Machine Learning Classifier: Cut Web-Fuzzing false positives by 50%

50% drop in Website Scanner noise

20% fewer irrelevant URL Fuzzer findings

Weighted F1-score rises from ~75% to ~92%

Instant value: the ML classifier runs automatically inside every scan

The hidden cost of rule-based vulnerability scanners

Traditional rule-based scanners promise coverage and speed. They do the job, but with unnecessary noise. They flag too much, miss too often, and generally slow security teams down.

Their detection logic relies on brittle, RegEx-heavy rules. As a result, they generate lengthy vulnerability reports, full of false positives, that force analysts to spend hours reviewing benign results and tuning filters.

That lost time could go towards analysis, remediation, or delivery.

Accuracy shouldn’t be a bonus.
It should be the baseline.

Moreover, these reports often result in missed threats:

- Static logic can’t adapt to modern application complexity, meaning multilingual error messages, dynamic content, or soft 404s slip through undetected.
- Sensitive content, like login portals, exposed backups, or API keys, is often buried or ignored.

As a result, the scanners create bloated reports, delay remediation, and ultimately diminish analyst trust in their findings. Accuracy shouldn’t be a bonus; it should be the baseline. Tools that miss the mark create downstream pain at every stage.

Engineered ML, not buzzword AI

To fix this problem, we created a purpose-built solution, not just bolted AI onto an existing one. Our web application security engineers engineered the ML Classifier behind Pentest-Tools.com to solve a specific challenge: reducing false positives in fuzzing workflows while maintaining high coverage.

We fine-tuned LLaMA models versions 3.1 and 3.2 using 3B and 8B parameter configurations, balancing intelligence with performance. The models run locally for fast, efficient inference during active scanning sessions, meaning there are no cloud overhead or latency tradeoffs.

More accuracy, less noise: what the ML classifier does

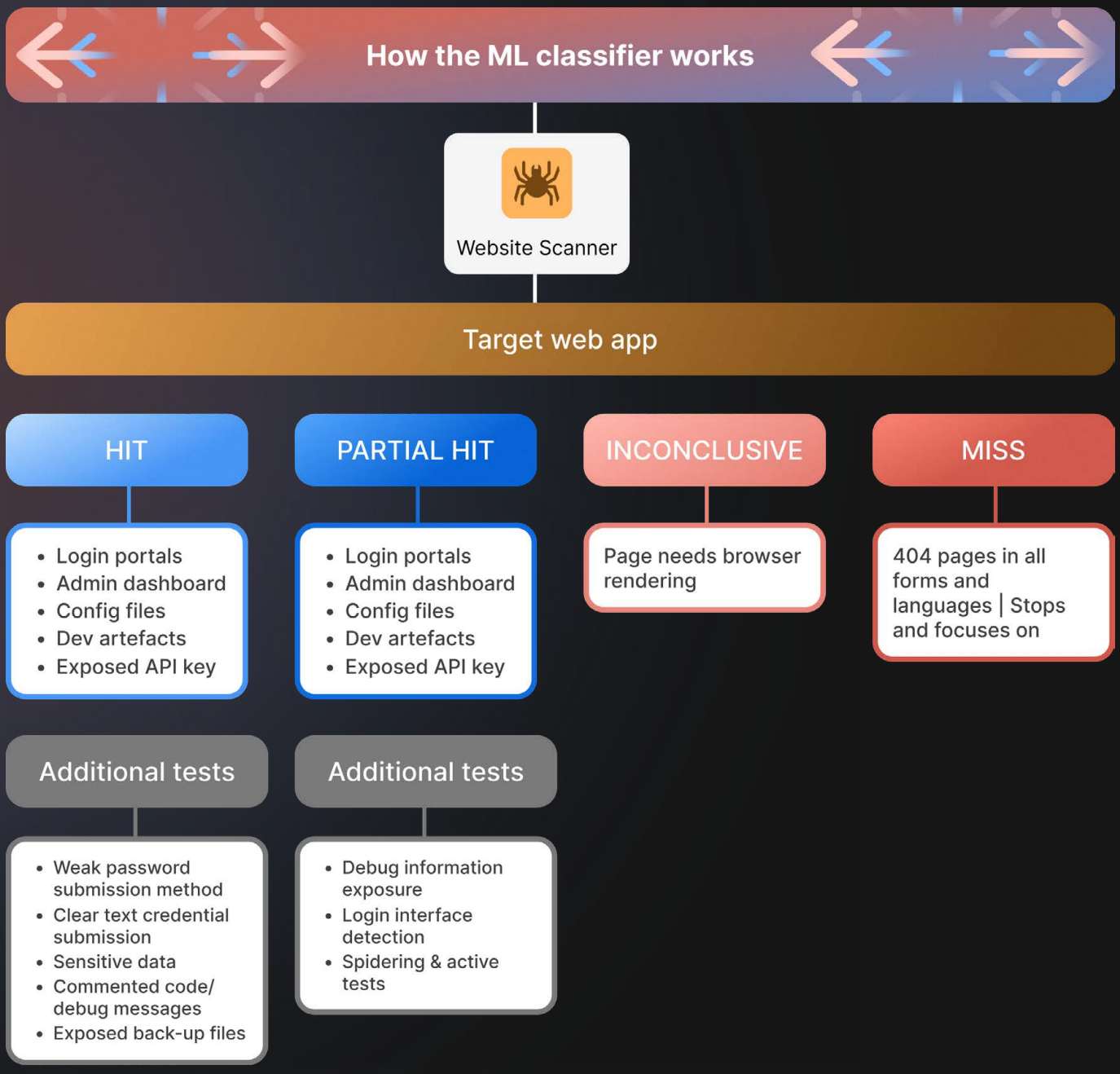
The ML classifier processes every HTML response during a scan and assigns it to one of four clear categories:

- **HIT**: High-value targets such as login pages, exposed configuration files, backups, admin portals, code fragments, and secrets.
- **MISS**: Confirmed dead-ends such as 404s and equivalent “not found” responses, even when status codes are misleading.
- **PARTIAL HIT**: Ambiguous but potentially interesting results such as firewall responses, generic templates, or redirect chains.
- **INCONCLUSIVE**: Pages, just as JavaScript-heavy content, that require browser rendering to classify accurately.

This structured triage cuts down on false positives without introducing false confidence.

It also solves a longstanding problem: multilingual recognition of error patterns. The classifier can spot “not-found” pages in dozens of languages - Spanish, French, German, Chinese, Farsi, and more - without needing a separate rule set.

Static logic can’t do that. The ML Classifier can.



How the ML Classifier does it

1

Extracts & normalizes tags

We pull out the essential HTML tags and simplify them into a normalized structure. This gives the model clear, standardized input that's easy to interpret and compare.

2

Removes unnecessary noise

Then, we intelligently filter out irrelevant elements that would confuse the model. We strip noise like third-party scripts, repeated boilerplate, or advertising code before classification, preventing the model from overfitting on meaningless elements.

3

Handles the curveballs

Our preprocessing carefully handles edge cases and unexpected structures - like malformed HTML, incomplete responses, or deliberately obfuscated content - ensuring a smooth and consistent input for the classifier, no matter how quirky the original HTML.

4

Keeps things private

We anonymize any references to specific domains or resources during this process, removing and masking URLs, hostnames, or brand-specific strings before training and inference. This ensures that known data doesn't bias classification.

5

Ensures a fair learning environment

To minimize bias and ensure our model learns generalizable patterns, we meticulously curated our training data, using clever heuristics to remove duplicates and create a diverse range of examples.

We wanted our model to learn from a balanced and representative dataset, so training included thousands of real-world examples across stacks, frameworks, and languages. As a result, the model generalizes well and adapts to the web's unpredictable nature.

How to use the ML Classifier: click 'Scan'

Using the ML Classifier is easy. It's baked in, so all you need to do is click scan. It works out of the box, and no configuration is needed for either the Website Vulnerability Scanner or URL Fuzzer.

What to expect

Improved risk posture with cleaner findings

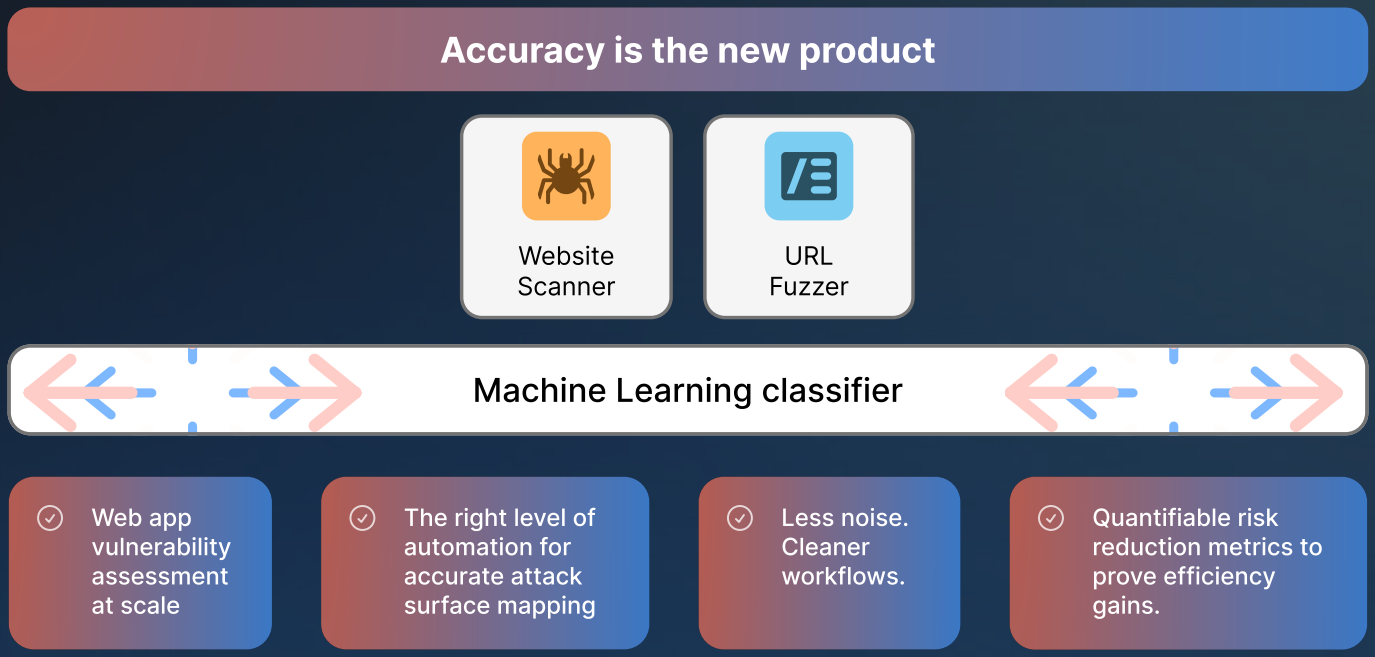
Accurate scan outputs yield more actionable reports, enabling teams to fix issues faster. Classifiers that distinguish threats from noise contribute to shorter, clearer, and more defensible reports. With a weighted F1-score improvement from ~75 % to ~92 %, the increase in precision isn't just noticeable; it's measurable.

Accelerated remediation by reducing noise

Halving noise results accelerates remediation cycles. By automating triage at the HTML response level, the classifier reduces fuzzing false positives by 50% and irrelevant findings in the URL Fuzzer by 20%.

Increased engagement capacity through automated triage

Reclaim time for client-ready deliverables instead of manual triage. Reports contain vetted, contextual, and relevant findings, eliminating wasted time investigating non-issues and allowing teams to concentrate on reducing real risk.



Why Pentest-Tools.com?

Pentest-Tools.com is built for actual security testing, not just detection.

We provide the **coverage, consolidation, and automation** security teams need to optimize vulnerability assessment workflows. And we ensure the **depth, control, and customization** on which professional pentesters count to increase engagement quality and profitability.

- **6,3 million** Vulnerability scans per year
- **1,6 million** Scheduled, cloud-based scans per year
- **611,000+** Automated, multi-tool scans with Pentest Robots
- **15,000+** Detected vulnerabilities & custom exploits

Pentest Tools

Discover what's possible. Prove what's **real.**
With proprietary tech and key **offensive security** experts.

Europe, Romania, Bucharest
48 Bvd. Iancu de Hunedoara

support@pentest-tools.com
pentest-tools.com

Join our community of ethical hackers!