

Article

# A Multi-Objective Bayesian Approach with Dynamic Optimization (MOBADO). A Hybrid of Decision Theory and Machine Learning Applied to Customs Fraud Control in Spain

Ignacio González García <sup>†</sup> and Alfonso Mateos Caballero <sup>\*,†</sup>

Decision Analysis and Statistic Group, Departamento de Inteligencia Artificial, Campus de Montegancedo s/n, Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Spain; ignacio.gonzalez.garcia@alumnos.upm.es

\* Correspondence: alfonso.mateos@upm.es

† These authors contributed equally to this work.

**Abstract:** This paper studies the economically significant problem of the optimization of customs fraud control, which is a critical issue for many countries. The European Union (EU) alone handles 4693 tons of goods every minute (2018 figures). Even though 70% of goods are imported at zero tariff, the EU raised EUR 25.4 billions in 2018, and customs-related income transferred by member states to the EU accounts for nearly 13% of its overall budget. In this field, (a) the conflicting objectives are qualitative and cannot be reduced to a common measure (security and terrorism, health, drug market access control, taxes, etc.); (b) each submitted item has dozens of characteristics; (c) there are constraints; and (d) risk analysis systems have to make decisions in real time. Although the World Customs Organization has promoted the use of artificial intelligence to increase the precision of controls, the problem is very complex due to the data characteristics and interpretability, which is a requirement established by customs officers. In this paper, we propose a new Bayesian-based hybrid approach combining machine learning and multi-objective linear programming (MOLP), called multi-objective Bayesian with dynamic optimization (MOBADO). We demonstrate that it is possible to more than double (with a 237% increase) the precision of current inspection systems, freeing up almost 50% of human resources, and outperform past results with respect to each of the above objectives. MOBADO is an optimization technique that could be combined with any artificial intelligence approach capable of optimizing the quality of multi-objective risk analysis in real time.



**Citation:** González García, I.; Mateos Caballero, A. A Multi-Objective Bayesian Approach with Dynamic Optimization (MOBADO). A Hybrid of Decision Theory and Machine Learning Applied to Customs Fraud Control in Spain. *Mathematics* **2021**, *9*, 1529. <https://doi.org/10.3390/math9131529>

Academic Editor: Juan Alfredo Aguarón and María Teresa Escobar

Received: 24 May 2021  
Accepted: 23 June 2021  
Published: 29 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** optimization; Bayes; customs; fraud; decision theory; machine learning

## 1. Introduction

The action of making the best use of resources for a purpose is called optimization. In the field of social sciences, many problems are characterized by the coexistence of multiple criteria and conflicting objectives. This paper studies the economically significant problem of the optimization of customs fraud control. Even though 70% of imports now enter the EU at zero tariff, nearly EUR 25.4 billions were raised in 2018, and the amount transferred by member states to the EU accounts for nearly 13% of its overall budget [1]. This economic importance is greater in South America and less advanced economies.

The decision objects, namely imported goods, are specified on customs declarations, known as single administrative documents (SADs). SADs contain information about the characteristics—type, origin, value, etc.—of the goods, and customs offices must decide which have to be inspected, taking into consideration many conflicting objectives (trade facilitation, tax enforcement, security, defense of intellectual property, drug smuggling prevention and many others). In other words, the decision has multiple attributes.

In this problem, (a) the conflicting objectives are qualitative and cannot be reduced to a common measure (security and terrorism, health, drug market access control, taxes, etc.); (b) each item has dozens of characteristics; (c) there are constraints; and (d) the

decision-making process is continuous, and successive decisions must be made continuously. Therefore, the risk analysis system has to make decisions in real time.

We propose a new Bayesian-based hybrid approach combining machine learning and multi-objective linear programming (MOLP), called multi-objective Bayesian with dynamic optimization (MOBADO).

In this paper, we provide a detailed description of the problem in Section 2 and a description of the MOBADO methodology in Section 3. Results are reported in Section 4 and are discussed in Section 5. The conclusions are presented in Section 6.

## 2. Description of the Problem

In the past, customs offices inspected each and every declared import. The globalization of commerce imposed a clearance philosophy dominated by risk analysis (World Customs Organization, Compendium, (I), p. 3). In the 1970s, customs administrations deployed the concept of channels. A rule-based system sends declarations (SADs) to different channels: green for immediate clearance, orange for document control and red for physical control.

Most risk analysis systems, such as the United Nations ASYCUDA system used by more than 90 countries, include a list of Boolean conditions (e.g., if code of the good = 84.02.01.01.00 and country of origin is China) associated with actions. Each rule is associated with a distribution. For a specific combination expressed with a rule, the system assigns, for example, 85% of declarations to the green channel, with goods going through customs formalities without inspection, 10% to the orange channel for document inspection, and 5% to the red channel for physical inspection. The systems provide the list of actions to be performed by the inspector in each case. Specific controls have been established on the grounds of health (e.g., vaccines, mask or protective materials quality), tax enforcement or security. More specifically, the orange channel leads to document controls, where the inspector must check certificate validity or declared values by inspecting invoices and price lists. The system is maintained by groups of experts (e.g., the Spanish Tax Agency's system contains more than 1200 rules) that translate the information gathered by risk analysis units and promote administrative cooperation.

These rules are independent and the output is a "list of things to do", control actions, inspections, which can produce many types of results: suspension and seizure in the case of drug smuggling or terrorism, settlement of customs or other excise/alcohol duties, VAT, sanctions, etc.

We considered two problems:

- **Problem 1.** Create an evidence-based algorithmic method to optimize decisions in this type of multi-attribute problem.
- **Problem 2.** Quantify the benefits in terms of precision and efficacy of the proposed solution with respect to current techniques.

To sum up, we studied a dynamic, real-time, multi-objective risk optimization problem in the field of customs risk analysis.

This research reviewed the following:

1. The literature on analytical optimization problem-solving methods and decision-making methods, which is included in [2–4], and the methods for solving MADM problems [5].
2. Risk analysis approaches [6], specifically when applied to customs [7–9], including the Bayesian approaches [10], and many studies addressing customs fraud control [11–15].
3. Studies related to the use of machine learning and artificial intelligence (AI) techniques, using decision trees, neural networks and hybrid systems [16,17].
4. Best practices, including smart customs in Korea, the use of deep learning in Nigeria [14], machine learning in New Zealand [18,19] and the Netherlands [20–22] using Bayesian networks and deep learning, AI use in Canada [23], the Connect system of

the U.K. [24], and Brazil’s SISAM system [25,26] as well as China’s and Indonesia’s approaches [27,28].

We found that decision makers use many approaches to search for the best solution. These approaches are described in the literature as multi-criteria decision making (MCDM), multi-attribute decision making (MADM), multi-criteria decision analysis (MCDA), multi-attribute decision analysis (MADA) or multi-criteria analysis (MCA). They use a plethora of methods: AHP (analytic hierarchical process), MAUT (multi-attribute utility theory), SAW (simple additive weighting), TOPSIS (technique for ordering preference by similarity to an ideal solution), COPRAS (complex proportional assessment), MOORA (multi-objective optimization on the basis of ratio analysis) or PROMETHEE (preference ranking organization method for enrichment evaluation).

We studied the evolution of best practices. In a continuing trend since the 1980s, the number of physical controls has been reduced [29,30]. Some countries physically control 25% of imports, but the rate of physical and document control is very low in most of the advanced customs administrations, for example, 5.53% in Brazil [31]. The problem of control optimization is becoming critical. According to the Pitney Bowes Index, more than 100,000 million parcels will be sent globally in 2021, which is a 35% increase over 2016.

### 3. Problem Settings and Methodology

#### 3.1. Problem Settings

We use as a toy example of the problem, see Table 1, a case with two features (HS10 and country), two attributes (precision in tax levied (T) or suspension (S)), and three constraints.

Table 1. Precision estimated in toy example.

N°	HS	C	Rule	R	Precision		Results		P(HS10)	P(HS * C)	P(T)	P(S)
					ch.	nc	T	S				
1	11	A	[90;10]	Yes	1	1		1	25/61	1/1		100
2	24	B	[90;10]	Yes	1	1		1	1/1	1/1		100
3	37	C	[90;10]	Yes	30	5	3	2	$\frac{5}{30}$	$\frac{5}{30}$	$\frac{3}{30}$	$\frac{2}{30}$
4	14	D	V	Yes								
5	11	E		Yes	60	24	12	12	$\frac{25}{61}$	$\frac{24}{60}$	$\frac{12}{60}$	$\frac{12}{60}$
6	90	F		Yes	100	20	2	18	$\frac{23}{120}$	$\frac{20}{100}$	$\frac{2}{100}$	$\frac{18}{100}$
7	84	G	V	Yes								
8	87	H										
9	91	I		Yes	1	1	1		1/1	1/1	1/1	
10	90	K	[50;50]	Yes	20	3		3	23/120	3/20	3/20	

Declarants submit SADs to customs continuously. Each SAD can contain one or more types of goods in which case they are declared as successive *sad.items*. A subset of ten *sad.items* is shown in the rows of Table 1. Many variables, including HS10 code, are declared for each *sad.item*, indicating product type (e.g., a phone) and country of origin (e.g., China).

Customs offices must consider unrelated objectives (i.e., tax enforcement, drug detection). When a forbidden good is detected, the good is seized and no taxes are settled, but both objectives can, in some exceptional cases, be achieved, for instance, when a drug is hidden in a container of undervalued fruit without the owner’s knowledge.

There are many constraints:

1. The control of some specific objectives is mandatory (i.e., all controls related to terrorism must be executed, and 100% are channeled to red).
2. The total percentage of *sad.items* inspected cannot exceed a specified percentage of all documents. Customs have to reconcile control with trade facilitation.
3. There is an established level of service, and decisions must be made in a time window due to logistical constraints.

Suppose that a set of 100 *sad.items* is submitted to the customs office and entered in the rule-based system. The customs office cannot accumulate more than a set of, for example, 10 documents pending decision because it has to maintain the level of service. The decision maker must optimize decision making and should control a maximum (e.g., 8%) of documents within this group.

In Table 1, *sad.item* 1 has the code  $HS10 = 11$ , and country of origin A. The rule-based system establishes that this item should be cleared, that is, assigned to the green channel, 90% of the time. There is past evidence that this type of good with this origin was inspected once and was found to be non-compliant. It was a case of drug smuggling, and the result was a seizure with suspension of clearance (S). The decision maker should decide, taking into account the evidence and subject to the constraints, which documents of the group should be inspected, and repeat the decision afterward for successive sets of documents, optimizing control.

Using the precision of the controls established in the past as risk criteria, a *sad.item* declared with  $HS10(11)*Country(A)$ , should be associated with a precision of 100%. The decision can be made taking in account the probability of non-compliance using a combination of one feature or several features. If the decision maker were to consider only the feature  $HS10(11)$ , the available data are with  $nc = 24 + 1 = 25$ , and inspected items =  $60 + 1 = 61$ ; therefore, the estimated precision would be  $25/61$ .

The decision maker has to deal with at least five issues:

1. Absence of a vector of preferences: It is not reasonable to ask customs experts to create a vector of preferences for such a complex combination of qualitatively diverse objectives (decide between security risks vs. health risks) because these are political and ethical questions to which no firm answer can be obtained in writing. In other words, it is impossible to get a vector of criterion weights or a prior articulation of preferences.
2. A sparse scoring matrix: The decision will be based on evidence of fraud probabilities. Due to their high cardinality, however, the evidence is sparse. For example, there were 131,854 possible combinations of  $HS10$  and country in 2019, of which 44,111 were inspected, with non-compliance being found in only 5320 cases, that is, 4.03 % of the combinations. If we consider a three-element feature cross, such as an  $HS10*country*exporter$ , we will have a scoring matrix where only 1/10,000 of its elements have non-zero content.
3. Levels of service: Declarants should be given an answer in more or less real time. It is not acceptable for documents to be retained pending the best decision. At all customs offices, there is a limit on the number of documents or the time that they have to wait for a decision.
4. What are the best predictors of fraud?
5. What are the criteria for deciding the best option?

Therefore, the problem can be stated as follows.

**Problem:** Given an import trade flow  $t$ , along with its importer  $u$ , declarant  $d$ , and HS code  $c$  and country of origin, optimize the controls, maximizing  $X_r$  objectives subject to constraints.

### 3.2. Methodology

Precision ( $p$ ):

$$p = \frac{\text{non-compliances}}{\text{total inspections}}. \quad (1)$$

Efficacy ( $e$ ): The efficacy of  $b$  with respect to  $a$  is the ratio of non-compliance:

$$e_{b/a} = \frac{\text{non-compliances}_b}{\text{total inspections}_a}. \quad (2)$$

*Features and cross-features of a sad.item.* One SAD contains more than 50 boxes with data, such as the  $HS10$  code, country of origin, declarant, importer, customs value, etc. We use the term feature to denote an attribute of the *sad.item* (e.g.,  $HS10$  code or tax identification

number), and cross-feature for a combination of two or more features (e.g., HS10\*Country of Origin). For simplicity of notation, we will always use cross-feature ( $f \in F$ ), even if there is only one feature. Therefore, each *sad.item* ( $s$ ) is defined by its cross-feature ( $f$ ).

*Flow of commerce.* It is the flow of goods submitted to the customs office. Each element of this flow is identified by the combination of the customs office code (e.g., 2801 Madrid airport, type (e.g., I = Import), year and a sequential number. We simplify the notation by referring to customs code and type of commerce.

Suppose that we have a continuous import trade flow documented with *sad*  $s_i$ ,  $i = \{1, 2, \dots, n_t\}$ , each of them composed of  $k_i$  *sad.items* that are cleared with a stopping time, and  $n_t$  the total number of SADs so far  $t$  from the opening of the customs office (i.e.,  $s_i$  would be the  $i$ th *sad* submitted at a customs office).

*Alternatives.* In each time period  $t = (t_1, t_2]$ , the customs office must select a subset  $A_t \subset \{s_i : i = n_{t_1} + 1, \dots, n_{t_2}\}$ , verifying a set of constraints for control (e.g., the number of *sad.items* with cross-feature  $f$  must represent a percentage of the total). Each group of selectable elements is an alternative. Therefore,  $A_t = \{a_{it} : it = n_{t_1} + 1, n_{t_1} + 2, \dots, n_{t_2}\}$ , where  $a_{it}$  is composed of  $n_{f_{itk}}$  *sad.items* with cross-feature  $f_{itk}$ ,  $a_{it} = (f_{it1}, \dots, f_{itn_i})$ .

*Objectives.* There are  $X_j$  objectives  $j = \{1, 2, \dots, n\}$ , such as tax, health, drugs control, etc.

*Estimated Bayesian precision.* We use a Bayesian approach. In each period  $t$ , the ratio between non-compliance ( $nc_{jt}$ ) and inspected *sad.items*  $ch_{jt}$  is known for each objective  $X_j$ . Let us suppose that, at the beginning of the period, the precision for objective  $X_j$  of cross-feature  $f$  was  $nc_{j0}/ch_{j0}$  and we know that the results of controlling these goods in the period is  $nc_{jt}$  after inspecting  $ch_{jt}$ . According to Bayes, the estimated precision for objective  $X_j$  at the end of  $t$  is as follows:

$$\hat{p}_{j(t+1)}(f) = \frac{nc_{j0} + \dots + nc_{jt}}{ch_{j0} + \dots + ch_{jt}} \tag{3}$$

With this definition and the current system, the precision values may be greater than 100%. The current system sends *sad.items* to the red channel, in which case it is mandatory to physically inspect the goods. In orange or green channels, however, the inspector can decide to change the original channel and inspect the goods. If non-compliance is found in these cases, we have results that do not have a prior inspection order handed out by the system. The inspector can decide to control four parcels, leading to three results, whereas the system ordered only one parcel to be controlled. In this case, precision is 300% (three results for one inspection ordered by the system).

*Precision for an objective ( $X_j$ ) of an alternative ( $a_{it}$ )* is as follows:

$$p_j(a_{it}) = \frac{\text{non-compliances for objective } X_j}{\text{total inspections in } a_{it}}$$

*Precision of an alternative ( $a_{it}$ )*, given by the precision resulting from the control of each objective  $X_j$  in  $t$  is as follows:

$$p(a_{it}) = (p_1(a_{it}), \dots, p_n(a_{it})), \forall i \in \{1, 2, \dots, n_{t_2} - n_{t_1} - 1\}.$$

*Record.* Overall precision for the previous year,  $(r_1, \dots, r_n)$ , where each  $r_j = (r_{j1}, \dots, r_{jN})$ , that is, precision for objective  $X_j$  in the different cross-features.

We use precision as a bridge for evidence optimization. We choose an alternative in  $A_t$  in each period  $t$  using three elements:

1. Cross-features  $f$  of the *sad.items*  $a_i$ .
2. Bayesian evidence for each cross feature  $f$  and objective  $X_j$ .
3. Record for the objectives  $(r_1, \dots, r_n)$ .

The problem is how to choose the best  $f$ , that is, the best characteristics of the documents, to capitalize on the evidence and predict  $(\hat{p}_1(a_{it}), \dots, \hat{p}_n(a_{it}))$ .

*Optimal cross-features*

An optimal cross-feature maximizes the discrimination of distributions of precision. We first studied *HS10 \* country*, which is used by current systems. Afterward, we stated the problem in terms of signal detection theory (SDT). Detection theory is a means to measure the ability to differentiate between information-bearing patterns. It is used by psychologists to measure how we make decisions under conditions of uncertainty, such as how we perceive distances in foggy conditions or during eyewitness identification [32]. This approach considers the distribution of something (precision, intensity) produced by two sources (signal and noise or, in this case, compliances and non-compliances), represented in Figure 1 (left) by two normal distributions.

A receiver operating characteristic curve or ROC is a graph of a function that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. In this case, the variable used is precision. We use this theory, considering, by analogy, non-compliance as a stimulus and *inspection decision* as signal present [33,34].

Different cross-features, see in Figure 1  $\{f_1, f_2\}$  have different functional relationships with precision, and a different discrimination threshold should be established for each one. Figure 1 (left) illustrates that a *sad.item* can be qualified with a different  $eB_p$  using different cross-features ( $f_1$  or  $f_2$ ) with a unique discrimination threshold.

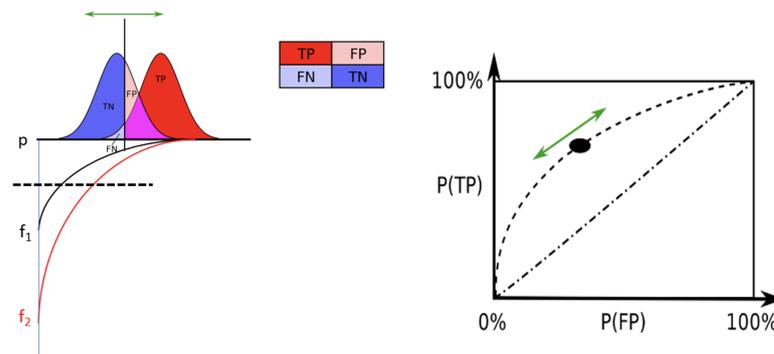


Figure 1. SDT concepts.

Table 2 shows the outcomes of a binary classifier. The classical psychological approach is shown on the left side (grey), and our problem is illustrated on the right side (white).

Table 2. Outcome of a binary classifier.

Observed	Prediction		Observed	Prediction	
	Signal Present	Signal Absent		Fraud	Compliance
Stimulus present	Hit	Miss	Fraud	True positives (TP)	False negatives (FN)
Stimulus absent	False alarm	Correct rejection	Compliance	False positives (FP)	True negatives (TN)

The notation that is used in customs analysis is as follows:

$$\begin{aligned}
 \text{Inspection rate} = IR &= \frac{TP+FP}{TP+TN+FP+FN}, \\
 \text{Accuracy} = A &= \frac{TP+TN}{TP+FP+TN+FN}, \\
 \text{Recall} = R &= \frac{TP}{TP+FN}, \\
 \text{and } F1 &= \frac{2RP}{R+P}.
 \end{aligned}$$

Based on the proportions of these types of trials, numerical estimates of sensitivity can be output using statistics, such as the sensitivity index [35].

The capability of each feature to discriminate distributions differs. In other words, for some features, the ROC is further away from the non-discrimination diagonal, and the AUC is bigger. Using other words, some variables are better at predicting fraud than others. Using different thresholds, we get different inspection and seizure rates [36], and it is possible to select the threshold that optimizes different criteria. The more false positives there are, the more seizures take place at the cost of reducing control precision. This is illustrated in the bottom right-hand corner of Figure 1 based on data output by the control of electronic commerce in China [36].

### Modeling the problem

This type of problems cannot be regarded in strict machine learning terms, that is, as classification problems where the system is trained, using techniques such as discrimination trees or neural networks, to inspect documents that have a higher probability of detecting fraud, because this approach is unable to process the problem constraints and the existence of multiple attributes.

These constraints include, at least, the minimum and maximum number of controls for each attribute, and maximum delay for clearance. On the other hand, a decision has to be made on how to express the structure of the problem. We evaluated several approaches:

1. A MCDM problem, traditionally expressed as follows:

$$\text{Max}(p_1(a_{it}), \dots, p_n(a_{it})), \text{subject to } a_{it} \in A_t.$$

This approach cannot be used to address our problem because it is impossible to obtain a function from  $(p_1(a_{it}), \dots, p_n(a_{it}))$  without a vector of preferences and a function for maximization.

2. An optimization problem, expressed in mathematical terms as follows:

$$\text{opt}_{x \in \mathcal{D}} f(x)$$

where  $f(x) : \mathcal{D} \rightarrow Y$  is the objective function or criterion;  $D$  is the set of permissible areas of the optimized objects,  $x \in \mathcal{D}$  is an element, and  $\text{opt}$  is the minimum or maximum value of the function  $f(x)$ . We conclude that it is possible to address the problem as a continuous optimization if each decision is optimal. Bayesian decision theory is concerned with choosing between different decisions based on one specific type of information: evidence. The information space is  $\Theta$ , and each piece of information in this space is  $\theta$ , such that  $\theta \in \Theta$ . To solve this problem, we need to define a loss function:

$$\mathcal{L} : \Theta \times \mathcal{D} \rightarrow \mathcal{R}.$$

We propose a five-stage methodology (MOBADO) as depicted in Figure 2.

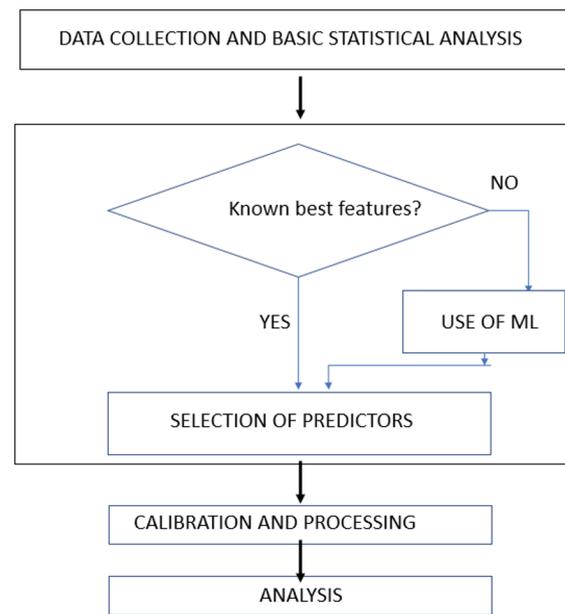


Figure 2. MOBADO stages.

1. *Data collection.* The source of the data are SAD documents, retrieved from the Spanish Tax Agency database. We have used only one table containing all the information related to SADs and clearances. It contains eight million rows/year, but the number of data related to inspections and non-compliance is around 0.5 million (see Table 3). In this specific case, the data were downloaded in CSV format from a SAP warehouse and were processed with R. In other cases (Section 5) requiring the use of all the data, we used Python.

We conducted a basic descriptive analysis of variables (see Table 3).

2. *Objective definition.* We believe that MOBADO is the first approach to multi-objective optimization in customs. Until now, customs considered each objective (control of drugs, health, etc.) separately because mostly customs offices have to run checks ordered by other authorities.

Looking at the collected data, data scientists should consider the different types of output from different inspections, number of records processed by the customs office, sample sizes, etc., to decide the best approach.

Table 3. Example of decision-making process.

	<i>a</i>	<i>b</i>	<i>r(a)</i>		<i>r(b)</i>		Total		$\Sigma$
			<i>nc(x<sub>1</sub>)</i>	<i>nc(X<sub>2</sub>)</i>	<i>nc(x<sub>1</sub>)</i>	<i>nc(X<sub>2</sub>)</i>	<i>nc(x<sub>1</sub>)</i>	<i>nc(X<sub>2</sub>)</i>	
	0	100							
	10	90							
Feasible	20	80	2	12	24	32	26	36	0.1381
	30	70	3	18	21	28	24	39	0.1179
	40	60		24	18	24	22	42	0.0976
	50	50	5	30	15	20	20	45	0.0774
	60	40	6	36	12	16	18	48	0.0571
	70	30							
	80	20							
	90	10							
	100	0							

We suppose that, when inspecting a *sad.item* with rules created with features  $f_1 = \text{“Coffee from Brazil”}$  and  $f_2 = \text{“imported by a new postal commerce declarant”}$ , we obtain  $p(f_1) = (0.1, 0.6)$  and  $p(f_2) = (0.3, 0.4)$ , always as a percentage of inspected *sad.items*, where 0.1 is the probability of *nc* for objective  $X_1$  and 0.6 for  $X_2$  when  $f_1$  is inspected.

Let us suppose that the flow of imports was 2500 *sad.items*. Therefore, 200 is the maximum number of documents to be inspected.

Let us suppose that the risk analysis units have already assigned 100 *sad.items* to the red channel for control. Therefore, the customs office must optimize the control of an additional number of 100 documents. Suppose also that the riskiest *sad.items* are of two types, *a* and *b*. Items labeled *a* are selected using  $f_1$  and those labeled *b* using  $f_2$ . The customs office must decide to control  $(x, 100 - x)$  items of each type. What would be the best decision? Table 3 shows the expected results using Bayesian inference for each objective for several values of  $x(20, 30, 40 \dots)$ . For example, when inspecting 60 items of class *a*, we could expect 18 non-compliance cases for the first objective and 48 for the second one. It is a matter of selecting the best  $x$ , and this requires knowledge of which of the possible expected outcomes (e.g.,  $A = (26, 36), \dots, B = (18, 48)$ ) is optimal.

The selected criteria is to outperform the last “record” obtained in the previous year. If decision makers cannot express their preferences, it is a good strategy to accept criteria that continuously outperform past results. Figure 3 illustrates the rationale. The ordinates on the plane are counts of the control of the attributes. Customs offices must control a minimum number of *sad.items* looking for drugs and a minimum of such items for tax enforcement (undervaluation), and the total number of controls cannot be above a specified threshold. Therefore, the feasible region for control is a triangle. Given a new *sad.item*, its control will be feasible if every constraint is satisfied.

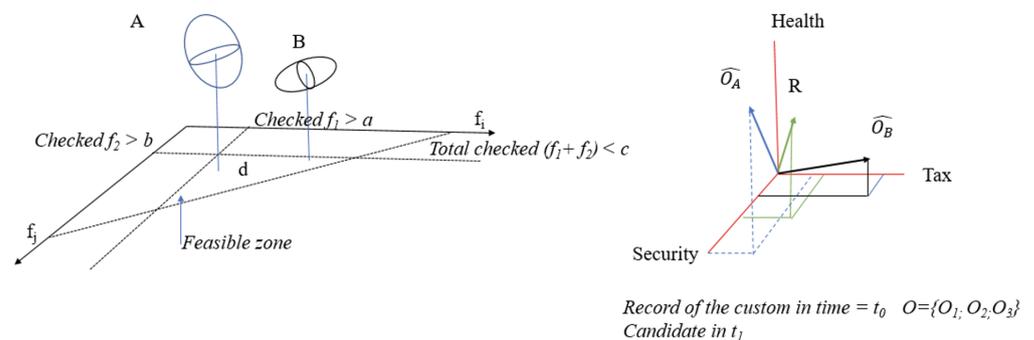


Figure 3. Dimensions of the decision.

Let us suppose that there are two possible options at any time, A and B, and we must decide which of them should be inspected with a record of  $(0.21; 0.40)$ . Let us suppose that the control of subset A will get an expected output of  $(26, 36)$ , and the control of subset B yields  $(18, 48)$ . Which subset should be inspected? For each of these results, only one subobjective can be improved. We select the option that most reduces the distance to the objective in the most delayed dimension.

We represent non-compliance vs. inspected features for each element on the  $z$  ordinate—precision. It is possible to study the precision for this point in each dimension.

If we have  $A_t = \{a_1, a_2, \dots, a_m\}$  candidates to be controlled, each  $a_i$  with  $(p_1(a_i), \dots, p_n(a_i))$  attributes and each  $p_j(a_i) = (p_j(f_{i1}), \dots, p_j(f_{in_i}))$ , where  $(f_{i1}, \dots, f_{in_i})$  are the cross-features of  $a_i$ :

$$\text{Min}_j \left\{ \text{Min}_i \left\{ \frac{p_j(a_i) - r_{ij}}{r_{ij}} \right\} \right\}.$$

In our case,

$$\text{Min} \left\{ \text{Min} \left\{ \frac{26 - 21}{21}; \frac{36 - 40}{40} \right\}; \text{Min} \left\{ \frac{18 - 21}{18}; \frac{48 - 40}{40} \right\} \right\} = \frac{18 - 21}{18} = -0.16.$$

Therefore, we should choose B.

3. *Selection of a fraud predictor.* We need (Figure 1 left) to decide the best feature, in other words the best function to discriminate compliance and non-compliance and afterwards to get the best threshold. We will use machine learning techniques to select the best predictor.
4. *MOBADO calibration and processing.* We will specify the threshold, that is, the value of the predictor that optimizes the decision, using SDT. This threshold is affected by the distribution of precision (in the classical approach to Gaussian curves, and in this case, a Bayesian estimation of these values calculated from the experience of past years at each customs office), that is, the stopping rule established to satisfy the level of service and constraints.

For the decision on the stopping rule, we used a strategy that applies the secretary problem concept built in decision theory [37]. The problem is to select the best of a set of candidates in a succession of interviews after which each candidate is immediately accepted or rejected. The best strategy is to select the first candidate that outperforms the first group passing the interview. On the boundary, we know that the probability of getting the highest-ranking list item is  $1/e$ , where compliance is randomly distributed over time [37].

If, as in our case, we want to select the best candidate to be inspected from a set of documents submitted in succession and we are obliged to decide whether or not each one is the best without any knowledge about the future *sad.items*, the solution can be to apply the above mentioned rule.

Figure 4 illustrates the comparison between the optimal strategy and random selection from  $N = 1$  to  $N = 50$ . The values when B changes from 1 to 9 are shown in the top left corner. The solution converges toward  $1/e$ , that is, 0.368.

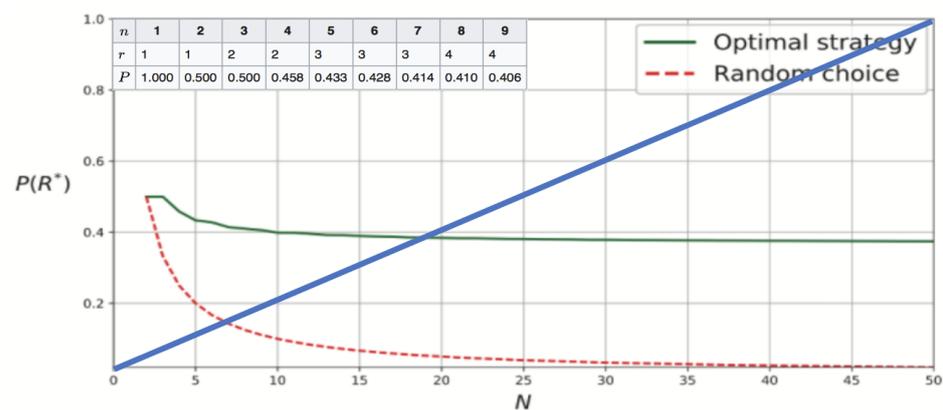


Figure 4. Probabilities and costs associated with different stopping times.

Figure 4 represents the cost of withholding the release of up to 50 documents. The document submitted first would have to wait 49 time intervals, the second one 48 and so on. For trade facilitation purposes, customs offices must balance the cost of the decision to impose a stopping time, for example, a half-hour delay, against the improvement of the optimized record.

5. *Validation of results and analysis of external validity.* The findings are evaluated according to a two-part process:
  - (a) *Counterfactual evaluation of benefits and calibration using training data.*

The proposed contrafactual was as follows: what should the results of using MOBADO in 2019 have been? For each day in the past, we know what documents were inspected following the instructions of the current rule-based system, the number of associated compliance and non-compliance cases and the precision for that day.

We can estimate what should have happened using MOBADO in the feasible region to calculate the following:

$$\hat{R}_{MOBADO} = \frac{\hat{nc}}{inspected} \quad (4)$$

and estimate the improvement as a ratio and its distribution:

$$\hat{p} = \frac{nc + \hat{nc}}{Checked_{rule} + Inspected_{MOBADO}} \quad (5)$$

$$\hat{e} = \frac{nc + \hat{nc}}{nc} \quad (6)$$

(b) Testing of MOBADO using 2020 test data

We studied the external validity of calibrating the 2019 training data with 2020 test data.

In our research, we did the following:

- (i) Calibrated and analyzed MOBADO with 2019 data.
- (ii) Tested against current practices with 2020 data.
- (iii) Estimated the potential benefits.

After assessing the external validity of a classical approach verifying the increase in precision (training and test), we provide a counterfactual estimation of what could have been the benefits of MOBADO.

## 4. Results

### 4.1. Calibrating MOBADO (2017–2019)

#### Step 1. Data collection

Table 4 shows Spanish customs data for the period 2017–2019. Microdata that do not affect privacy, as is the case of sanctions, can be downloaded. In any case, we take full responsibility for the conclusions of this paper, which were not reviewed by the Spanish Tax Agency.

The bottom two rows list the number of *sad.items* and the number of inspections performed when goods are directed along the orange or red channels, that is, whenever clearance is not automatic:

1. The percentage of controls (inspections) is of the order of 7% of submitted *sad.items*. Customs offices cannot perform more than the above rate of inspections for each period  $t$ .
2. The analysis includes four measures and three objectives  $X_j$ : (i) initiated sanction procedures; (ii) total number of settlements of tax due after inspection (termed *actas tributarias* in Spanish), composed of two subgroups customs tax non-compliance cases (b1) or additional tax due (b2); and (iii) number of cases of clearance suspensions after the discovery of forbidden goods, mostly on security or drugs-related grounds. The records are listed as percentages on the right of the table, first with respect to the number of inspections and second with regard to the total number of declared *sad.items*.
3. The overall precision of the system is around 4%. Data by years are shown in Table 4 (see row % *nc* vs. inspections).

The efficacy ( $\bar{e}$ ) for 2018 and 2019 is calculated with respect to 2017 (the baseline).

**Table 4.** Spanish customs office results in the period 2017–2019 (own elaboration based on Spanish Tax Agency data).

	#				% Checked				% Declared		
	2017	2018	2019	2020	2017	2018	2019	2020	2017	2018	2019
Sanction procedure	3,713	3,613	3762	2448	0.67	0.64	0.64	0.45	0.05	0.05	0.05
LRD	11,495	16,644	20,872	17,624	2.08	2.95	3.54	3.21	0.15	0.21	0.25
LC	838	842	545	609	0.15	0.15	0.09	0.11	0.01	0.01	0.01
Suspended clearance	1283	1962	1826	1459	0.23	0.35	0.31	0.27	0.02	0.03	0.02
Customs tax non-compl.	17,329	23,061	27,005	19,292							
SAD.items	7,864,768	7,865,024	8,270,449	8,229,925							
Inspections	553,300	565,248	589,373	548,312							
% of control	7.04	7.19	7.13	6.66							
% nc vs insp.	3.13	4.08	4.58	3.52							
e(t/2017)		1.33	1.56	1.12							

The items under the % inspected heading in Table 3 are considered the records (R) that must be outperformed in the next period. For example,

$$R_{2018} = p_{2018} = \{4.08; 0.639; (2.945 + 0.15); 0.347\}.$$

Total precision (% of non-compliance vs % inspected in 2019 was 4.58. We have identified three objectives: sanctions, tax enforcement (summing the cases of sub-objectives b1 and b2), and suspended clearance. This is the 2018 record that should be outperformed in 2019.

We apply the proposed methodology subject to the constraints below:

- Every document in the red channel must be controlled.
- The maximum control percentage is 8% of *sad.items* for each alternative, that is, 8% for the office as a whole.
- Maximum elapsed time, *t*, one hour for a customs office open to taxpayers for eight hours.

*Step 2. Predictor selection and calibration*

We explained that the conventional approach is to evaluate risks in terms of the *HS10 \* country* cross-feature. In other words, rule-based systems aim to control combinations such as *iphone \* China* based on evidence that many taxpayers undervalue their declaration of these goods purchased over a platform. We must optimize the inspections because *HS10 \* country* is useful for establishing rules, but it is not a good predictor of fraud.

Figure 5 shows the observed distribution of the precision. In 2019, 332,057 shipments arrived from China. Point A represents the cross-feature (99.90.00.01.00 \* *China*), which was controlled 4,997 times with 826 non-compliance cases ( $\bar{p} = 16.52\%$ ), and point B, 64.02.99.98.00 \* *China* with a  $\bar{p}$  of 0.643%. Customs offices inspect, for diverse reasons, goods with high and low non-compliance risks (i.e., looking for weapons of terrorists).

Figure 5 (left) shows that the estimated  $\bar{p}$  is greater than 30 % for only 282 cross-features. Figure 5 (right) crosses precision with effort (# of inspections). There is a region of low precision inspections and a high number of controls that could be improved. According to the rule-based system, some goods must be inspected on security grounds, even though there is a low probability of non-compliance, but the decision should be optimized within the bounded feasible region open to the decision maker.

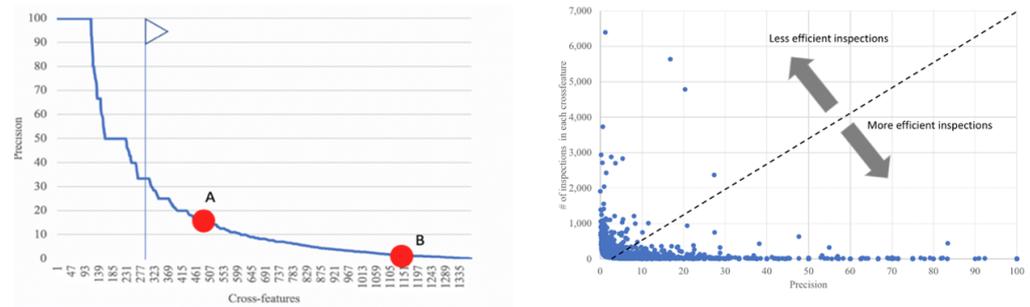


Figure 5. Precision.

We observed several important facts:

1. The type of products imported each year change with globalization and show new cross features; see Table 5 column *Tot.* in *HS10 \* Country*.

Table 5. Control results for different cross-features.

	Country			HS10			Declarants			HS10*Country			100%
	Tot.	Ch.	Res.	Tot.	Ch.	Res.	Tot.	Ch.	Res.	Tot.	Ch.	Res.	
2017	234	208	98	11,639	8447	2231	9617	9617	486	128,385	44,155	4216	768
2018	233	233	114	11,766	8419	2647	7002	7002	492	129,315	43,880	5335	1123
2019	239	239	113	1173	8451	2624	9223	4766	517	131,854	44,111	5320	576

2. While the number of cross-features with a precision of 100% is high, Figure 6 left, it is very uncertain. For example, there were 1123 cross features in 2018 (with only one non-compliance case per inspection in many cases).
3. Customs offices use inspection rules that do not consider all cases of risk. Many high-risk cross-features are not inspected because the rule-based systems cannot provide the granularity to cross features with a cardinality in the range of 130 K.
4. The lessons learning process could be better. For example 38.22.00.00.00 is the code for laboratory reagents. When combined with country of origin U.S.A., it was inspected 8256 times with  $\bar{p} = 0.37\%$ . It is clear that, after verifying that a provider is regularly compliant, rules should be updated using a Bayesian strategy to avoid the reiteration of useless inspections. The rule-based systems does not have this flexibility.

There is room for improvement, but the solution to the optimization is not trivial for many reasons:

1. There is not enough evidence to optimize the decision. In Figure 6 (right), points are cross-features (e.g., 99.90.00.01.00\* China) for the goods submitted on a random day (2 December 2019). Some of them were compliant, and others, non-compliant. The ordinate shows the estimated Bayesian precision prior to their control. We find that there were cases of compliance and non-compliance for a single level of estimated precision, and also many non-compliance cases were found for high values of expected fraud (precision).

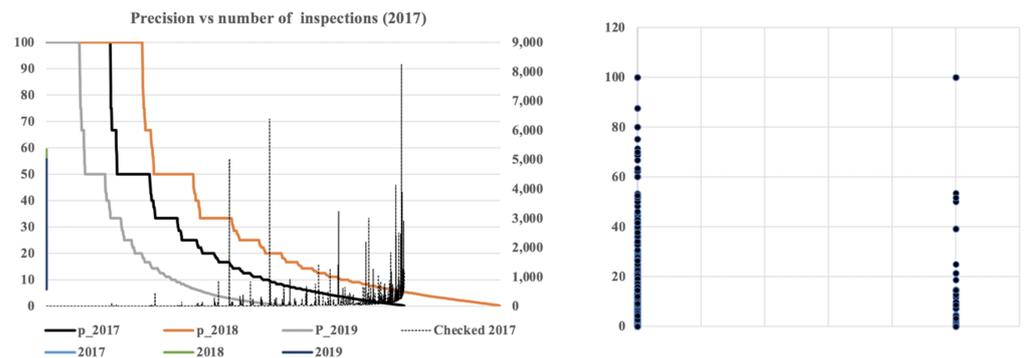


Figure 6. Evolution and distribution of precision.

2. A switch to low cardinality variables, such as country of origin, does not solve the problem because the highest probability of fraud is for countries of origin with low traffic (e.g., Australia with 77 non-compliance cases for 82 *sad.items*), and the country with the most detected non-compliance cases is China, with 7784 non-compliance cases in 332,057 *sad.items* ( $p = 2.34\%$ ), that is, a low probability.
3. There are conflicting objectives. Figure 7 (left) shows the top ten countries by flow of imports and their respective percentage with respect to each type of results (i.e., China accounts for 60% of all sanctions and is followed in order of importance by the U.S.A.).

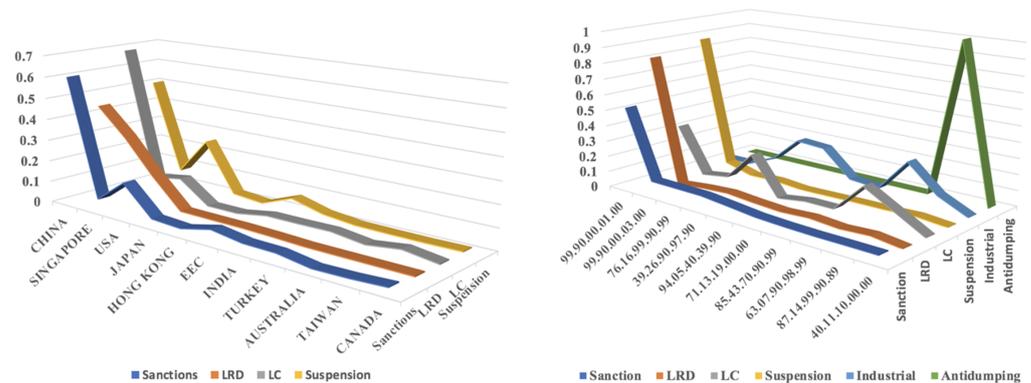


Figure 7. Top countries and codes.

4. When we include these and other possible attributes, such as the control of antidumping taxes, identified by their *HS10* code, we find, for example, that the control of position 87.14.99.90.89 is important for antidumping, although it is irrelevant for other objectives. It is complicated to optimize all objectives.

We resumed the facts and the obstacles, and we now describe the solution proposed.

In order to select a predictor of fraud, we search the optimal cross-feature using SDT. Figure 8 (left) represents precision (with the *HS10 \* country* cross-feature). We use the analogy of stimulus intensity with the possibility of distinguishing signal noise, to study compliance vs. non-compliance, as shown in Figure 1. Figure 8 (right) represents the ROC curve.

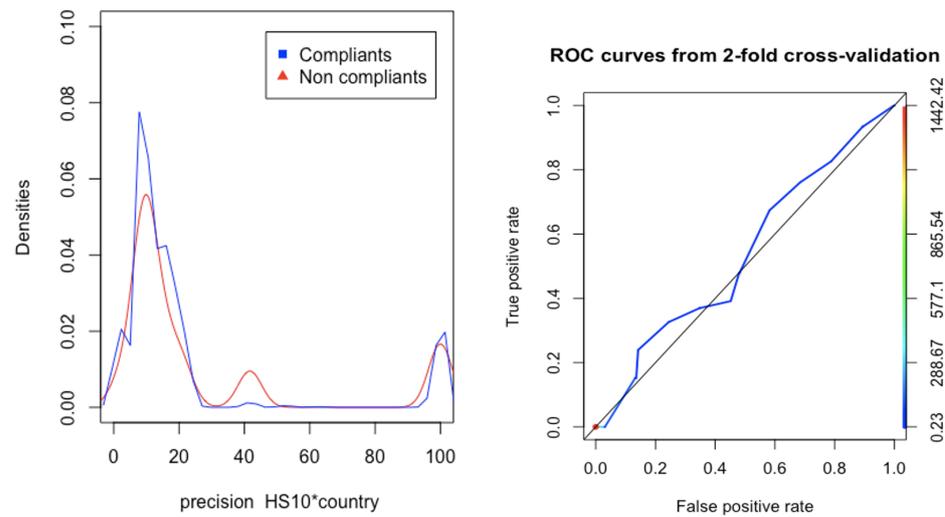


Figure 8. Classifier using  $HS10 * country$  as the cross-feature.

Figure 8 illustrates that, using this cross-feature, there is no clear distinction between the distribution of compliant, and non-compliant documents, the AUC in the ROC is low, and we get a result that is not much better than a random classifier.

The strategy adopted to improve the results was to increase granularity. The first step was to look for a better cross-feature combining four variables:  $HS10 * Country * Typifiers * RangeOfCustomsValue$ .

Figure 9 shows the results of several tests using weighted precision values, applying a linear transformation, followed by an exponential transformation (green curve) of precision. We see that the density of the non-compliant distribution has moved upward. Therefore, it is easier to identify compliance in the interval of low values of weighted precision. In a third step, we used a logistic regression to search for the optimal classification for *sad.items*, using a combination of variables as the predictor. Precision, importer status as private person (Yes/No), importer status as Authorized Economic Operator (Yes/No), importer legal status (natural person, PLC, corporation), weight and value of the goods. The final results are illustrated in Figure 9 (right), where the blue line provides a better classification of non-compliance cases at the top of the curve.

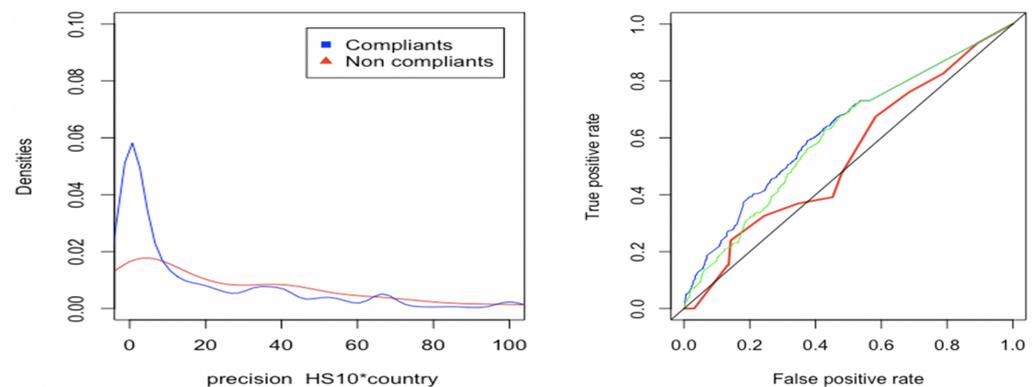


Figure 9. Using high-order cross-features.

With an answer to the problem of the selection of features we move to the problem of dynamic optimization.

*Calibration of MOBADO for Adolfo Suarez Airport*

We studied the characteristics of the selected customs office, Adolfo Suarez Airport, in Madrid, see Table 6. We selected one type of traffic because the type and rate of goods

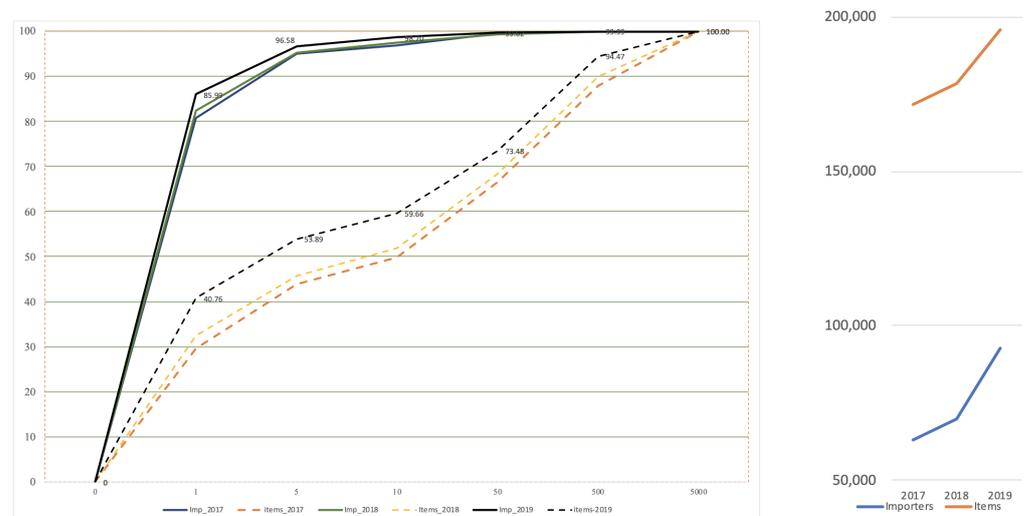
cleared differs from one office to another, and a joint analysis of the mixture would be less precise. The system has to be adjusted for each problem because non-compliance is local.

**Table 6.** Example of data submitted to Adolfo Suarez Airport.

HS10	Country	SAD.items	#	Sanc.	LRD	LC	S	NC	SANC.	LRD	LC	S
99.09.00.00.10	Singapore	15,187	2	0	0	0	2	0.01	0.00	0.00	0.00	0.00
99.09.00.00.10	USA	10,769	1	0	0	1	0	0.01	0.00	0.00	0.01	0.00
99.09.00.00.10	China	10,189	1	0	1	0	0	0.01	0.00	0.01	0.00	0.00
03.02.89.90.90	Morocco	8199	2	1	1	0	0	0.02	0.01	0.01	0.00	0.00
08.10.20.01.00	Marocco	6387	1	0	1	0	0	0.02	0.00	0.02	0.00	0.00
03.02.89.90.90	Mauritania	4953	1	0	1	0	0	0.02	0.00	0.02	0.00	0.16
03.02.85.90.00	Morocco	4755	1	0	0	0	1	0.02	0.00	0.00	0.00	0.00

A total of 1,812,269 *sad.items* were submitted at this office, with a mean of 4952 per day. Of the 55,874 different cross-features presented to inspectors at this customs office, 1432 were found to be fraudulent. This accounts for  $1432/5320 = 27\%$  of the total national non-complaint cross-features, see Table 5 (last ine).

Figure 10 (left) represents the number of *sad.items* by importer on the abscissa. Of the total traffic, citizens that use this customs service only once a year, for example, to buy a book or a phone, account for 85%, that is, most are low-value imports by natural persons. The number of non-compliance cases by type is listed in the columns. The evolution of the number of items and importers is shown on the right.



**Figure 10.** Importers at Adolfo Suarez Airport.

Figure 11 illustrates the data for this airport in 2020. This illustrates the reduction of traffic due to the COVID-19 pandemic, and the partial recovery at the end of the year.

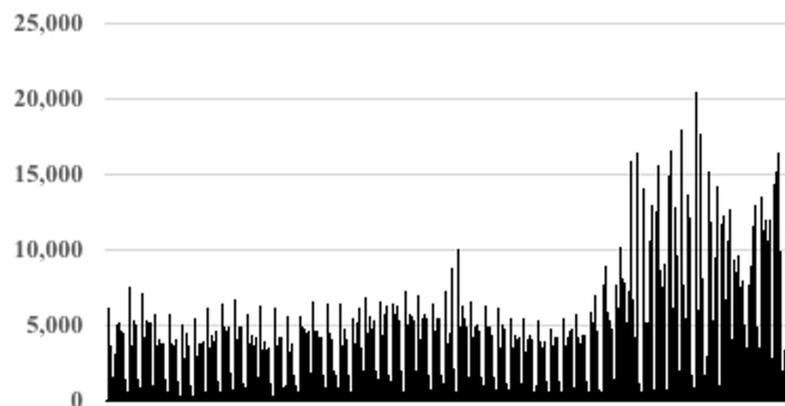


Figure 11. Number of *sad.items* over the year 2020 at Adolfo Suarez Airport.

The calibration has three components:

1. Decision about the *feature* used as a predictor.  
 Figure 12 (right) shows the distribution of the  $HS10 * country$  cross-feature precision with the histogram and a multimodal curve that fits the density distribution in 2017 (blue). The fitted distribution changes over time, red (2018) and green (2019), showing that the use of an evidence-based Bayesian approach improves the quality of the evidence over time.

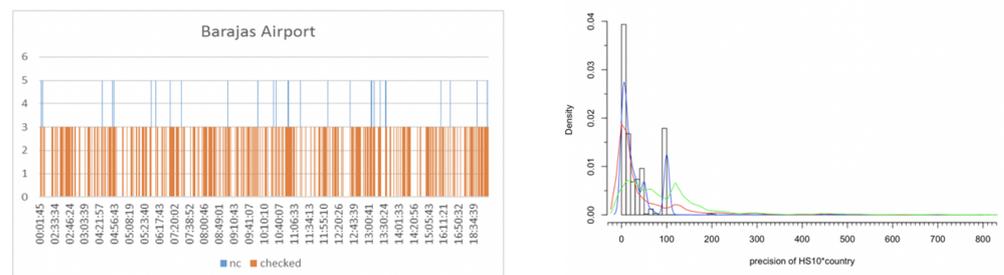


Figure 12. Rate of clearance and characteristics of precision.

We know that the mean value of the number of *sad.items* submitted at the airport is 4952. The interval within which declarants can submit documents is from 12:00 a.m. to 6:35 p.m. Therefore, the rate of flow is 43 documents per minute.

2. Setting of constraints.  
 The number of controls must not account for more than 8% of documents.  
 All documents in the red channel must be inspected.
3. Decision on the optimal stopping rule 1 hour.
4. Decision about the threshold.

In our case, every document must be channeled during the day, so we must study the problem in a day by day basis. We show one working day of each month of 2019 at Adolfo Suarez Airport in Madrid. We report the results for the current system. Tables 7 and 8 show the control date (1), the number of inspections established by the current system, red channel (2) and channels other than green (3), the number of non-compliances detected via the red (4) and other channels (5).

The current rule-based system uses the variable  $p$  of the  $HS10 * country$  cross-feature, although we proved that it is not a very good predictor. We obtain the number of estimated results that would have been obtained using the estimated Bayesian probability when inspecting only the *sad.items* with a  $p$  higher than each of the thresholds considered. We made an assay with several discrimination thresholds and we present in Table 8 several of them.

For example, on 2 January 2019, 239 *sad.items* surpassed the threshold of  $p = 250$ , and we can estimate that their control would have identified 28 non-compliance cases. In this case, the customs office would have been identified 3 non-compliance cases in red channel controls and 28 in other inspections, totaling 31 non-compliance cases with the stated precision and efficacy.

Table 7. Analysis with one day per month sample for 2019

Checks		Results			Low Levels					
Day	Imposed	Other	Red	Others	Level 10		Level 50		Level 100	
2-January	3	374	3	6	2554	12.85	1750	15.34	1201	17.43
1-February	26	462	4	30	2208	12.64	1518	15.25	923	18.13
2-March	25	458	1	26	942	13.60	673	16.24	487	18.37
1-April	23	404	1	24	3264	12.90	2315	15.31	1645	17.20
2-May	20	206	1	21	1972	12.85	1359	15.34	963	17.30
3-June	32	355	4	36	3209	12.94	2286	15.41	1631	17.31
2-July	41	309	4	45	1897	12.75	1235	15.91	866	18.13
2-August	20	318	2	22	2303	12.55	1574	15.16	991	17.56
2-September	27	443	6	33	3061	12.40	2147	14.71	1353	17.05
2-October	37	371	3	40	2226	12.77	1520	15.39	975	17.97
1-November	8	129	0	8	1270	13.28	928	15.61	694	17.28
2-December	35	568	7	42	4015	12.96	2902	15.20	2056	17.09
Total	297	4397	36	333	28,921	12.83		15.32		17.46

Table 8. Analysis with one day per month sample for 2019 (Continue)

Checks		Level 200			Level 250			
Day	Check	Est.	$p$	$e$	Check	Est.	$p$	$e$
2-January	328	25	8.48	3.09	239	28	12.91	3.43
1-February	277	27	19.04	1.55	217	29	25.55	0.98
2-March	127	29	42.75	2.01	101	33	57.16	1.25
1-April	437	25	10.93	1.91	290	29	17.85	1.19
2-May	277	25	16.08	2.02	216	27	21.61	1.26
3-June	392	25	14.63	1.43	271	29	22.60	0.83
2-July	272	25	24.35	1.35	195	29	35.71	0.67
2-August	259	25	17.48	1.89	200	28	23.94	1.24
2-September	337	24	15.25	1.32	216	29	25.82	0.89
2-October	322	25	19.21	1.44	224	28	29.23	0.73
1-November	158	26	21.46	4.24	116	29	32.06	3.65
2-December	459	25	13.16	1.23	330	29	19.35	0.73
Total	3645	25.3	18.57	1.96	2615		26.98	1.40

We find that the number of inspected *sad-items* decreases (i.e., from 2554 to 239) as the threshold increases (50, 100, 200,...) in the data for 2 January 2019, and the mean estimated number of non-compliance cases increases (from 12.85, (level 10), to 28 (level 250)). The estimated number is greater because we are selecting only *sad.items* with a higher probability of fraud.

Figure 13 summarizes the efficacy (auxiliary ordinate axis) and precision (principal ordinate axis) for each of these twelve days presented in the table—samples of the months of the year. With a threshold of  $eB_p = 200$ , the customs office would have inspected 3645 *sad.items*, that is, 82.89% of the inspections (4397) that were performed by the customs office on these days), increasing the precision fourfold from 4.58 (see Table 1) to 18.57 (mean value for the specified dates). Looking back at Figure 1, a different threshold changes the ratio of the inspection to seizure rate. With a threshold of 250, the customs office would have inspected 2615 documents with a precision of 26.98.

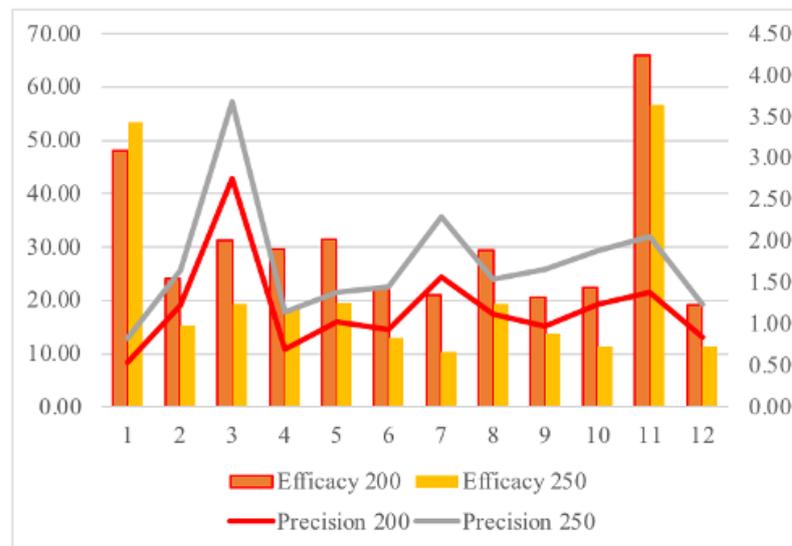


Figure 13. Adolfo Suarez Airport Customs Office clearances (2020).

4.2. Validation of MOBADO with Evidence from 2020

We use 2020 data to test the external validity of the solution using MOBADO trained with 2019 data. We used the same days, that is, a sample with 53,336 elements for the test. Customs offices inspected 4433 of the elements, 228 of which were assigned to the red channel with mandatory control, and we detected 36 non-compliance cases.

We found 201 non-compliance cases for a MOBADO selection of 2605 inspections. Therefore, total precision was as follows:

$$p = \frac{36 + 201}{228 + 2605} = 0.0837(8.37\%).$$

The solution satisfies all the constraints, and records from previous years.

$$R_{2018} = \{4.08; 0.639; (2.945 + 0.149); 0.347\}$$

$$R_{2019} = \{4.58; 0.638; (3.541 + 0.092); 0.310\}$$

$$R_{2020} = \{3.52; 0.446; (3.214 + 0.111); 0.266\}$$

$$MOBADO - R_{2020(BARAJAS)} = \{8.37; 1.092; (7.108 + 0.092); 0.5189\}$$

The 2020 optimization more than doubled (with a 237% increase) the precision of the customs 2020 record. The number of controls non ordered by the rule-based systems decreases from 4397 to 2605 and that frees up 40.52% of the resources dedicated to control.

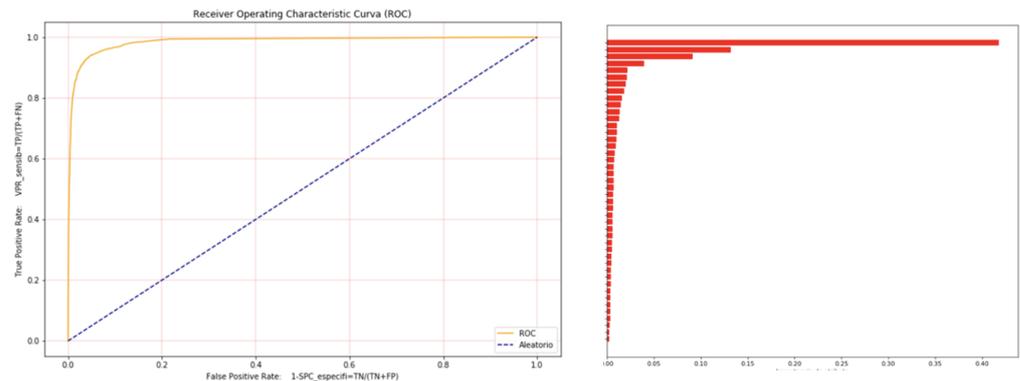
The results could be optimized even further if the customs office were to adopt a new inspection strategy, explaining a new method that combines stopping time and lines of desire to increase precision. The result is a precision of 23.23% with an increase of 1.8 times in expected efficacy.

These results were achieved using a stopping time (one hour for orange and red SADs) instead of continuous decision making. However, the combination of both papers provides the framework for decision learning, considering costs and benefits.

5. Discussion

There remain many open questions. These are good results, but could they be even better? What determines the quality of the results? To answer these questions, we need to return to MOBADO stage and discuss two issues.

1. What is the best possible result with only one objective (e.g., tax inspection)? Figure 14 shows the results for the Adolfo Suarez Airport in Barajas with all the data for 2020, using the gradient boosting classifier and all the useful features of a SAD (43 items). The resulting ROC-AUC was 0.999965 for the training set (70% with a sample size of 81.655) and 0.98755 for the test set. The right side depicts the relative importance of features (box 4: type of previous document or summary declaration; box 8: legal personality (natural or legal person); box 29: customs office; box 1: country, etc., ranked in order of importance).



**Figure 14.** Classification with gradient boosting (Barajas 2020).

At the optimal point, we output 2083 true positives and 31,152 true non-compliance cases, 1630 false positives and 130 false negatives with a precision of 56.10%. These are better results than those reported in Section 4 and are consistent with or better than the best results reported in the literature [17]. The crux of the matter, however, is that this approach, despite raising precision from 3.57 to 56.10 (an incredible 1571%), does not provide a real solution to the problem.

In customs control, the observed fraud (see Table 3) with respect to inspected documents is in the range of 3%. Therefore, the precision of a classifier predicting that every SAD is compliant would be 97%. We did not apply gradient boosting to outperform other ML tools, nor did we aim to select only the “most promising SADs” because customs offices have to investigate terrorism or drug smuggling, even if the risk is low. We wanted to optimize control under real conditions in a multi-objective context with real constraints.

Therefore, the answer to the question is as follows:

- (a) The results could be improved by forsaking “low precision controls”, such as those to drug smuggling- and terrorism-related inspections.
  - (b) It only makes sense to compare results for controls by different customs administrations that have the same mix of objectives and a similar fraud baseline.
2. What is the best result achievable by any country in a multi-objective context using a less detailed document, while meeting commerce facilitation constraints?

We must take into account that the control of low-risk, albeit dangerous, shipments is high in terms of precision.

Figure 15 shows the cost of misclassification in the control of China’s e-commerce. As the inspection rate grows (health- and terrorism-related checks) or due to extensive customs checks, the seizure rate decreases. There is a tradeoff between inspection extension and precision.



**Figure 15.** Cost of misclassification. (Source [36]).

How is it possible to manage all these elements? This is the question that MOBADO solves, as it does the following:

1. Increases the precision by 237%.
2. Outperforms the current system in each and every regard (tax collection, drugs, etc.).
3. Is applicable with just the basic features used in every country.
4. Can be combined with ML approaches and bridges with current rule-based systems.
5. Provides an iterative method to increase inspection quality because performance is better year on year, which frees up resources that can be used to combine the “exploitation” of evidence and the “exploration” of new forms of fraud.

Finally, what determines the quality of practical results?

1. The quality of previous work of inspection done by customs offices because the training set used in the analysis contains the data previously selected and inspected by customs offices.
2. The quality of the practical results depends on the possibility of building a bridge between MOBADO results and the real-time rule-based transactional system performing risk analysis.

## 6. Conclusions

The use of machine learning techniques in the field of customs control from its origins [25] to present day [17] has had to overcome two type of difficulties: the adaptation of legacy rule-based systems and the need to fulfill the mission of customs, which must meet multiple objectives simultaneously, some of which are quantifiable and respond to measurable risks and others not.

This paper reports an original approach, setting out a hybrid method that is compatible with any artificial intelligence approach to optimize the inspection task at a customs office within a multi-attribute and multi-objective setting.

**Author Contributions:** Conceptualization, I.G.G. and A.M.C.; methodology, I.G.G. and A.M.C.; software, I.G.G.; validation, I.G.G. and A.M.C.; formal analysis, I.G.G. and A.M.C.; investigation, I.G.G.; resources, I.G.G.; data curation, I.G.G.; writing—original draft preparation, I.G.G.; writing—review and editing, I.G.G. and A.M.C.; visualization, I.G.G.; supervision, A.M.C.; project administration,

A.M.C.; funding acquisition, A.M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Ministry of Economy and Competitiveness, project MTM2017-86875-C3-3-R.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. European Commission. The European Union Explained: Customs. Available online: <https://op.europa.eu/en/publication-detail/-/publication/98af8cf1-93fa-11ea-aac4-01aa75ed71a1/language-en> (accessed on 22 April 2021).
2. Vinogradova, I. Multi-Attribute Decision Making Methods as a Part of Mathematical Optimization. *Mathematics* **2019**, *7*, 915. [CrossRef]
3. Mardani, A.; Jusoh, A.; Nor, K.; Khalifah, Z.; Zakwan, N.; Valipour, A. Multiple criteria decision-making techniques and their applications—A review of the literature from 2000 to 2014. *Econ. Res.* **2015**, *28*, 516–571. [CrossRef]
4. Belton, V.; Stewart, T.J. *Multiple Criteria Decision Analysis: An Integrated Approach*; Kluwer: Alphen van den Rijn, The Netherlands, 2002.
5. Hwang, C.L.; Yoon, K. *Multiple Attribute Decision Making: Methods and Applications a State-of-the-Art Survey*; Springer: Berlin/Heidelberg, Germany, 1981.
6. Baesens, B.; Vlasselaer, V.; Verbeke, W. *Fraud Analytics Using Descriptive Predictive and Social-Network Techniques*; Wiley: Hoboken, NJ, USA, 2015.
7. Vanhoeyveld, J.; Martens, D.; Peeters, B. Customs fraud detection: Assessing the value of behavioral and high-cardinality data under the imbalanced learning issue. *Pattern Anal. Appl.* **2020**, *23*, 1457–1477. [CrossRef]
8. Canrakerta, A.; Hidayanto, A.N.; Ruldeviyani, Y. Application of Business Intelligence for Customs Declaration: A Case Study in Indonesia. *J. Phys. Conf. Ser.* **2020**, *1444*, 012028. [CrossRef]
9. Desiderio, D. Data Analysis Techniques for Enhancing the Performance of Customs, Big Data and Government: How the Public Sector Leverages Data Insights. Hortonworks. 2020. Available online: <https://hortonworks.com/article/big-data-and-government-how-the-public-sector-leverages-data-insights/> (accessed on 22 April 2021).
10. Filho, J.J.; Wainer, J. Using a hierarchical Bayesian model to handle high cardinality attributes with relevant interactions in a classification problem. In Proceedings of the 20th International Joint Conference on Artificial Intelligence IJCAI, Hyderabad, India, 6–12 January 2007; pp. 2504–2509.
11. Biljan, J.; Trajkov, A. Risk management and Customs performance improvements: The case of the Republic of Macedonia. *Procedia Soc. Behav. Sci.* **2012**, *44*, 301–313. [CrossRef]
12. Geourjon, A.M.; Laporte, B.; Coundoul O.; Gadiaga M. Inspecting Less to Inspect Better: The Use of Data Mining for Risk Management by Customs Administrations. Working Papers. Fondation Pour les Études et Recherches sur le Développement International. 2012. Available online: [http://www.ferdi.fr/sites/www.ferdi.fr/files/publication/fichiers/P46-eng\\_WEB\\_0.pdf](http://www.ferdi.fr/sites/www.ferdi.fr/files/publication/fichiers/P46-eng_WEB_0.pdf) (accessed on 22 April 2021).
13. Chermiti, B. Establishing risk and targeting profiles using data mining: Decision trees. *World Customs J.* **2019**, *13*, 39–58.
14. Kavoya, J. Machine learning for intelligence-driven Customs management. *Afr. Tax Cust. Rev.* **2020**, *3*, 50–58.
15. Laporte, B. Risk management systems: Using data mining in developing countries customs administrations. *World Customs J.* **2011**, *5*, 17–27.
16. Regmi, R.H.; Timalisina, A.K. Risk Management in customs using Deep Neural Network. In Proceedings of the IEEE International Conference on Computing, Communication and Security, Kathmandu, Nepal, 25–27 October 2018; pp. 33–137.
17. Kim, S.; Tsai, Y.C.; Singh, K.; Choi, Y.; Ibok, E.; Li, C.T.; Cha, M. DATE: Dual Attentive Tree-aware Embedding for Customs Fraud Detection. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Diego, CA, USA, 23–27 August 2020.
18. New Zealand Customs Service. Available online: <https://www.customs.govt.nz/about-us/about-customs/our-history/> (accessed on 22 April 2021).
19. New Zealand Customs Service. Privacy by Design. Available online: <https://www.customs.govt.nz/about-us/your-information/your-personal-information/privacy-by-design/> (accessed on 22 April 2021).
20. Triepels, R.; Feeldeers, A.; Daniels, H. *Uncovering Document Fraud in Maritime Freight Transport Based on Probabilistic Classification*; Computer information systems and industrial management, Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9339, pp. 282–293.
21. Triepels, R.; Feeldeers, A.; Daniels, H. Data-driven fraud detection in international shipping. *Expert Syst. Appl.* **2018**, *99*, 193–202. [CrossRef]

22. Tsois, A.; Lema, C.; Makridis, M.; Checchi, E. Using container status messages to improve targeting of high-risk cargo containers. In Proceedings of the Research Track at the 5th World Customs Organization Technology and Innovation Forum, Rotterdam, The Netherlands, 27 October 2015.
23. Slowey, C. Data Analysis for Effective Border Management. The Canadian Experience. *WCO News* **2017**, *82*. Available online: <https://mag.wcoomd.org/magazine/wco-news-82/data-analysis-for-effective-border-management-the-canadian-experience/> (accessed on 22 April 2021).
24. Rigney, P. The All-Seeing Eye—An HMRC Success Story? Available online: <https://www.ifa.org.uk/media/653935/Tax-HMRC-Connect-system.pdf> (accessed on 22 April 2021).
25. Jambeiro, J.; Wainer, J. A model for handling BN nodes with high cardinality parents. *J. Mach. Learn. Res.* **2008**, *9*, 2141–2170.
26. Jambeiro, J. Artificial intelligence in the customs selection system through machine learning (SISAM). *Receita Fed. Brasil* **2015**, *1*, 1–49.
27. Zhou, X. Data mining in customs risk detection with cost-sensitive classification. *World Cust. J.* **2019**, *13*, 115–130.
28. Zhu, F.; Ma, Z.; Li, X.; Chen, G.; Chien, J.T.; Xue, J.H.; Guo, J. Image-text dual neural network with decision strategy for small-sample image classification. *Neurocomputing* **2019**, *328*, 182–188. [[CrossRef](#)]
29. Castilla, M. Mejores Practicas y Futuro de las Aduanas. Interamerican Development Bank. 1998. Available online: <https://publications.iadb.org/publications/spanish/document/Mejores-prácticas-y-futuro-de-las-aduanas.pdf> (accessed on 22 April 2021).
30. Appeals, T.; Swimland, H. Rolling back the frontiers: The customs clearance revolution. *Int. J. Logist. Manag.* **1998**, *9*, 111–118. [[CrossRef](#)]
31. Coelho, F. *From “Red Tape” to “Red Carpet”: How the Brazilian AEO Programm Has Brought Customs Procedures to an Entire New Level*; WCO News: Brussels, Belgium, 2019; Volume 88. Available online: [https://mag.wcoomd.org/magazine/wco-news-88/brazil\\_ao/](https://mag.wcoomd.org/magazine/wco-news-88/brazil_ao/) (accessed on 23 April 2021).
32. Clark, S.E.; Benjamin, A.S.; Wixted, J.T.; Mickes, L.; Gronlund, S.D. Eyewitness Identification and the Accuracy of the Criminal Justice System. *Policy Insights Behav. Brain Sci.* **2015**, *2*, 175–186. [[CrossRef](#)]
33. Stanislaw, H.; Todorov, N. Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* **1999**, *31*, 137–149. [[CrossRef](#)] [[PubMed](#)]
34. Green, D.M.; Swets J.A. *Signal Detection Theory and Psychophysics*; Wiley: New York, NY, USA, 1966.
35. Swets, J.A. *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1996.
36. Quing, H.; Zheng, G.; Fu, D. Risk Data Analysis of Cross Border E-commerce Transactions Based on Data Mining. *J. Phys. Conf. Ser.* **2011**, *1744*, 032014. [[CrossRef](#)]
37. Ferguson, T.S. Who Solved the Secretary Problem? *Stat. Sci.* **1989**, *4*, 282–289. [[CrossRef](#)]