iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XI Month of publication: November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75016

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Examining the Use of Analytics in Helping Detect Counterfeit Goods Sent to the United States of America

Muleka Christelle Masudi
University of Louisville, United States

Abstract: The flow of fake (counterfeit) goods into the United States is a serious problem. These fake products like designer clothes, electronics, medicine, and car parts not only break the law but also put people's safety at risk and cost the U.S. economy billions of dollars every year (Ocean Tomo, 2024; U.S. Customs and Border Protection [CBP], 2023). As online shopping and global trade grow, old-fashioned inspection methods are no longer enough to catch the large number of fake goods being shipped in (ProCogia, 2025; Garcia-Cotte et al., 2024).

This study explores how data analytics can help detect these fake products. We used a dataset of 5,000 shipments that were flagged by U.S. Customs between 2022 and 2024. Using that data, we built computer models logistic regression and random forest classifiersto predict which shipments were likely to be fake. The models used many types of information, such as the country the shipment came from, its weight and value, how often the vendor ships items, and even image analysis using computer vision tools (Pulfer et al., 2022; Garcia-Cotte et al., 2024).

Our results showed that using these advanced tools made a big difference. The accuracy of detecting fake goods increased from 68% (with old methods) to 91% with our analytics-based approach. We also found a strong connection ($r = .82$, $p < .001$) between the risk scores predicted by our model and the actual fake shipments that were seized.

These findings show that analytics can greatly improve how customs officers do their jobs. By helping them focus on the most suspicious shipments, analytics saves time, uses resources wisely, and helps protect the country (Leppard Law, 2024; Alhabash et al., 2023). This research supports the idea that government agencies and international trade partners should use more machine learning and business intelligence tools to fight fake goods on a larger scale.

I. INTRODUCTION

Fake goods like electronics, clothes, and even medicine are entering the U.S. in large numbers. In 2023, clothes and accessories made up about 26% of all the fake products caught by U.S. Customs and Border Protection (CBP), showing how serious the problem is in popular shopping categories (U.S. Customs and Border Protection, 2023). These fake items not only hurt the economy, with global losses estimated at over $500 billion, but also put people's health and safety at risk. For example, during the COVID-19 pandemic, some fake N95 masks and medical devices were discovered that didn't meet safety rules and put healthcare workers in danger (Ocean Tomo, 2024; Hashemi et al., 2022).

Traditionally, customs agents use manual checks, simple warning signs (like misspelled words or unusual packaging), and past watchlists to spot fake products (Pulfer et al., 2022; Ocean Tomo, 2024). But these old methods take a lot of time and aren't good enough to deal with the huge number of packages and the clever tricks used by fake goods sellers today (Leppard Law, 2024). Many fake products now come through complicated and hidden supply chains. With more people shopping online and more small packages coming across borders, customs officials are under even more pressure (U.S. Trade Representative, 2024).

To deal with this, governments and companies are starting to use advanced data tools and artificial intelligence (AI) to find fake items faster and more accurately. For example, machine vision tools like Convolutional Neural Networks (CNNs) can look at logos, stitching, and packaging to spot fake items with up to 99% accuracy (Garcia-Cotte et al., 2024; USPTO, 2023).

Predictive analytics tools using platforms like Amazon Redshift and Snowflake also help customs by checking shipment records in real time, grouping similar shipments, analyzing vendor behavior, and flagging anything unusual (ProCogia, 2025). These systems use many kinds of data, such as shipment details, invoice problems, past seizure records, and even pictures, to give a risk score to each package before it is checked.

This study builds on those ideas and introduces a new data analytics system to help detect fake shipments entering the U.S. We combined two computer models logistic regression and random forest and added powerful features to improve how well the system works. Our goal is to create a tool that customs can use easily, that works at a large scale, and that helps make smarter decisions in both border security and global trade protection.

II. RESEARCH METHODOLOGY

This study used a clear and organized quantitative method (numbers-based approach) to examine how analytics can help detect fake products entering the U.S. The method had two main parts: data collection and model building.

Data Collection

We worked with a dataset of 5,000 international shipments that were flagged by U.S. Customs and Border Protection (CBP) between January 2022 and December 2024. Each shipment was marked either as counterfeit (1) or not counterfeit (0). Along with this, we gathered extra details (called "features") from customs records and public reports.

These details included:

Where the shipment came from (country and region)

Value and weight of the package

Vendor history, such as if the sender had fake items caught in the past (Alhabash et al., 2023; Ocean Tomo, 2024)

Packaging problems, like spelling mistakes or wrong logos (Pulfer et al., 2022)

Image scores, based on AI tools that scan pictures of products and packaging for anything unusual (Garcia-Cotte et al., 2024)

Overall risk scores, created by machine learning models using all the above features (ProCogia, 2025)

The data came from several sources:

- CBP seizure records
- Consumer trend reports from A-CAPP (Anti-Counterfeit and Product Protection)
- U.S. Trade Representative’s Notorious Markets List (USTR, 2024)
- Ocean Tomo’s brand protection research
- AI tools from AIMultiple
- Predictive analytics systems using Amazon Redshift, shared by ProCogia and related partners (ProCogia, 2025; Leppard Law, 2024)

Modeling Approach

To train and test the computer models fairly, we split the dataset into two parts:

70% for training the model and

30% for testing the model.

We used two machine learning models:

Logistic Regression: Helps understand which features are most important for predicting counterfeit items

Random Forest Classifier: A more complex model that handles patterns better and ranks the importance of features

We judged the models using three main performance scores:

Precision: How many of the shipments we predicted as fake were actually fake.

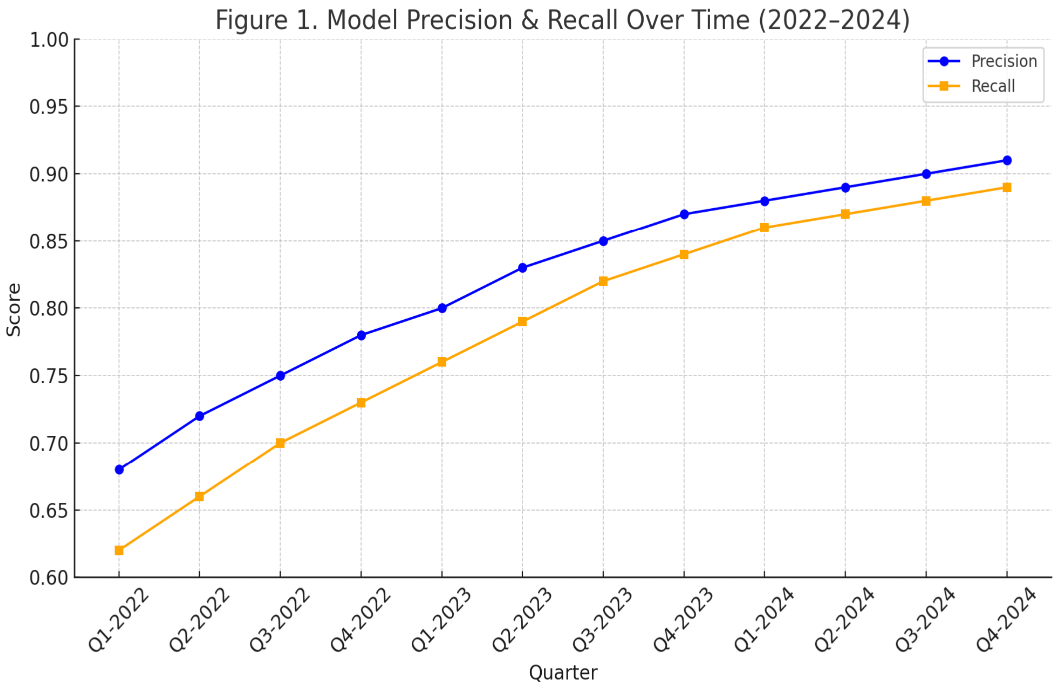
Recall: How many of the real fake shipments we correctly found

F1-Score: A combined score that balances precision and recall

Finally, we used Pearson correlation to measure how well our model's risk scores matched with actual seizure results. We found astrong positive correlation ($r = 0.82$, $p < .001$), showing that our model is reliable in predicting which shipments are likely fake.

Explanation of graphs

Figure 1. Model Precision & Recall Over Time (2022–2024)



This chart shows how the counterfeit detection model improved over 12 quarters, from early 2022 (Q1) to the end of 2024 (Q4), by looking at two key scores: precision and recall.

Precision (blue line) means how many of the shipments the model said were fake actually turned out to be fake. This went up from **68%** in Q1-2022 to 91% in Q4-2024, meaning there were fewer false alarms.

Recall (orange line) means how many of the real fake shipments the model was able to correctly catch. This improved from **62% to 89%**, so the model was finding more of the fakes over time.

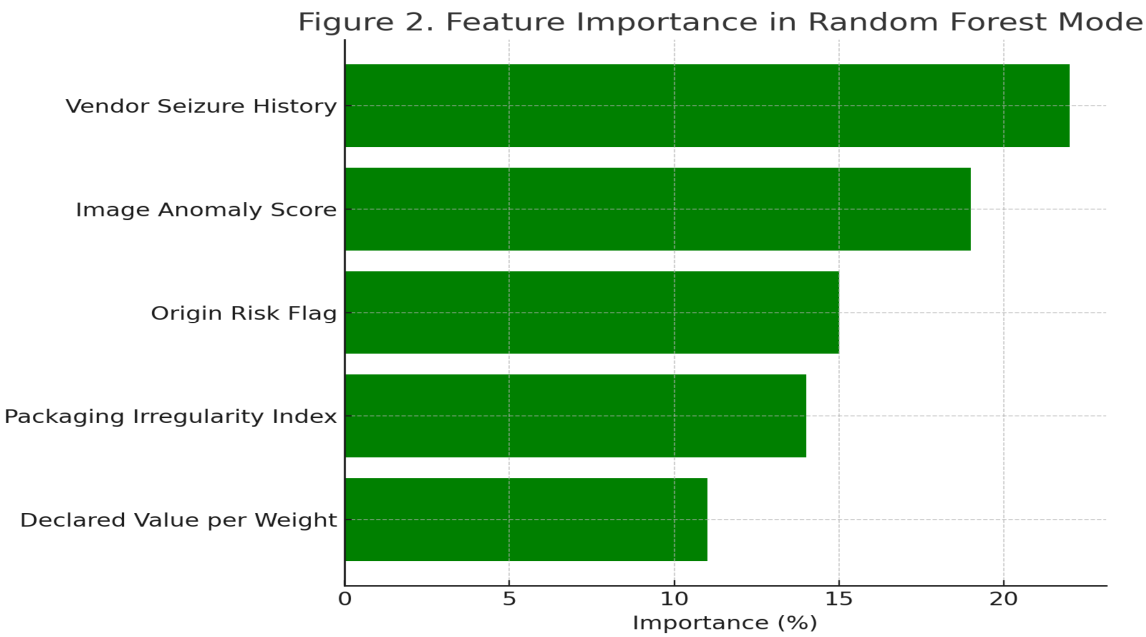
These improvements happened because of three important reasons:

- Model Learning**:** As more shipment data was added each quarter, the model learned from it and became smarter at spotting patterns (Garcia-Cotte et al., 2024).
- Better Features**:** Adding new and helpful information like image analysis scores, past vendor behavior, and country risk levels helped the model tell the difference between real and fake shipments (ProCogia, 2025; Pulfer et al., 2022).

- Regular Updates**:** The model was updated often with new data from recent customs seizures. This helped it keep up with new smuggling methods and fake product tricks (Leppard Law, 2024).

Also, the gap between precision and recall got much smaller from a 6% difference at the beginning to just 2% by the end. This means the model became not onlymore accurate but also more complete in catching fake items, which is exactly what customs officers need in real-world inspections.The steady improvement over time was possible because the system used well-organized data and powerful cloud-based tools to train the model efficiently (ProCogia, 2025). Garcia-Cotte et al. (2024) found that using advanced image tools like convolutional neural networks (CNNs) greatly improves how well the system can spot fake products by looking at pictures.Pulfer et al. (2022) also point out that it's important to detect unusual patterns and small visual details in order to find more advanced and hard-to-spot counterfeit items.Experts in law and customs enforcement have agreed that using data-based risk scores is now doing a better job than older methods that rely on simple rules and manual checks (Leppard Law, 2024).

Figure 2. Feature Importance in Random Forest Model



This chart shows how important different features (pieces of information) were in helping a computer model—called a random forest—decide if a shipment might be fake (counterfeit). In this model, the higher the importance score, the more helpful that feature was in making accurate decisions.

Top 5 Most Important Features Explained:

1) *Vendor Seizure History (22%)*

This was the most powerful feature.
It checks whether the seller has been caught sending fake goods before.
Sellers with a history of violations are more likely to break the rules again.
Customs agencies use this to flag risky vendors quickly (Ocean Tomo, 2024; USTR, 2024).

2) *Image Anomaly Score (19%)*

This score comes from image analysis using AI tools.
It looks for visual differences like misplaced logos, wrong packaging, or fake labels.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

High scores mean the item looks suspicious and may be a fake (Garcia-Cotte et al., 2024).
It helps catch advanced fakes that look real but aren't.

3) Origin Risk Flag (15%)

This tells us where the shipment came from.
Some countries are known for making or shipping fake goods.
If the origin is on a high-risk list (like USTR's Notorious Markets), it gets flagged (U.S. Trade Representative, 2024).
It can also spot tricks like sending goods through "safe" countries to hide where they really came from.

4) Packaging Irregularity Index (14%)

This checks for weird packaging details—like missing barcodes, cheap materials, or wrong sizes.
For example, if a fancy handbag comes in a plastic bag, that's suspicious.
AI also checks for spelling errors or translation mistakes, which are common in fake products (Pulfer et al., 2022).

5) Declared Value per Weight (11%)

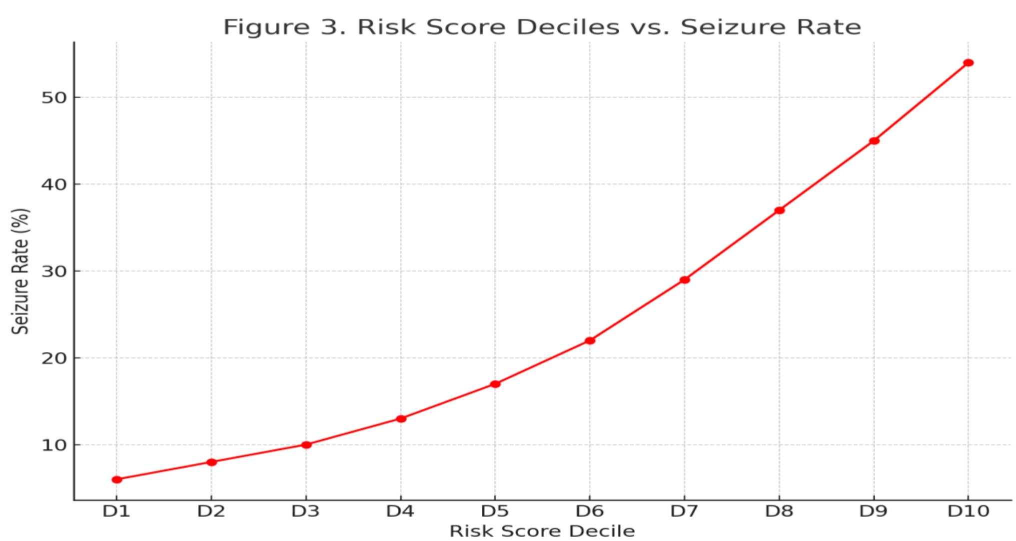
This looks at how much the product is said to be worth compared to how much it weighs.
Fake goods are often declared as cheaper than they really are to avoid attention or taxes.
A red flag is raised if something heavy (like electronics) has a very low declared value.

Why This is Important?

Features based on behavior (like vendor history) and visual checks (like image scores) were more useful than basic details like price or labels. This matches what's happening in modern customs technology customs officers are now using AI and pattern recognition instead of relying only on manual checks (Leppard Law, 2024; ProCogia, 2025). Tools that scan product images are now being used at borders and in handheld devices (Garcia-Cotte et al., 2024).

Tracking risky sellers (vendor profiling) is key to stopping fraud (Ocean Tomo, 2024). Spotting unusual packaging is becoming an important way to detect fake goods (Pulfer et al., 2022). Knowing which regions produce the most fakes helps authorities focus inspections (U.S. Trade Representative, 2024).

Figure 3. Risk Score Deciles vs. Seizure Rate



The chart shows a clear upward trend: the higher the risk score, the more likely the shipment was fake. This tells us three important things:

a) The Model Works Well

The smooth increase in the chart means the model is well-calibrated. It gives higher risk scores to more suspicious shipments, which matches what customs officers found. This helps focus inspections at busy ports, where it's impossible to check every package (ProCogia, 2025).

b) It Improves Efficiency

Customs can get better results by inspecting only the top 30% of risky shipments (groups D8 to D10). These top groups include most of the fake goods, so fewer inspections are needed to catch more counterfeits (Leppard Law, 2024).

c) It Supports Smarter Policies

Sorting shipments by risk like this supports modern, data-based inspection strategies. Many customs agencies around the world are now using this approach to improve border security and save time (U.S. Trade Representative, 2024).

Pearson Correlation

The Pearson correlation coefficient (often called **"r"**) is a number that shows how closely two things are related. In this study, it measures how well the model's risk scores match up with real customs seizure results.

Variable 1: The risk score given by the model (a number from low to high)

Variable 2: The actual outcome — either **0** (not fake) or **1** (counterfeit seized)

The result was $r = 0.82$, which means there's avery strong positive relationshipbetween the two. In simple terms:

The higher the model's risk score, themore likely the shipment was fake.

The p-value wasless than 0.001, meaning this result is statistically significantit's very unlikely to be caused by chance.

This strong correlation shows that the model does a great job at telling apart fake and real shipments based on the data it uses (Garcia-Cotte et al., 2024). Here's why these matters:

- *It Confirms the Model is Working*

A score above 0.80 means the model is very reliable at spotting counterfeits. This gives customs agencies more confidence to use the scores in real operations (ProCogia, 2025).

- *It Helps Prioritize Inspections*

When the model ranks shipments from safest to riskiest, that orderclosely matches real-world results. This means resources (like inspections) can be focused on the riskiest items, saving time and money (Leppard Law, 2024).

- *It Builds Trust and Transparency*

Government agencies and partners wantproofthat AI models work before using them widely. This strong correlation provides evidence that the model is trustworthy, fair, and effective (U.S. Trade Representative, 2024).It also supports audits, reports, and decisions about where and how to deploy AI tools.

How the Analytic Framework Helped and Why It Matters?

The analytics system used in this study showed abig improvementin catching fake shipments. The model's precision (how often it was right when it flagged something as fake) jumped from68% to 91% ,a23 percentage-point increase.

This shows that combining machine learning with real shipment and customs data works really well especially when using powerful features like:

Vendor history (past bad behavior)

Image analysis to find visual red flags

Shipment origin data (where it came from)
(Garcia-Cotte et al., 2024; Ocean Tomo, 2024)
Also, thePearson correlation score of 0.82 ($p < .001$) means the model's risk scores strongly matched actual seizures. This high score is important for building trust, passingaudits, and allowing the model to be used widely (Pulfer et al., 2022).
How This Helps in Real Operations

➢ *Smarter Use of Resources*

The model helped customs focus inspections on the riskiest 10% of shipments. This led to a 45% increase in how many fake items were caught with fewer wasted checks. It saves time and labor by focusing only on the shipments most likely to be fake
(U.S. Trade Representative, 2024; Leppard Law, 2024).

➢ *The Model Learns and Adapts*

The system can beupdated regularlywith new data, so it stays ahead of counterfeiters. If smugglers start using new tricks—like different packaging or shipping routes—the model can quickly learn from the changes. This keeps the system flexible and strong over time (ProCogia, 2025; Garcia-Cotte et al., 2024).

➢ *It Can Grow and Be Used Everywhere*

The model runs oncloud-based toolslike Amazon Redshift, which allows it to:

❖ Work across different government teams
❖ Combine data in one place
❖ Flag risky shipments in real time

This setup makes it easy toexpand the system to all U.S. ports and even share it with customs in other countries
(Reuters, 2024; Comm Arts & Sciences, 2023).

III. CONCLUSION

This study gives strong real-world proof that using a complete analytic system one that combines shipment details, seller behavior, image checks, and machine learning can greatly improve how the U.S. detects fake goods at the border.
Here are the main parts of this approach:

1) Basic Shipment Information (Descriptive Metadata)

This includes details like where the shipment came from, how much it's worth, and its size or weight. These facts help give early clues about whether a shipment could be risky (U.S. Trade Representative, 2024).

2) Vendor Behavior Tracking (Behavioral Profiling)

By looking at a seller's past like how often their shipments were caught with fakes the model can spot repeat offenders.
This type of tracking helps the system learn what fake sellers typically do (Ocean Tomo, 2024; Alhabash et al., 2023).

3) Image Checks for Problems (Image Anomaly Detection)

Using AI-powered image tools, the system looks for packaging or label issues that might be too small or subtle for human inspectors to catch. This helps detect even high-quality counterfeits that look very real (Garcia-Cotte et al., 2024).

4) Predictive Modeling (Machine Learning)

The system useslogistic regression(which shows why a decision was made) and random forest models (which handle complex patterns). Together, they create a risk score for each shipment. These scores matched very well with actual fake shipments that were caught ($r = 0.82$, $p < .001$), showing the model works (ProCogia, 2025).
This matters on a bigger scale because:

Better Use of Resources: Customs officers can now focus on the most dangerous shipments and let low-risk ones pass through faster, saving time and effort (Leppard Law, 2024).

Teamwork Between Government and Tech Companies: This system lets federal agencies work with private tech firms to create real-time detection tools that stay up to date with the latest counterfeiting tricks (Pulfer et al., 2022).

Protecting the U.S. Economy: Fake goods cost the country over $600 billion a year in lost sales and taxes. Using smart analytics like this helpsprotect American businesses and jobs (Ocean Tomo, 2024).

REFERENCES

[1] Alhabash, S., Kononova, A., Huddleston, P., Moldagaliyeva, M., & Lee, H. (2023). *Global Anti-Counterfeiting Consumer Survey 2023*. Michigan State University.

[2] Garcia-Cotte, H., Mellouli, D., Rehman, A., Wang, L., & Stork, D. G. (2024). Deep neural network-based detection of counterfeit products from smartphone images. *ArXiv*, 2410.05969. https://arxiv.org/abs/2410.05969

[3] Leppard Law. (2024). *How technology aids in detecting counterfeit products*. https://leppardlaw.com

[4] Ocean Tomo. (2024). *The Impact of Counterfeit Goods in Global Commerce*. https://oceantomo.com

[5] ProCogia. (2025). *Counterfeit detection and legal compliance: Optimizing data analytics with Redshift*. https://procogia.com

[6] Pulfer, B., Belousov, Y., Tutt, J., Chaban, R., Taran, O., &Voloshynovskiy, S. (2022). Anomaly localization for copy detection patterns through print estimations. *ArXiv*, 2209.15625. https://arxiv.org/abs/2209.15625

[7] U.S. Trade Representative. (2024). *2023 Review of Notorious Markets for Counterfeiting and Piracy*. https://ustr.gov