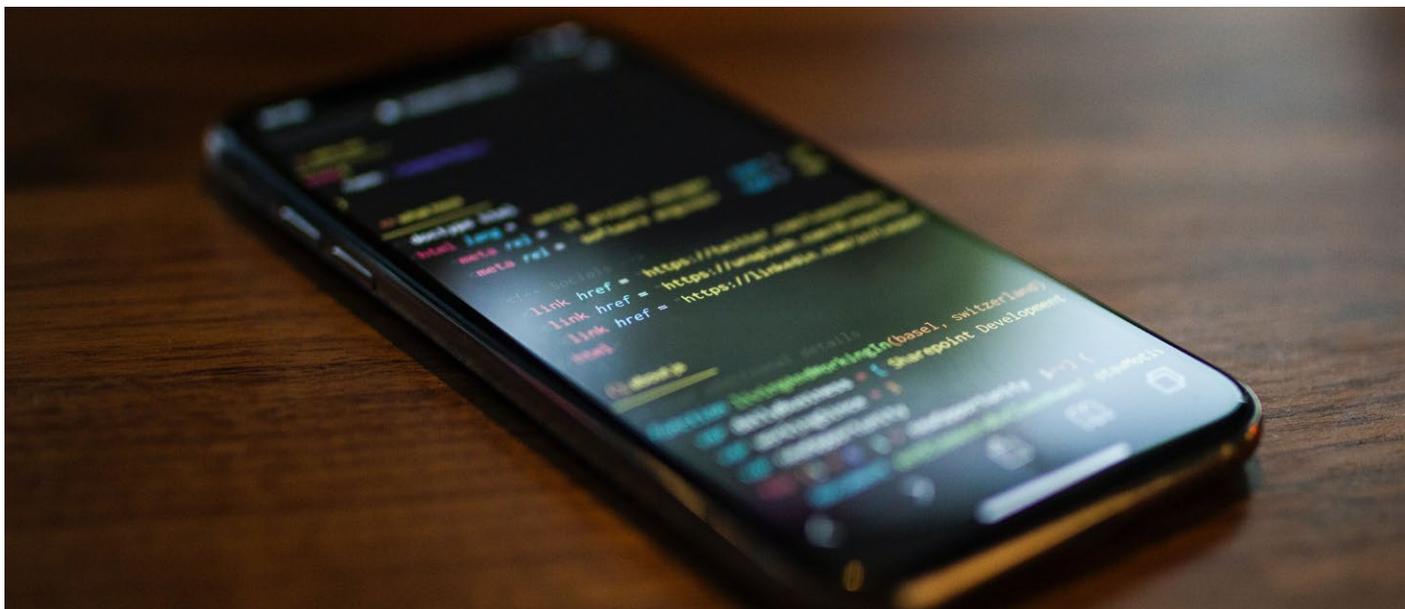# pasabi

THE SCIENCE OF FAKE REVIEW DETECTION

By Chris Downie

Whether you are an online retailer, a travel website or a comparison service, trust is at the heart of your business. Authenticity is key and if consumers can't trust the reviews on your website, how can they trust the products and services you are selling?

The challenge isn't at the human level, it is at a data scale

When you know what you are looking for, the signs of fake reviews are easy to spot. As humans we are pretty good at identifying the tell-tale signs of fabrication. From badly worded reviews, to common spelling mistakes and content that is clearly biased or even written about a totally different product, the signs are there for most to see. However, for a business to tackle this problem it is not a question of individual quality control, it is a question of scale.

Websites and apps are 24/7 and so are your customers. They are generating vast quantities of content and you can't possibly keep tabs on all of this input. Thankfully technology that was previously the domain of industry giants, such as Tripadvisor and Trustpilot, is now available to businesses of all sizes to take on the fraudsters. From the latest AI technology to tools used by banks to detect patterns of fraud, your business can now benefit from fake review detection at scale.

"For a business to tackle this problem it is not a question of individual quality control, it is a question of scale."

Chris Downie, CEO Pasabi

pasabi

What is fake content?

Before diving into the science behind the detection of fake reviews, let's look a little at the classifications of fake or fabricated content and what challenges they present to a business.

Spam & Scam

You create a top quality website to engage your audience and the first thing you may find lurking in the corners are the spammers and scammers. From hacking services to insurance scams to pyramid schemes, unscrupulous individuals and groups are riding on your success to peddle their wares to an unsuspecting audience. Whilst consumers may not fall for these schemes, the very presence of spammers and scammers on your website hurts your brand and challenges your audience's trust in your service.

Biased positive or negative reviews

It's a battleground out there for competing businesses and some of the weapons they deploy can be seen in the review sections of your websites.

Biased positive reviews come from companies who incentivise their employees, friends, family and customers to boost their business or product profile. This can be very direct, where they simply ask for a positive review in return for discounts and other benefits, or it can be more subtle such as "review gating". This is where companies engaging in customer service phone calls or email exchanges, send the good responders to your site to post an inevitably positive review to boost their profile, whilst directing bad responders into a black hole. Either way, they are creating a false impression which your website or app is unwittingly promoting.

Biased negative reviews come from one of two sources: a rival company who is looking to negatively impact the profile of another; or a malicious consumer, who is looking to blackmail a company for their own benefit. As a website or app promoting the services and products of other companies, you will want to help protect your customers from such nefarious behaviour.

Paid for reviews

Finally there are "services" out there that are targeting businesses like yours, specialising in massaging the profiles of their customers so that their businesses look like they genuinely perform well on your website when in fact it is being manipulated by a third party. From reporting genuine negative reviews as false, to automating the writing of a mixture of positive and neutral reviews that will result in a steady increase in profile, these service providers are paid to "game the system"; your system.
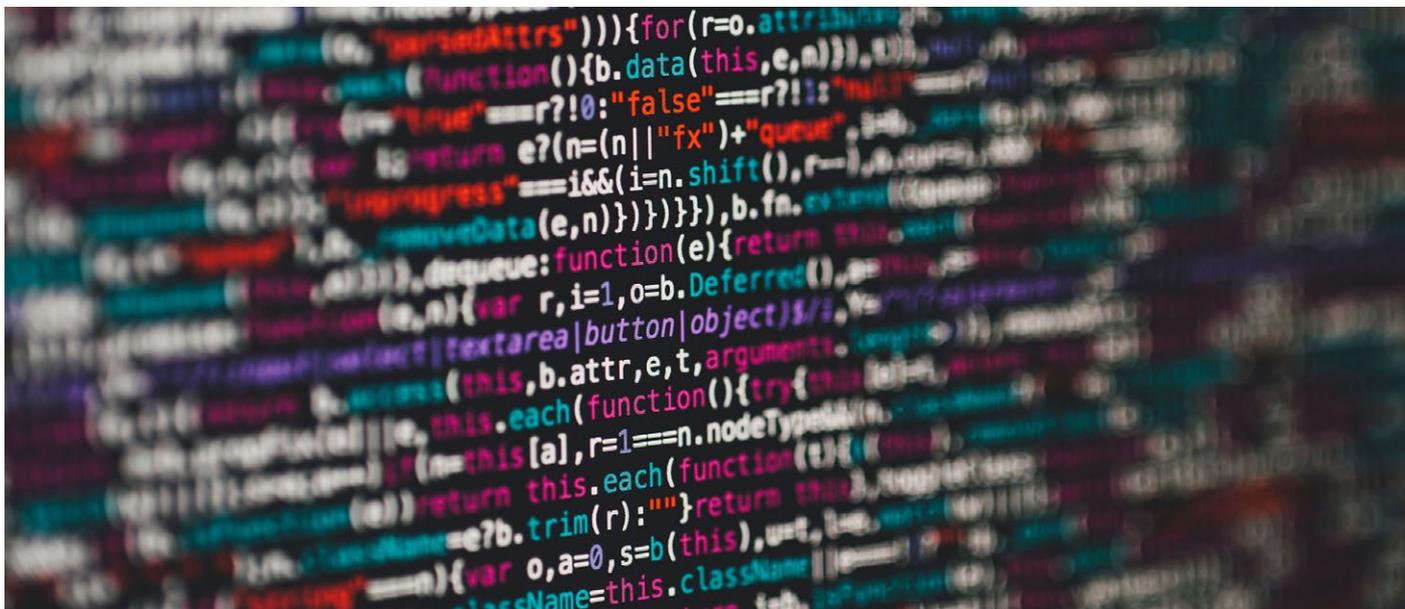
How can technology help?

Thankfully, everything that happens within your user-generated data leaves traces that technology can follow.

At Pasabi, we have worked with some of the world's biggest brands, from luxury giant LVMH to consumer review industry leaders like Trustpilot, to develop a state-of-the-art fraud and fake content detection system that reads these signals and shines a spotlight on the fraudsters who are trying to game your system.

With the culprits identified and constantly monitored, the system can use automation to clean up your user generated content and, if necessary, gather evidence to allow you to pursue action against high-value targets.
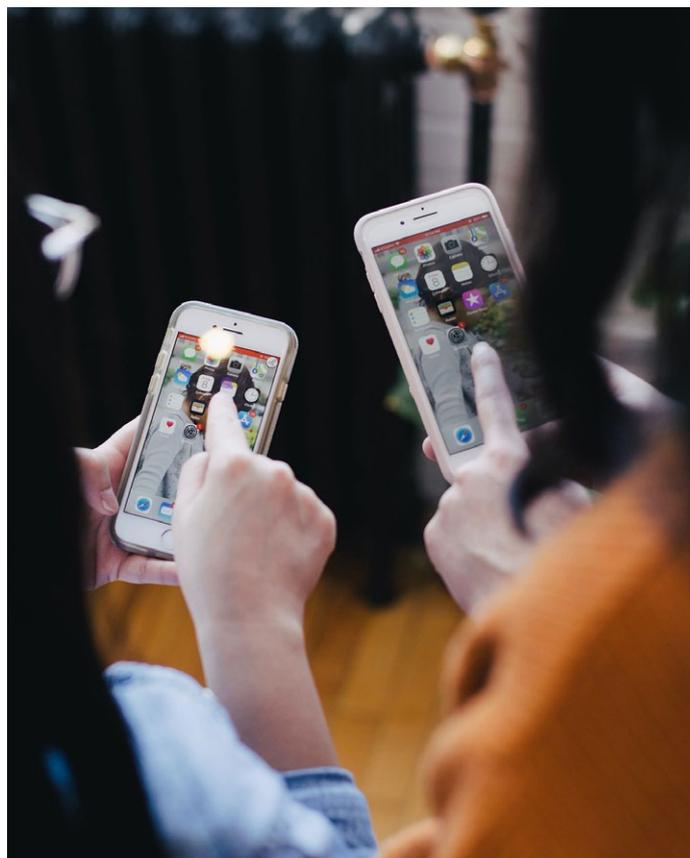
How does it work?

The Pasabi system works like a high-tech digital production line. We automatically take the review content that is posted to your site or app and run it through a series
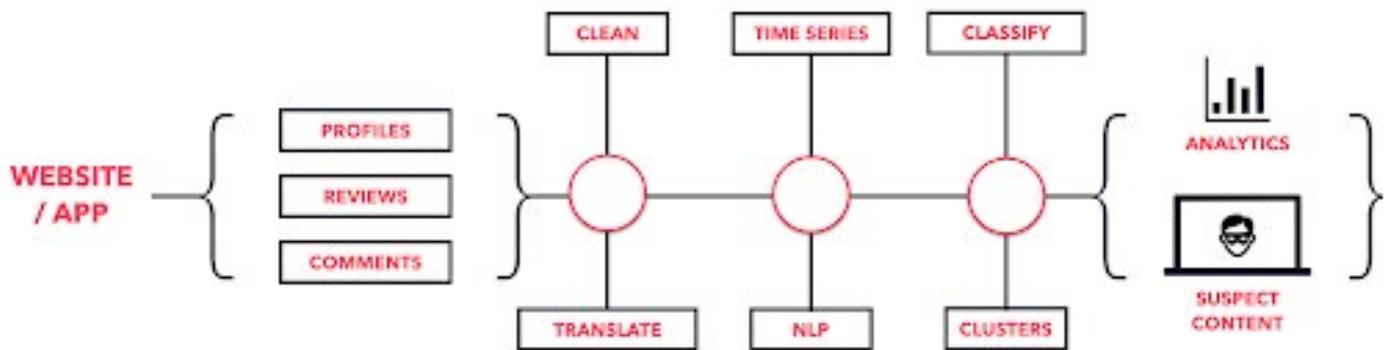
Here are some of the tools and techniques we use:

Natural Language Processing - This is a classic artificial intelligence technique that we have built from scratch and honed to break down the language used within reviews, highlighting key phrases that help us identify nefarious activity.

Machine learning classifiers - trained using hundreds of thousands of examples, our classifiers spot the language style of fabricated, spam and scam reviews.
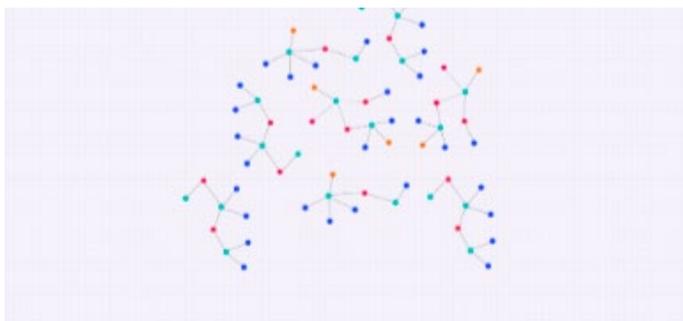
Time series analysis - fraudsters tend to post in patterns that we can spot over time. By tracking the behaviour of both reviewers and the companies they review over time, we can see spikes in the data that are the red-alerts for suspicious activity. If it looks too good to be true (or too bad to be true!),

then it probably is.

Cluster identification - the worst culprits tend to work together. Whether it is groups of individuals who are working as one organisation, or a series of "bots" (automated systems that use fake accounts) being deployed alongside real users, Pasabi uses the latest cluster detection technology and bespoke algorithms to spot these groups to help you either block their activity or provide you with the evidence to take action against them. By combining all the individual data points we have gathered, we look for patterns within that data which shines a light on the groups who are taking advantage of your service.

What is the end result?

The output of all this work is verified, quality reviews. And the benefit to you? Trust. Trust from consumers who rely on your service to make informed decisions and also from the businesses who pay to benefit from being on your platform.

How can this work for me?

Whether you want to clean up your content, protect your customers, or are looking to take direct action against groups who are abusing your service, the Pasabi platform can support your needs.

ⓞ pasabi

Request an Online Review Insights Report

Understand fabricated reviews, scam & spam, and misbehaving companies on your platform and how to take effective action at scale.

Request Report

About the Author

Chris Downie

Chris Downie is Pasabi CEO. A technical leader with over 20years' experience he has delivered major digital projects for Schroders, the Rank Group and Skyscanner where he achieved 20 Million app downloads. Leading Pasabi, his vision is for an AI technology platform that supports a range of use cases in delivering authentic online user experiences.

Pasabi

Pasabi supports the world's largest brands, deploying unique AI technology solutions that detect and prevent the proliferation of counterfeit goods and fake reviews online.

We help brands establish the true scale of their counterfeit goods problem online, identify offenders and groups acting together, supporting takedowns and offline legal action.

Pasabi defends your brand's authenticity online.

pasabi