Customs Valuation Assessment Using Cluster-based Approach

Nur Siyam (✉ nur.siyam@gmail.com)

The British University in Dubai <https://orcid.org/0000-0003-1625-4892>

Omar Alqaryouti

The British University in Dubai <https://orcid.org/0000-0002-8538-8344>

Khaled Shaalan

The British University in Dubai <https://orcid.org/0000-0003-0823-8390>

Research Article

Keywords: Customs Valuation, Clustering, Local Outlier Factor, Value Manipulation, Distance Measures, Density Measures

Posted Date: January 24th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1288941/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Customs Valuation Assessment Using Cluster-based Approach

Omar Alqaryouti¹ [0000-0002-8538-8344], Nur Siyam^{1,*} [0000-0003-1625-4892], and Khaled Shaalan¹ [0000-0003-0823-8390]

¹ The British University in Dubai, Dubai United Arab Emirates

Faculty of Engineering and Information Technology

* nur.siyam@gmail.com

Abstract. Customs duty obligations are primary revenue sources for nations. Any under or over value can result in economic instability. The role of customs is safeguarding society and local businesses from criminally guided value manipulations. In this paper, a cluster-based approach to detect the customs commodity value manipulation is proposed. This approach represents the shipment related information into 3-dimensional space to determine any wrongdoing through two stages comprising distance- and density-based techniques. This representation simplifies the process by identifying relationships among the various shipment features. These two stages work together in order to analyse the shipments information and determine any abnormal behaviour for these shipments. The results of the proposed approach achieved an accuracy of 86%. This technique will provide much needed capabilities to protect the revenue and securing the trade supply chain from illegitimate activities.

Keywords: Customs Valuation; Clustering; Local Outlier Factor; Value Manipulation; Distance Measures; Density Measures.

1. Introduction

The role of customs Administrations is vital in protecting the country borders from hazardous people and materials. Customs Administrations encounter challenges in keeping balance to sustain the economic growth through facilitating the movement of cargo and passengers while applying controls to detect customs fraud and other offences and therefore protect the society. One of the essential controls at the customs administrations is the Customs Valuation.

The Customs Valuation function is responsible for validating and determining the correct declared values in the customs declaration. This includes the process of estimating the values of the declared imported goods. This aims at protecting the revenue and establishing effective collection mechanisms and identifying any fraud activities. Additionally, value manipulation highlights the potential of smuggling goods. Therefore, determining whether the declared value for the goods is correct has an important impact on protecting the local society and economy.

The customs valuation process is triggered whenever there is a doubt on the declared value of goods. This can be identified through the established mechanisms based on the pricelist that the customs administrations maintain along with the tariff and the origin information. Upon the trigger of the valuation process, the valuation officers verify the declared value using any of the following six methods; namely; Transaction Value, Transaction Value of Identical Goods, Transaction Value of Similar Goods, Deductive Method of Valuation, Computed Method of Valuation, and Derivative Method of Valuation. The transaction value method is the most commonly used method among other methods. The transaction value method takes into consideration the showing cost provided in the invoice in addition to any proof of sale to verify the authenticity of the provided cost (Rosenow & O'Shea, 2010).

The customs valuation process also handles the follow-ups related to the customs valuation developments regarding the implementation of the World Trade Organization (WTO) Valuation Agreement. This agreement aims for establishing a comprehensive system for determining and validating the customs value in a fair and neutral manner (WCO, 2019).

The process of validating the goods value is normally performed manually and is dependent on the valuation officer experience. It also requires tremendous human efforts to determine the correct value. The result of this process is either there are issues found with respect to the declared value or there are no issues found. The issues are usually related to fraud activities, human errors while submitting the goods value, legitimate trade agreements, or seasonal trades that cause variations in the values. The importance of this process in protecting the local society underlines the demand of establishing accurate mechanisms that can be used to verify the declared goods value.

In this paper, a cluster-based approach that aims to identify whether the declared goods value raises any suspicious behaviour is proposed. The proposed approach consists of two stages. In the first stage, the investigation takes into account all declared shipments that share the same goods HS-Code. In this stage, the shipments are represented as points in multidimensional clusters. Each cluster consists of customs declarations that share the same HS-Code. Figure 1 illustrates the representation in multidimensional space. In this representation, the price for the new shipment can be analysed by using statistical measures that capture the locations of the new shipment inside the cluster. This stage labels the shipment as either normal or outlier. When the new shipment is labelled as normal, the status of this shipment will be confirmed in terms of values. The outlier shipments will be sent to the second stage for further analysis. In the second stage, an outlier detection mechanism is employed for the clustered points to determine whether the new shipment can be considered as an outlier or not. In this case, the results of the second stage override the results of the first stage. This mechanism is adopted to identify whether a new shipment can be considered as an outlier. It is required to have an agreement between both stages. Both of these stages take into consideration the distribution of the points to determine whether a shipment should be considered as an outlier or not. However, the first stage validates based on distance measures while the second stage validates based on density.

Figure 1: Multi-dimensional space representation example

This paper comprises six sections, including the introductory section. The second section explores the related work and identify the research gap. The third section is concerned with the methodology used for this research. The fourth section illustrates the proposed algorithm and the different stages that are adopted. The fifth section discusses the various experimental settings and presents the research findings. Finally, the sixth section provides brief summary of the proposed work and the research findings. In addition, it states the areas for future works.

2. Related Work

In several application domains, the use of outlier detection mechanism to detect the abnormal activities have been studied extensively. For instance, in credit cards fraud domain, a k-nearest neighbour (K-NN) mechanism has been proposed to ensure the legitimacy of the credit cards transactions and detect anomalies (Ceronmani Sharmila et al., 2019; Malini & Pushpa, 2017).

To address the problem of detecting the money laundering related activities, a distance-based approach was developed to capture the likelihood of money laundering behaviours (Gao, 2009). The proposed approach groups the transactions based on the distances between the points within the group. Then, the points are analysed based on pre-defined threshold to determine legitimacy of these transactions. In the same vein, a study that addresses the financial behaviours were conducted (Chen et al., 2007). In this study, the authors focused on Taiwanese companies and investigated the benefits of employing outlier detection approach such as Local Outlier Factor (LOF) and K-NN.

In the cars insurance domain, a K-NN approach has been also adopted to detect fraud activities (Badriyah et al., 2018). The proposed approach represents the transactions in multi-dimensional space, where the analysis is performed based on the actual distances between the points. The new transaction will be investigated with its k-

nearest neighbours to confirm the status of the transaction. The value of k is determined by performing a sensitivity analysis.

In the computer network domain, a mechanism to capture the abnormality in network traffics was proposed (Gan & Zhou, 2018). This mechanism represents the network traffic as points in multi-dimensional space. Then, the points are analysed using clustering techniques to determine the similarity between these points. In the transportation domain, several outlier detection approaches were proposed to optimize the bus root planning (Almiani et al., 2018, 2014). In these approaches, the authors established a detection mechanism to determine the bus stops that are considered as an outlier. Removing the outliers will results in optimizing the overall bus root path.

To optimize the distance-based outlier detection converging time, a multi-core clustering algorithm was proposed (Bhaduri et al., 2011). This algorithm uses pruning method to determine the outlier factor value for a given point based on pre-defined threshold value. In this algorithm, any point that reaches the pre-defined threshold will not be considered. A hierarchal sorting technique was employed to update the threshold value. Additionally, the proposed algorithm adopted the concept of leadership points that is used to monitor the transactions set and the process of determining the likelihood at each point being an outlier.

In a more related study, Vanhoeyveld et al. (2020) proposed a customs fraud detection application that supports vector machine base learner in a confidence rated boosting algorithm. The study analysed a dataset of 9,624,124 records obtained from the Belgian customs administration. The results revealed that high AUC (Area under the ROC Curve) and lift-values have a strong impact on customs fraud detection. The study also emphasised the importance of fine-grained customs features, such as consignee, consignor, and declarant.

In this line, this study proposes multi-level outlier detection approach to determine any value manipulation from customs perspective. The main idea of the proposed approach is to combine the LOF algorithm with the distance-based approach. This combination is selected to capture the differences between the transactions in terms of density and distance.

3. Methodology

The problem of ensuring the accuracy of the declared goods is a challenging problem. This paper aims to study how clustering techniques can be used to detect abnormal declared goods values. This section illustrates the proposed cluster-based approach that is performed on two stages of analysis namely; distance-based and density-based analysis. These stages aim to determine whether any manipulation has occurred in terms of the goods declared value.

There are several methodologies that are used to perform data mining tasks such as Knowledge Discovery in Databases (KDD) (Fayyad et al., 1996; Siyam et al., 2020), SEMMA (Azevedo & Santos, 2008) and the Cross Industry Standard Process for Data Mining (CRISP-DM) (Pete et al., 2000). These methodologies have common objectives to uncover hidden knowledge and gain insights.

In this work, the CRISP-DM is adopted. CRISP-DM is neutral in terms of industry, technology and application. It is one of the most popular data mining methodologies that is considered as a data mining *de facto* standard for developing knowledge discovery projects (Mariscal et al., 2010).

As illustrated in Figure 2, The CRISP-DM methodology comprises six stages. The first stage entails obtaining clear understanding of business requirements, challenges and opportunities. The second stage deals with studying and comprehending the collected data. The third stage goes into preparing the data in order to provide a cleansed and harmonized dataset. The fourth stage takes the prepared data and starts the modelling exercise represented in the proposed cluster-based approach. The outcome of the modelling exercise is evaluated in the fifth stage. Finally, in the sixth stage, the model is packaged for deployment to production to handle real-time data.

The modelling stage is concerned about building the Data Mining model. In this study, clustering approaches were adopted. The focus in this study is to address real-life scenarios related to customs post clearance audit and value manipulation. The obtained dataset is divided into 90% safe customs declarations and 10% risky customs declarations. This creates a challenge to reduce the learning bias through controlling the outliers and class imbalance. Traditional classification algorithms are biased toward the majority class. This may lead to degrade the classifier performance. According to the literature, using clustering techniques proved its efficiency in various domains (Olson, 2007). Thus, this study adopts clustering techniques as a promising starting point, since converting the problem into geometrical space is expected to reduce the problem complexity and avoid bias. In our problem, the presence of the HS-Code hierarchy underlines the expected benefit of using clustering since the presence of such controlling parameters already divided the input data into isolated subspaces. By transforming it in a way where the relationship between the points is represented by using Euclidian distance and this simplifies the complexity of the problem and give the ability on how to control performance and achieve the desired outcomes.

The data that is used in this work is obtained from Dubai Customs; a government entity in Dubai. The dataset represents the shipments declarations information. The dataset comprises 500,000 shipment declaration records. These records were performed in the year of 2018 and labelled to identify goods value manipulations. This research captures the corresponding Country of Origin (COO), Commodity Type and CIF value (Cost, Insurance-Freight Forwarder) features from declaration information that have the most significant impact on the goods declared value. The following section discusses the details of the proposed cluster-based with the various experiment settings to capture the performance of this approach.

4. Cluster-based Approach

The proposed approach considers each goods type as an isolated cluster. This isolation aims to simplify the analysis process since the attention is to eventually investigate the goods values that share the same commodity type. The proposed approach analyses the shipment in terms of behaviours and patterns. Behaviour analysis is performed during the first stage by taking into consideration the *Euclidian* distance between the points. On the other hand, pattern analysis is performed by analysing the distribution of the points inside each cluster. Once the first stage completes its execution, the new point is classified into either normal or abnormal. If a shipment is classified as abnormal, the first stage has concluded that based on the locality of the new shipment in its cluster, this shipment has abnormal features. This will trigger the second stage. If the first stage tags a shipment as normal, this result will be confirmed and the second stage will not be triggered. In both of these stages, the following features are used to represent the shipments: HS-Code, Commodity Value (Cost), COO, Declared CIF Value. HS-Code will be used for clustering purposes where other three attributes have a clear relationship that can be used to determine if a value manipulation has occurred. For instance, there a strong relationship between the COO and the other two values Commodity value and CIF value. COO is considered as significant indicator about the commodity value and CIF value.

Once the clusters are constructed, the distance-based analysis starts processing the entered goods value. During this stage, a *similarity* value for the new entered goods value will be calculated. This value determines whether the new goods value can be considered as normal or not. The calculation of this value takes into consideration the distribution of the points inside the clusters and the locality of the new price value inside the cluster. If the calculated value denotes a normal price, the status of the new entered goods value will be confirmed as normal.

On the other hand, when the similarity value highlights that the entered goods value is abnormal, the density-based analysis stage will be triggered to confirm the status of the new price. The density-based analysis stage examines the new goods value against all available historical approved declarations that have been processed.

The core of the density-based analysis is to use an outlier detection mechanism to either confirm the abnormality of the entered goods value or not. Using this mechanism, the status of the new goods value will be confirmed as *normal* when the density-based analysis determines an active goods value pattern.

4.1. The Distance-based Analysis

In this stage, each goods type that shares the same HS-Code is represented as a single cluster (partition). The objective of this step is to determine whether the declared goods price can be considered normal. To achieve this objective, this step takes into consideration the distribution of the processed declaration values that are represented as points in each cluster. This step uses a distance-based mechanism to detect the abnormality of the new declaration entered value.

For each cluster, this stage starts by determining the cluster centroid, which is defined as the closest point to all other points within the cluster. The distance between any point and the cluster centroid can be considered as a reasonable indicator whether the new point (goods value) can be considered normal. However, the distance between the same cluster points must be also considered in the process of determining the normality of any given point. Algorithm 1 illustrates the process of this stage. Once the clusters representation is constructed. The distance (dc) between the new point (goods price), and the cluster centroid is calculated. This distance alone represents how far the new point from the rest of the cluster points. Beside this distance, it is also required to determine the distance between the new point and its neighbouring points. This can be calculated by determining the average distance between the new point and all other points inside the cluster (dn). These two distances values are used to calculate the similarity values as follows:

$$Similarity = \left(\frac{dc}{mc}\right) \times \left(\frac{dn}{md}\right)$$

Where md represents the maximum average distances between the cluster points, and mc represents the maximum average distances between the cluster's points and the cluster centroid. For any new shipment point, if the similarity value of this point is less than one (<1), the declared goods value point will be considered as normal. In case that the similarity value is higher than one (>1), the declared goods value will be considered as an abnormal and this will trigger the density-based analysis stage.

Algorithm 1: The Proposed cluster-based algorithm

<i>DistanceBasedAnalysisStage</i> (p)	
<i>Input</i> p: New Shipment Point Information	
<i>Output</i> status	
1	$c = \text{LocatePointCluster}(p)$
2	$dc = \text{FindDistancetoCentroid}(p, c)$
3	$dn = \text{AvgDistancetoNeighbors}(p, c)$
4	$md = 0$
5	$mc = 0$
6	For each point (j) in c
7	If ($\text{FindDistancetoCentroid}(j, c) > mc$) then
8	$mc = \text{FindDistancetoCentroid}(j, c)$
9	If ($\text{AvgDistancetoNeighbors}(j, c) > md$) then

10		$md = AvgDistanceToNeighbors(j, c)$
11	End For	
12		$s = (dc/mc) \times (dn/md)$
13	If ($s \leq 1$) then	
14		Return "Normal"
15	Else	
16		Return <i>DensityBasedAnalysisStage</i> (p)

Figure 3 illustrates an example about the distance-based analysis stage. From the figure, the point that is highlighted in orange colour is more likely to be considered as an outlier compared to the point highlighted in green colour.

Figure 3: Example of distance-based analysis stage

4.2. Density-based Analysis Stage

Once the distance-based analysis stage has labelled a shipment as an outlier, the density-based analysis stage will be triggered to examine the new shipment in terms of density. The distance-based analysis uses distance measures to determine whether any shipment is normal or an outlier. Accordingly, in order to confirm the outlier label for any shipment in the density-based analysis stage, the density of the cluster is examined beside the new shipment point location. In this stage, the LOF algorithm is employed. The LOF algorithm determines whether any given point is an outlier based on the distribution of the points and the density of the points across the entire cluster. For each point, the LOF algorithm returns an outlier factor value greater than or equal to zero (≥ 0). In an instance when the outlier factor value for a new given point is less than one (< 1), this point will be considered as an inlier (normal). However, the new point will be considered as an outlier (abnormal) when the outlier factor value is greater than one (> 1).

In this stage, once the new shipment cluster is identified, all of the cluster points in addition to the new points will be used as inputs for the LOF algorithm. This algorithm will analyse the shipment information in order to determine whether any pattern that validates the declared value is present in the historical data.

5. Experiments and Discussion of Results

To evaluate the performance of the proposed approach, various experiments with different settings have been conducted. The dataset used in this work represents 500,000 declarations applications obtained from Dubai Customs for the year of 2018. These declarations are labelled to normal and abnormal with respect to the declared goods value. These declarations are pre-processed to eliminate all records that have duty exemptions. In addition, the least 10% of the clusters with minimum number of declarations are eliminated. The normal declarations are considered as the historical available shipments' declarations, and these declarations will be used to construct the clusters. The testing sample was prepared formulating random 10 groups, where each groups includes 200 items divided into 100 abnormal items and 100 normal items. Through these experiments, the following experiment parameters are used:

- **Number of declarations:** the number of used declaration groups for evaluation purposes.
- **The size of used clusters:** the percentage of HS-Code clusters used for evaluation purposes.

To evaluate the performance for each experiment settings, the accuracy and the precision are calculated. The accuracy is the ratio of the correctly predicted shipments value manipulation status to overall predictions. And the precision is the ratio of correctly predicted shipments with value manipulation to the total shipment with value manipulation. Various experiments were conducted with varying the size of the used clusters using the full testing sample (10 groups).

Figure 4 illustrates the accuracy and precision of the proposed approach against the size of the used clusters. In these experiments, the clusters are grouped in ascending order based on the number of available declarations for each cluster. Accordingly, the percentage of used cluster is varied with highest number of points as recorded in the experiment. According to the results, reducing this percentage results in reducing the gap between the accuracy and the precision.

In addition, reducing this percentage will have significant impact on the overall performance of the proposed approach. This is due to several factors; reducing the percentage of the selected clusters results in selecting the clusters with the largest sizes. Having clusters with large sizes improve the classification performance of the proposed approach. The reason is that the size of each cluster plays a major role in determining the authenticity of the obtained result. The size plays a major role in the first stage since it is a distance-based mechanism and impacts the result of the second stage since it will influence the density of the resulted clusters. This explains the seen performance and clarifies the gap between the accuracy and the precision.

Figure 4: Performance Results against the Size of the Used Clusters

Another set of experiments were conducted with different settings that consider the size of used clusters.

Figure 5 illustrates the accuracy and precision of the proposed approach against the number of used testing sample groups using the largest 10% of the clusters. The figure shows that increasing the number of used testing samples results in improving the overall performance of the proposed algorithm. In addition, reducing the number of used testing samples increases the gap between the accuracy and the precision.

In other words, the number of declarations affects the performance of the proposed approach. For instance, increasing the number of used declarations stabilizes the performance of the proposed approach since this will increase the likelihood to capture several experimental scenarios. This is evident in the results since increasing the number of used declarations to over eight groups have no clear influence on the behaviour.

In addition, as shown in Figure 5 at lower number of groups with less than or equal to three, the performance of the proposed algorithm was not stable. Regarding the gap between the accuracy and the precision, performing experiments using low number of declarations will increase the probability of inaccurate classifications. This explains the reason behind the seen gap between these two metrics.

Figure 5: Performance Results against the Number of Used Declarations

6. Conclusion and Future Works

The customs valuation is a critical function in any customs administration that serves the objectives of society protection and trade facilitation. Furthermore, it protects the revenue by identifying and preventing the leakage caused due to fraud activities by providing undervalued goods declarations. In this paper, a cluster-based approach to detect customs value manipulation is proposed. The proposed approach consists of two stages; the distance-based analysis stage and the density-based analysis stage. These two stages work to capture the abnormality of the shipments in terms of declared goods value. The first stage focuses in its analysis on the distances between the points that represent the shipments. While, the second stage focuses on the density aspects of these points to confirm result of the first stage. The results achieved an accuracy of 86% which shows the importance of combining both behavioural and pattern analysis to process shipments declaration. This also shows the potential of the proposed algorithm to simplify and automate the process of detecting the value manipulation as part of the declaration process at customs domain.

In this paper, LOF algorithm has been used to capture the relationship between the points in terms of density. Exploring the use of different outlier detection algorithms may have significant improvement on the performance of the proposed algorithm. In addition, there is an area of research to expand the scope of the proposed approach in this work to also cover the problem of goods smuggling. This problem has a direct relationship with the value manipulation mechanism since traders who try to smuggle undeclared goods will most likely to falsify the declared goods values.

Compliance with Ethical Standards

Conflict of Interest: The authors declare that they have no conflict of interest.

References

- Almiani, K., Chawla, S., & Viglas, A. (2014). The Effect of Outliers in the Design of Data Gathering Tours. *2014 Sixth International Symposium on Parallel Architectures, Algorithms and Programming*, 209–214. <https://doi.org/10.1109/PAAP.2014.23>
- Almiani, K., Viglas, A., Lee, Y., & Abrishambaf, R. (2018). Peripheral nodes and their effect in path planning in networks. *International Journal of Ad Hoc and Ubiquitous Computing*, 27(3), 157–170. <https://doi.org/10.1504/IJAHUC.2015.10001796>
- Azevedo, A., & Santos, M. F. (2008). KDD, semma and CRISP-DM: A parallel overview. *MCCSIS'08 - IADIS Multi Conference on Computer Science and Information Systems; Proceedings of Informatics 2008 and Data Mining 2008, January 2008*, 182–185.
- Badriyah, T., Rahmaniah, L., & Syarif, I. (2018). Nearest Neighbour and Statistics Method based for Detecting Fraud in Auto Insurance. *2018 International Conference on Applied Engineering (ICAE)*, 1–5. <https://doi.org/10.1109/INCAE.2018.8579155>
- Bhaduri, K., Matthews, B. L., & Giannella, C. R. (2011). Algorithms for speeding up distance-based outlier detection. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '11*, 859. <https://doi.org/10.1145/2020408.2020554>
- Ceronmani Sharmila, V., R., K. K., R., S., D., S., & R., H. (2019). Credit Card Fraud Detection Using Anomaly Techniques. *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, 1–6. <https://doi.org/10.1109/ICIICT1.2019.8741421>
- Chen, M.-C., Wang, R.-J., & Chen, A.-P. (2007). An Empirical Study for the Detection of Corporate Financial Anomaly Using Outlier Mining Techniques. *2007 International Conference on Convergence Information Technology (ICCIT 2007)*, 612–617. <https://doi.org/10.1109/ICCIT.2007.4420326>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence*, 17(3), 37–54. https://doi.org/10.1007/978-3-319-18032-8_50

- Gan, Z., & Zhou, X. (2018). Abnormal Network Traffic Detection Based on Improved LOF Algorithm. *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 1, 142–145. <https://doi.org/10.1109/IHMSC.2018.00040>
- Gao, Z. (2009). Application of Cluster-Based Local Outlier Factor Algorithm in Anti-Money Laundering. *2009 International Conference on Management and Service Science*, 08, 1–4. <https://doi.org/10.1109/ICMSS.2009.5302396>
- Malini, N., & Pushpa, M. (2017). Analysis on credit card fraud identification techniques based on KNN and outlier detection. *2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, 255–258. <https://doi.org/10.1109/AEEICB.2017.7972424>
- Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *Knowledge Engineering Review*, 25(2), 137–166. <https://doi.org/10.1017/S0269888910000032>
- Olson, D. L. (2007). Data Mining in Business Services. *Service Business*, 1(3), 181–193. <https://doi.org/10.1007/s11628-006-0014-7>
- Pete, C., Julian, C., Randy, K., Thomas, K., Thomas, R., Colin, S., & Wirth, R. (2000). Crisp-Dm 1.0. *CRISP-DM Consortium*, 76.
- Rosenow, S., & O'Shea, B. J. (2010). *A handbook on the WTO customs valuation agreement*. Cambridge University Press.
- Siyam, N., Alqaryouti, O., & Abdallah, S. (2020). Mining government tweets to identify and predict citizens engagement. *Technology in Society*, 60(November), 101211. <https://doi.org/10.1016/j.techsoc.2019.101211>
- Vanhoeyveld, J., Martens, D., & Peeters, B. (2020). Customs fraud detection. *Pattern Analysis and Applications*, 23(3), 1457–1477. <https://doi.org/10.1007/s10044-019-00852-w>
- WCO, W. T. O. (2019). *Customs Valuation*. https://www.wto.org/english/tratop_e/cusval_e/cusval_e.htm