking to evaluate the new Confidence Score quickly, our consumer product was determined to be the perfect test bed for developing a new version of our logistic regression model, and a great way to understand and improve our Pro offering. Using the Identity Check API to continually feed information from our Pro product into the logistic regression model we built for Premium, including the new Confidence Score; the fraud model now provides a clearer picture of our customers.

The model is just one of multiple pieces used for fraud detection- we also have a manual review team because human intervention is important to confirm that transactions flagged as fraud are truly fraud. However, the model is a critical piece of our whole fraud-fighting arsenal.

Whitepages Pro provides businesses with global identity verification solutions via enterprise-scale APIs and web tools.

pro.whitepages.com | +1.877.767.8052

AUTO-REJECT AND AUTO-ACCEPT RULES COULD PLAY A KEY ROLE

We did a great deal of analysis on auto-reject and auto-accept rules in relation to our Premium machine learning model. This discovered that there is a mathematically rigorous answer to where, in your model, you should put auto-reject rules.

Auto-reject is a costly process because on a per order basis, a chargeback may cost a few cents plus the cost of the product you are shipping - and you have to eat that cost. However, if in the process of auto-rejecting certain things, you also auto-reject a bunch of good orders, that could be expensive because you are potentially forfeiting the lifetime revenues from that customer. In most cases you must look at the cost of a chargeback, or effectively the cost of processing a bad order vs. the cost of rejecting a good order and you must look at those two things in tandem.

After weighing the pros and cons of implementing auto-reject rules, some merchants may find that they are better off not putting in auto-reject rules and instead introducing friction. For example, in our sign-up process for creating a new account for our Premium product, it only asks for email. But perhaps if the model said, "hey this looks fraudulent," we would also ask for a cell phone number to text the customer and make sure the new account is legitimate. Instead of rejecting that customer outright, it was found that we were better off putting more friction into the process, and that friction could be in the form of a manual review or 2-factor auth. You could start to introduce more layers of friction instead of putting in an auto-reject rule. Auto-reject rules could be expensive if not managed correctly.

In our analysis of our machine learning model, we predicted the **» savings opportunity would be 2-3x higher for cost optimization of auto-reject rules** vs. auto accept rules.

THE RESULTS

We are extremely pleased with the results of our Premium fraud model and the improved performance of the model after adding Pro Identity Check Confidence Score. Our original logistic regression model significantly reduced fraud and the number of chargebacks. However, when Pro Identity Check Confidence Score was added to the Premium fraud model, **» we estimated that it saves an additional 12% per order.**

AN EFFECTIVE MEANS OF FIGHTING FRAUD

Before our model was in place we were dealing with a high level of chargebacks due to fraud and were at risk of exceeding the 1% fraud threshold. Once we improved the accuracy of our model by adding Pro Identity Check Confidence Score; it improved our ability to fight fraud and significantly reduced the number of chargebacks to a manageable and acceptable level.

Dealing with fraud and chargebacks is a necessity for every growing business. We found that by using a machine learning model that takes in many diverse inputs, coupled with a manual review process, is an effective means of fighting fraud, and ultimately reducing the risk of chargebacks.

Whitepages Pro provides businesses with global identity verification solutions via enterprise-scale APIs and web tools.

pro.whitepages.com | +1.877.767.8052