

# **Sistema de Hiperautomação**

## **Detecção de Pirataria em Cartuchos HP**

Integração RPA + Inteligência Artificial

**Relatório Técnico**

Gerado em: 30/10/2025 21:44:25

Sprint 4 - Encerramento e Integração Inteligente

# 1. Introdução

Este relatório técnico descreve a implementação do sistema de hiperautomação desenvolvido para detecção automática de indícios de pirataria em produtos de cartuchos HP disponíveis na plataforma Amazon.com.br. O sistema combina técnicas de Robotic Process Automation (RPA) para coleta automatizada de dados com algoritmos de Inteligência Artificial leve para classificação e análise de risco dos produtos coletados. A solução foi desenvolvida seguindo as melhores práticas de automação corporativa, utilizando tecnologias open-source e abordando o problema de forma incremental através de sprints de desenvolvimento.

## 1.1. Objetivos

O sistema foi projetado para:

- Executar coleta automatizada de dados de produtos da Amazon.com.br
- Classificar produtos utilizando modelos de IA treinados
- Calcular níveis de risco baseados em regras de negócio e previsões da IA
- Gerar relatórios e alertas automáticos para produtos de alto risco
- Fornecer dashboard interativo para visualização dos resultados

## 2. Arquitetura do Sistema

O sistema é composto por três componentes principais que trabalham em conjunto: **2.1. Robô RPA (Amazon Scraper)** Responsável pela coleta automatizada de dados da Amazon.com.br. Utiliza Selenium WebDriver para simular navegação de navegador e extrair informações dos produtos, incluindo título, preço, vendedor, avaliações e URL. **2.2. Classificador de IA** Implementa algoritmos de Machine Learning (Random Forest + TF-IDF) para classificar produtos em três categorias: ORIGINAL, SUSPEITO ou COMPATIVEL. O modelo é treinado com base em regras heurísticas aplicadas aos dados históricos. **2.3. Analisador de Risco** Combina previsões da IA com regras de negócio para calcular um score de risco e classificar produtos em níveis ALTO, MÉDIO ou BAIXO risco.

## 3. Fluxo RPA + Inteligência Artificial

### 3.1. Pipeline Integrado

O pipeline completo de detecção segue os seguintes passos: Etapa 1: Coleta de Dados (RPA) O robô RPA navega pela Amazon.com.br buscando produtos relacionados a cartuchos HP. Para cada termo de busca configurado (ex: "cartucho HP 667"), o sistema:

- Navega pelas páginas de resultados
- Extrai informações básicas de cada produto (listagem)
- Acessa páginas individuais para obter detalhes completos
- Filtra produtos sem informações essenciais (ex: vendedor)

Etapa 2: Comparação com Base Consolidada Os produtos coletados são comparados com uma base de dados existente para identificar produtos já conhecidos e novos produtos que requerem análise. Etapa 3: Classificação com IA Cada produto coletado é analisado pelo modelo de IA que:

- Extrai features textuais (TF-IDF do título, descrição, vendedor)
- Extrai features numéricas (preço, confiança do vendedor)
- Gera predição de classificação (ORIGINAL/SUSPEITO/COMPATIVEL)
- Calcula score de confiança da predição

Etapa 4: Análise de Risco O analisador de risco combina múltiplos fatores:

- Predição da IA
- Confiança da predição
- Preço do produto (valores muito baixos ou altos são suspeitos)
- Vendedor (marketplace vs vendedor oficial)
- Presença de palavras-chave suspeitas

O resultado é um score de risco e classificação em ALTO, MÉDIO ou BAIXO risco. Etapa 5: Geração de Relatórios Os resultados são consolidados e apresentados em múltiplos formatos:

- CSV para análise em planilhas
- HTML para visualização web
- PDF para documentação formal
- Dashboard Streamlit para análise interativa

Etapa 6: Alertas Automáticos Produtos classificados como ALTO RISCO geram alertas automáticos que são:

- Registrados em arquivo de log
- Simulados para envio por email (mockado em ambiente de desenvolvimento)
- Salvos em arquivo de alertas para revisão manual

## 4. Exemplos de Detecção

### 4.1. Produto Classificado como SUSPEITO

**Título:** KIT 3x Refil Tinta Preta Compatível com 664 T664 100ml cada | 300ml no Total | L395 L380 L375 L220 L455 | Premium | KORA ou EVOLUT  
**Preço:** R\$ 24.0  
**Vendedor:** Kitts Mania BR  
**Predição IA:** SUSPEITO  
**Confiança:** 0.35 (se disponível)  
**Nível de Risco:** ALTO  
**Score de Risco:** 6

### 4.2. Produto de Alto Risco Detectado

**Título:** Tinta 100ml Black Compatível Impressora Cartucho 667xl  
**Preço:** R\$ 17.0  
**Vendedor:** Amazon.com.br  
**Razões do Alto Risco:** • Predição da IA: COMPATIVEL • Score de risco: 4

### 4.3. Estatísticas de Detecção

Métrica	Quantidade	Percentual
Total de Produtos Analisados	180	100%
Produtos Suspeitos	1	0.6%
Produtos de Alto Risco	5	2.8%

## 5. Regras de Negócio

O sistema utiliza regras heurísticas combinadas com IA para detectar produtos suspeitos:

5.1. Palavras-chave Suspeitas Produtos contendo as seguintes palavras-chave são sinalizados: • genérico, cópia, compatível, recondicionado, usado • refurbished, remanufactured, compatible, generic • não original, alternativo, substituto, imitação 5.2. Análise de Preço Preços muito abaixo do mercado (< R\$ 30) ou muito acima (> R\$ 200) são considerados suspeitos e aumentam o score de risco. 5.3. Confiança do Vendedor Vendedores oficiais (Amazon.com.br, HP Brasil) reduzem o score de risco. Vendedores de marketplace ou terceiros aumentam o score de risco. 5.4. Score de Risco O score é calculado considerando: • Predição da IA (SUSPEITO = +3, COMPATIVEL = +1) • Confiança da IA (< 0.7 = +1) • Preço (< R\$ 30 = +2, > R\$ 200 = +1) • Vendedor (marketplace = +1) Classificação final: • Score  $\geq$  4: ALTO RISCO • Score  $\geq$  2: MÉDIO RISCO • Score < 2: BAIXO RISCO

## 6. Modelo de Inteligência Artificial

6.1. Algoritmo Utilizado O sistema utiliza Random Forest combinado com vetorização TF-IDF para classificação:

- Features Textuais: TF-IDF aplicado ao título, descrição e vendedor
- Features Numéricas: Preço, tamanho do título, confiança do vendedor, contagem de palavras suspeitas/originais

6.2. Treinamento O modelo é treinado usando dados históricos rotulados por regras heurísticas. A acurácia típica é de aproximadamente 85% em dados de teste.

6.3. Classes de Predição

- ORIGINAL: Produto identificado como original/genuíno
- SUSPEITO: Produto com indícios de pirataria
- COMPATIVEL: Produto compatível (não original mas não necessariamente pirata)

6.4. Integração com Regras As previsões da IA são combinadas com regras de negócio para calcular o risco final, permitindo que o sistema aprenda com padrões complexos enquanto mantém controle através de regras explícitas.