# Final Admissibility – Using Artificial Intelligence and Machine Learning to Modernize and Optimize the Imports Screening Process

Gregory Parcover[1], Indu Konduri[1], Jason Sadler[1], Gregory Jackson[2], Alex Hanson[3], Camryn Bishop[3], Jacqueline Guill[3]

[1]US Food and Drug Administration, Division of Import Systems Solutions, OISM, ORA, [2]US Food and Drug Administration, Office of Transitional Sciences, OCS, CDER, [3] Booz Allen Hamilton

## Imports Screening Overview

### Problem Statement

**The FDA would like to create a more efficient imports screening process by using a predictive model to analyze the final admissibility statuses set by regulatory officers and learn the parameters that define whether an entry line is violative or non-violative. As each commodity has its own unique qualities, a total of 26 models will be created.**

**This analysis focuses on Devices, Dietary Supplements, and Cosmetics, which were all trained on the 36 months of data preceding their test month (May – September 2021).**

### Current Imports Screening Process

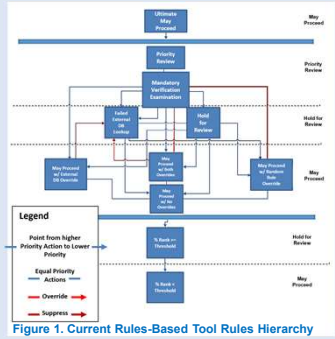


Figure 1. Current Rules-Based Tool Rules Hierarchy

- **As all foreign and domestic products regulated by the FDA are electronically screened, import oversight has proven to be an ever-expanding responsibility. Fiscal Year 2019 alone boasted 45 million imports, a steady increase of over 5%. With a limited staff, efficiency is more important now than ever.**
- **The current system in place begins with a manually-informed rules-based tool. This tool reviews each entry, allowing for multiple rules to fire for a single line. The recommendations assigned to a single rule are ranked in the hierarchy shown here as they generate a raw risk score for each entry line seeking passage.**
- **Using this risk score, the tool assigns each entry a risk percentile and compares these percentiles to a chosen threshold that varies due to the type of import that was screened. If the percentile is greater than the threshold, it is considered violative and recommends that the product be held for further review by a regulatory officer. Otherwise, it is denoted as non-violative and recommends that it is allowed to proceed. The regulatory officer receives the tool's recommendation, decide if they agree with it, and make a final admissibility recommendation for the import screening line.**

### Constraints with Current Rules-Based System

**While this application has worked well over the last ten years to streamline the previous imports screening process, there are two main constraints that an automated approach could rectify:**

**Lack of Assessment of Rules**

- **Difficult to simulate production data to test rules performance prior to implementation**
- **Lack of auditing protocol to ensure rules are updated and performing as intended**

**Infrequent Threshold Value Refresh**

- **Commodity thresholds are manually assessed for optimization yearly**
- **Thresholds are not updated immediately based on the assessment but rather are changed at the behest of the product center directors**

**Disclaimer: This presentation reflects the views of the authors and should not be construed to represent FDA's views or policies.**

## Preliminary Results from the Final Admissibility Model

### Current Rules-Based Tool's Performance

**Our metric of success here will be the F1 score, the harmonic mean of precision and recall where precision shows how many retrieved items are relevant and recall shows how many relevant items are retrieved.**

**While outputs vary from month to month, the current tool typically produces F1 Scores of 0.6 - 0.7. The goal for our models, is to improve upon those scores.**

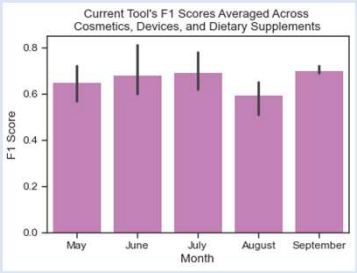


Figure 2. Current Tools' Average F1 Score by Month

### Finding a Suitable Classification Model

**To account for the unique natures of each commodity, four versions of Random Forest Classification were implemented:**

- **Random Forest Classifier (RFC – Base)**
- **Random Forest Classifier + Randomized Search Hyperparameter Tuning (RFC – Base Tuned)**
- **Random Forest Classifier + Undersampling (RFC – US)**
- **Random Forest Classifier + Undersampling + Randomized Search Hyperparameter Tuning (RFC – US Tuned)**

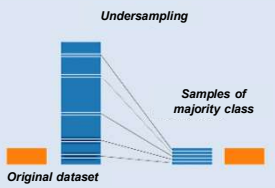


Figure 3. Undersampling Demonstration

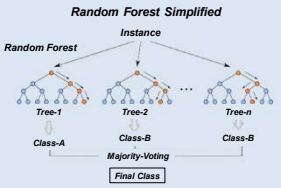


Figure 4. Random Forest Demonstration

### Rules-Based Tool vs Random Forest Classifier

**These three charts display the F1 scores over time for each of the models applied to each commodity. The current tool, in purple, and the best-performing model are bolded. While the tuned version of the base RFC produced the superior final output for cosmetics, the untuned versions happened to perform better for Devices and Dietary Supplements. Aside from May and June in the Cosmetics Commodity, all four versions of the Random Forest Classification Model outperform the current rules-based tool.**

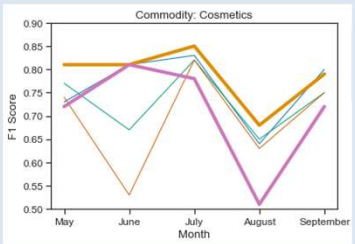


Figure 4. F1 Scores produced by all models in the Cosmetics commodity by month

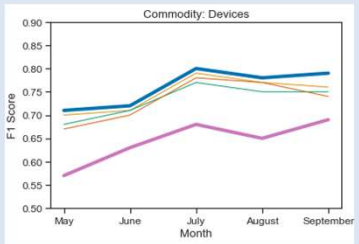


Figure 5. F1 Scores produced by all models in the Devices commodity by month

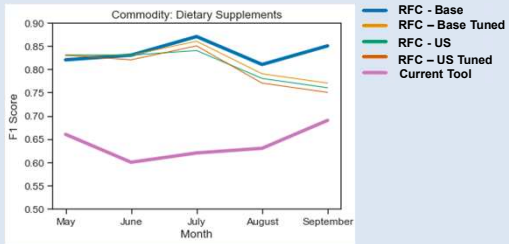


Figure 6. F1 Scores produced by all models in the Dietary Supplements commodity by month

### Random Forest Classifier – F1 Score Percent Increase

**To see the impact of these models, we analyzed the percent change of the F1 Score from the current tool to the consistent top two performers – base Random Forest Classifier, tuned and un-tuned.**

**While we see a lot of variability between the commodities, with Dietary Supplements displaying the biggest impact and Cosmetics displaying the least, we see the F1 scores increase by up to 40.3% compared to the current tool.**

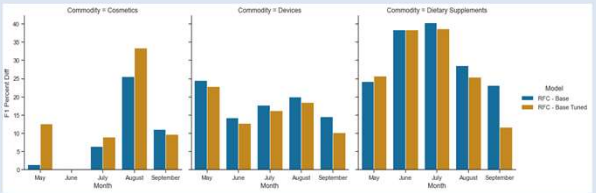


Figure 7. Percent increase of RFC Base tuned and untuned models' F1 Scores from the current tools' F1 Score by month

### Random Forest Classifier – Average F1 Score Percent Increase

**When averaged out between all the commodities, the percent change ranges from a 10-25.7% improvement on the current tool. We expect this range to be consistent as we expand the analysis to the other commodities.**

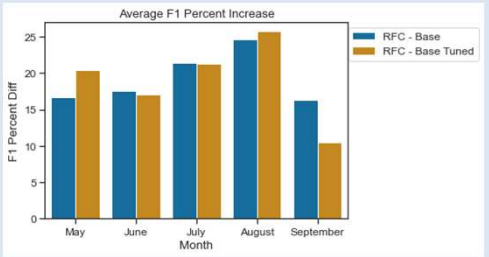


Figure 8. Average percent increase of RFC Base tuned and untuned models' F1 Scores from the current tools' F1 Score by month