Machine learning for detection of trade in strategic goods: an approach to support future customs enforcement and outreach

Christopher Nelson

Abstract

Customs authorities have a critical role in identifying trade in strategic goods that could have an adverse impact on international security. This study proposes a method of using data resampling and the Random Forest machine learning algorithm to model common patterns and characteristics that separate transactions involving strategic goods from broader international trade flows. By embracing advances in machine learning and computing power, customs authorities can leverage existing data to improve enforcement and outreach efforts related to strategic goods subject to international export control regulations.

1. Introduction

Customs authorities must balance facilitation of legitimate trade with the obligations of international security. As the volume of international trade has increased over the years, states have taken many steps to improve efficiency in both trade facilitation and detection of fraudulent activity. There remains, however, a gap in the ability of states to identify potential instances of illicit trade in strategic goods. The World Customs Organization (WCO) defines strategic goods as 'weapons of mass destruction (WMD), conventional weapons, and related items involved in the development, production, or use of such weapons and their delivery systems' (WCO, 2019, pp. 8–9). While comprising a small amount of overall international trade, these commodities represent an essential piece of the international customs security mission.

Advances in data collection and computing power provide the opportunity to analyse the vast amount of data collected by customs authorities every day through shippers' export declarations. This study proposes that machine learning techniques can profile trade in strategic goods, identifying patterns that may allow customs authorities to more effectively recognise these transactions. A longstanding issue with identifying trade in strategic goods has been the difficulty in correlating the Harmonized System (HS) codes to export control classification numbers. The approach proposed in this study would use transaction records to identify how HS codes are being utilised for strategic goods in order to detect this behaviour in the broader context of trade not subject to export control licensing requirements.[1]

Utilising resampling and the Random Forest algorithm, this study proposes that customs authorities can use supervised learning techniques to identify the signatures of strategic goods based on transaction data. This approach would create models for a portfolio of high-priority strategic goods, considering state-specific traits, such as trading partners and the domestic industrial base. These models could be used to identify entities that are shipping commodities that match the characteristics of a strategic good but have not sought out the proper export licence. In addition, they could be used to improve outreach efforts, identifying where entities did not seek a licence and promoting more accurate HS code classifications.

By embracing recent advances in machine learning techniques, data collection and computing power, customs authorities would be in a better position to identify and respond to strategic trade flows in the future.

2. The problem

Strategic goods comprise a wide range of dual-use commodities that have legitimate commercial uses, such as chemicals, materials, electronics and manufacturing equipment. This study considers strategic goods to be those items subject to national or international export control regimes, such as the Wassenaar Arrangement, Nuclear Suppliers Group, Australia Group, and others. These regimes outline the agreed upon commodities and the technical parameters that elevate the item to a controlled commodity requiring a review by state authorities before export. Examples of strategic goods include high-end machine tools, mass spectrometers, radiation hardened integrated circuits, carbon fibre and rocket propulsion systems.

In attempting to identify and regulate trade in these strategic goods, state authorities need to create an effective licensing and enforcement process that can undertake tasks such as identifying relevant transactions, assessing end-users and proposed end-use, and uncovering violations. Individual state regulations, international agreements and partnerships commit states to putting effective systems in place to control strategic goods. United Nations Security Council Resolution (UNSCR) 1540 was a major step forward in strategic trade controls, requiring states to prevent the unregulated transfer of WMD-related commodities. However, when it was initially implemented in 2004, most states lacked an effective system of identification and control, particularly if 'these goods were intentionally routed around national licensing authorities' (Perry, 2019, p. 10).

Identifying trade in strategic goods is a difficult problem for state authorities for many reasons. First, the volume of trade not subject to export control requirements is vastly larger than trade in strategic goods. As an example, in 2018, the United States had approximately $1.7 trillion in total exports. Of that, $44.6 billion, or 2.7 per cent, were exported under a government licence (US Department of Commerce, 2018). This imbalance, which varies by state and commodity, makes it very difficult to identify relevant transactions amid the noise. In addition, the main classification systems for international trade and strategic trade have different objectives and do not correlate well. The lack of a universally applicable and accepted system that can delineate trade not subject to export control requirements and strategic trade is one of the key drivers for the approach outlined in this study. State authorities must also consider efforts to obfuscate the commodities being traded, transshipment through different states, and lack of knowledge by the entities trading the goods. The massive amount of data collected in the course of international trade has compounded these issues, making traditional data or risk analysis an unwieldy undertaking.

State authorities have taken a variety of approaches to improving the effectiveness of customs targeting. In general, many state authorities focus on identifying misclassifications in order to ensure they collect proper import taxes and duties from goods entering their country. A traditional approach has been to identify trade gaps between the value reported by the exporting country and the value declared upon import. Using available international trade statistics, these gaps can help identify under or overvaluation for commodities and help target customs enforcement (Chalendard, 2017, p. 12). Another common approach is risk rating shipments based on factors such as country of origin, commodity type, entity profiles and tolerance measures. This relies on in-depth past knowledge to create risk profiles and accuracy in the declarations for the system to work reliably. State authorities can also utilise pre-shipment inspection processes in which importers (and to an extent the exporters) are required to engage private companies to inspect the cargo and ensure that it fits the declared quality, quantity and price (Dictionary of International Trade, n.d.).

The WCO has created the *Strategic trade control implementation guide*, which outlines the different stages of an effective state-level control system. These steps provide a starting point for states to set goals and measure progress toward an effective customs system. At the Established Capability phase, there are 'established outbound enforcement teams capable of targeting and inspections', 'national risk management (targeting) centers', and the ability to process intelligence tips to inspect and/or interdict (WCO, 2019, p. 16). An Enabled Capability goes further, adding synthesis between licensing and customs capabilities and regular feedback loops and coordination mechanisms. The WCO guide outlines best practices for these areas, but it is ultimately up to the state to implement the specific procedures.

Efforts in utilising past trade records and customs declarations have also moved forward. State authorities already collect and maintain a massive amount of trade data, from licensing information to bills of lading. For example, the Hong Kong China Customs' Central Information Repository data warehouse increased in size from three terabytes in 2012 to 12 terabytes in 2015 (Okazaki, 2017, p. 11). These records contain a wealth of data that can assist in identifying the signatures of a transaction, such as value, weight, quantity, HS code, destination, consignees, and export control numbers. The key question is how can state authorities get a handle on this increasing volume of data? In addition, how can the small amount of strategic trade be identified within the massive amount of transactions? Leveraging machine learning algorithms and the increasing computing power to handle larger and larger datasets can help state authorities take a meaningful step toward the future of more effective commodity identification and transaction targeting.

3. Identifying strategic goods amid the noise

The nomenclatures to track trade generally, and strategic goods specifically, were designed with different fundamental objectives and therefore do not correlate well. This non-alignment is the core of the strategic trade control problem and has been the focus of many research efforts and studies. First, the HS is a universal classification system designed to facilitate tariffs. Since these codes are uniform for all parties, 'the Customs' clearance process is expedited, tariff collections are readily determined, and commercial disputes like levying inappropriate duties due to misclassification of traded goods are avoided' (Kim, 2017, p. 71). On the other hand, strategic trade controls are designed with the objective of regulating the flow of specific commodities with specific parameters in specific use cases or destinations. The inherent structure of both systems are different:

> In [strategic] trade controls, multiple dimensions such as functions, raw materials, and industrial specifications are considered together in the licensing process while the HS begins by classifying goods based on a specific industrial sector and then hierarchically drills down (Kim, 2017, p. 74).

Attempting to correlate export control classification numbers (ECCNs) for strategic goods with their pertinent HS code is an extremely difficult task. The HS–ECCN relationships that would easily identify strategic goods are rarely one-to-one. In fact, a strategic good can be potentially shipped under a host of HS codes and an HS code could contain multiple ECCNs. There are almost never technical specifications in HS code descriptions, which are the key delineating factor on strategic control lists. In some cases, such as uranium, nuclear fuel, or certain materials, the HS code closely correlates to a controlled item. For example, depleted uranium matches exactly to HS code 284430, 'uranium depleted in U235'. Most strategic goods, however, are indistinguishable in the HS description from a mixture of other commodities included in the code that are not strategic goods. As a result, even states 'with advanced export control systems where companies have strong export compliance programs in place had been unable to prevent insiders from diverting illicit dual-use transfers' (Perry, 2019, p. 10).

On top of this ambiguity, entities self-declare the relevant HS code and can easily make a mistake or intentionally misclassify their goods, making identification even harder. Entities identify the HS code of the commodity they are trading in their shipping documents with little oversight as to the accuracy of

this declaration. If an error is made by the trading party due to a misreading of HS codes, it may not be detected by customs authorities. In addition, some entities may intentionally misclassify the HS code of their commodity to avoid export licensing requirements or tax burdens, or for general subterfuge. The nature of how HS codes are declared in shipping documents contributes to the difficulty of identifying trade in strategic goods.

Companies utilise the HS to identify their shipments and customs officers rely on these declarations to assess duties/taxes and potentially target shipments for inspection. Without a direct HS–ECCN connection, how are customs inspectors supposed to identify shipments of strategic goods when a shipper has not declared it as such? It is possible to create approximate correlation tables through research and subject matter expertise, as is the case with the European Commission's *Correlation list between TARIC and the dual-use annex of the Regulation 428/2009* (European Commission, 2020). These are extremely useful guides but are often not backed by real-world shipment data. They do not allow for any systematic review of transactions for misclassification. Strategic goods shipped under the incorrect HS code can only be found through other risk mitigation strategies, such as targeted inspections, end-use checks and entity profiles. These correlations cannot be deployed to find illicit trade in strategic goods on a large scale.

A potential solution lies in the data that states and their customs authorities currently maintain. States could exploit existing data gathered from shipping declarations to create a data-driven system to target strategic goods. Machine learning techniques can amplify traditional analytic approaches to provide predictive classifications to create strategic commodity profiles utilising attributes such as the ECCN, HS code, value and quantity. The approach detailed in this study would use previous export data of commodities that have an ECCN in the shipping documentation along with shipments of commodities not subject to export licensing requirements within the same basket of HS codes to create a model to identify the key characteristics of shipments of a specific strategic good.

4. Machine learning for detection of strategic goods

Machine learning is 'centered around making predictions, based on certain trends and patterns that have been already identified through data analytics' (Okazaki, 2017, p. 7). From this definition, we can create a path to what states would ideally like to do regarding strategic trade: predict whether a particular transaction matches a pattern that indicates it could be a strategic good that needs to be regulated.

In this case, states would take historical export transactions and use attributes such as value, quantity, destination, and HS code to train a model to classify whether a new transaction is likely to be subject to strategic trade controls. Doing so would allow authorities to be proactive in identifying critical transactions for enforcement or outreach on a wide scale. Since machine learning relies on models that can adapt to new data, it refines itself over time. In addition, since we can use transaction data that we know involve strategic goods, we will have various ways to check the performance of our model.

Applying big data analysis or machine learning to international trade data is not a new concept. One method to identify misclassification or tax avoidance is to utilise mirror trade statistics. By comparing an exporting state's declaration with information from the importing state, authorities may be able to detect patterns of under or overvaluation, identifying transactions or commodities where entities commonly avoid taxes or duties (see Chalendard, 2017; Cariolle et al., 2018). In other cases, the simple use of historical transaction data may allow authorities to recognise patterns in new transactions that indicate strategic goods or tax/duty evasion (see Digiampietri et al., 2008; Chalendard, 2017). Recent studies by Chermiti (2019) and Zhou (2019) apply different approaches to decision tree algorithms in customs risk detection and profiling. The approach proposed in this study expands upon aspects of these past efforts but focuses on the identification of strategic goods rather than the risk of duty evasion.

5. Data collection and preparation

Classifying transactions involving strategic goods lends itself to solutions developed for a relatively common machine learning problem: outlier detection. Outlier detection is used for a wide variety of applications, such as credit card fraud, suspicious traffic in cyber security, disease detection, and many other problems where we seek to find 'patterns in data that do not conform to expected behavior' (Singh & Upadhyaya, 2012, p. 307). Since strategic trade is such a small proportion of overall trade, we assume these transactions are not the norm and can be considered outliers. This approach in analysing international trade has been taken before, but primarily to detect tax/duty avoidance during the import of goods rather than the detection of strategic trade (see Digiampietri, 2008; Laporte, 2011).

The two key aspects of an outlier are that 'They are the minority consisting of few instances, and they have attribute-values that are very different from those of normal instances' (Liu et al., 2012, p. 2). As mentioned previously, in the United States in 2018, only 2.7 per cent of global exports were exported under a government licence. Trade in strategic goods is rare when compared to all international trade. Since strategic goods are controlled for their specialisation, advanced technology or technical parameters, they tend to have distinctive characteristics compared to commodities shipped under the same HS code or set of HS codes such as tending to be shipped in lower quantities at higher values. If strategic goods were not comparatively rare, the export controls involved would be mute. An international trade control 'becomes inapplicable when an item is too commonly traded, and items can be de-listed for this reason' (Chatelus & Heine, 2016, p. 52).

The first step is to identify the strategic good that the customs authorities want to prioritise and create a model for, based on its ECCN. After the ECCN is identified, data for controlled export transactions that identify this ECCN in the shipping documentation would be pulled together for a given timeframe. This timeframe could be decided based on the amount of data or other considerations, such as regulation changes, end-use destination or geopolitical events. Once this data is gathered, we would create a 'basket' of the different combinations of the ECCN and HS codes. This basket would show how often a particular HS code is utilised by exporters for transactions involving the particular ECCN (e.g. 45% of transactions involving strategic goods classified under ECCN X were shipped using HS code Y). This allows us to identify the HS codes that are *actively being used* for transactions involving a strategic good rather than relying on a correlation table that says what the HS code for a strategic good *should be*.

An example of a such a basket for maraging steel (ECCN 1C116) is in Table 1. Note that the percentages are examples only and not based on actual transaction data.

Table 1: Example HS basket for maraging steel (ECCN 1C116)

HS code	Percentage of transactions
720429	45%
722692	17%
722090	15%
722810	11%
720521	7%
721129	2%
722540	1%
731822	1%
210690	1%

Source: Zauba.com for HS codes

The gathered ECCN–HS correlations often contain codes that are utilised at a low rate. This may be due to misclassification, an outlier case, or other issues. In subsequent steps, transactions of commodities not subject to export licensing requirements for these HS codes will be gathered. As such, a reasonable cut-off point should be set so as to not include a large amount of transactions that may be irrelevant just because the HS code is in the basket. For example, a cut-off could be set at 10 per cent—only HS codes utilised in 10 per cent or more of the transactions with an ECCN would be included in further analysis. Table 2 shows the revised example basket for maraging steel with a 10 per cent cut-off.

Table 2: Example HS basket for maraging steel with cut-off

HS code	Per cent of transactions
720429	45%
722692	17%
722090	15%
722810	11%
~~720521~~	~~7%~~
~~721129~~	~~2%~~
~~722540~~	~~1%~~
~~731822~~	~~1%~~
~~210690~~	~~1%~~

Source: Zauba.com for HS codes

Once the most used HS codes are identified, data would be pulled from all transactions with commodities not subject to export licensing requirements with these codes for the same time period from which we drew the transactions with ECCNs. This assembles the universe from which we can model the characteristics of trade in the particular strategic good as opposed to the transactions with commodities not subject to export licensing requirements. This assembles a set of known transactions involving strategic goods and commodities not subject to export licensing requirements with known outcomes, becoming a supervised learning exercise. Since we know the ultimate content of each transaction, we can test how well our models perform in predicting classifications.

Customs authorities gather a large amount of information from international trade transactions. There are many attributes included in a shipper's export declaration or bill of lading that could be useful in the process of attempting to identify trade in strategic goods. Machine learning algorithms can handle large numbers of disparate features and there are techniques to select the most impactful factors in determining the ultimate classification through feature selection, which is beyond the scope of this study. As a baseline, some key features for analysis could be:

- quantity
- net shipping weight
- monetary value
- HS code
- ultimate destination.

6. SMOTE resampling

In gathering the set of transactions identified above, transactions with commodities not subject to export licensing requirements will outnumber the transactions that have an ECCN for the targeted commodity. In machine learning, imbalanced data can have adverse effects on modelling. For example, if it is assumed that every transaction does not involve a strategic good there may be a very high accuracy rate, but we will miss the objective of this approach to detect the minority class. To identify the outliers in the broader volume of international trade it is necessary to resample our data. In this case, the resampling will bring the minority class into balance with the majority class, which allows for better performance of modelling.

The techniques for resampling involve undersampling the majority class or oversampling the minority class. Undersampling would involve removing transactions with commodities not subject to export licensing requirements at random to bring the data in balance with the number of transactions involving strategic goods. This could omit valuable information that would distinguish the categorical differences that could delineate strategic goods. Oversampling would involve duplicating transactions involving strategic goods to match the amount of transactions with commodities not subject to export licensing requirements. This could create an overemphasis on certain characteristics, overfitting the model to specific transactions creating a model that might not fit in accurately predicting new cases.

A balanced approach that has shown impressive performance in handling imbalanced data is the Synthetic Minority Oversampling Technique (SMOTE). This approach identifies similar examples in the minority class and creates new instances that share the space between them. Rather than duplicating transactions to oversample, this technique provides 'new' examples of the minority class. As the originators of this approach summarise, 'With replication, the decision region that results in a classification decision for the minority class can actually become smaller and more specific…This is the opposite of the desired effect. Our method of synthetic over-sampling works to cause the classifier to build larger decision regions that contain nearby minority class points' (Chawla et al., 2002, p. 352).

7. Machine learning with random forest

Once prepared, the data is ready to be used to create a model for classifying transactions involving strategic goods. A minority portion of the collected dataset should be set aside for unbiased testing of the model once it is trained on the data, which will provide a measure of performance.

This study proposes the use of the random forest algorithm to predict whether a transaction involves a strategic good. Originally proposed by Breiman (2001), this model creates many decision trees based on randomly selected features and data samples to determine the classification of a transaction. A decision tree tests one feature at each decision node, splitting into sub-nodes in order to maximise the homogeneity of the resulting groups. It continues splitting until it can put the data, with a high probability, into a leaf that identifies the classification of the record. Decision trees have the advantage of being more easily visualised and interpreted than other classification algorithms. Random forest expands this model by creating hundreds or thousands of randomly generated trees using different features and subsets of data. In our case, each tree would generate a prediction for each transaction and the final classification would be decided by a majority vote. Figure 1 presents a simplified representation of the random forest algorithm. Our models would be a binary classification; either the transaction is classified as involving a strategic good or not. In Figure 1, two trees predict that the transaction does involve a strategic good and one does not, therefore the overall prediction is that there is a strategic good involved.

Figure 1: Random forest algorithm
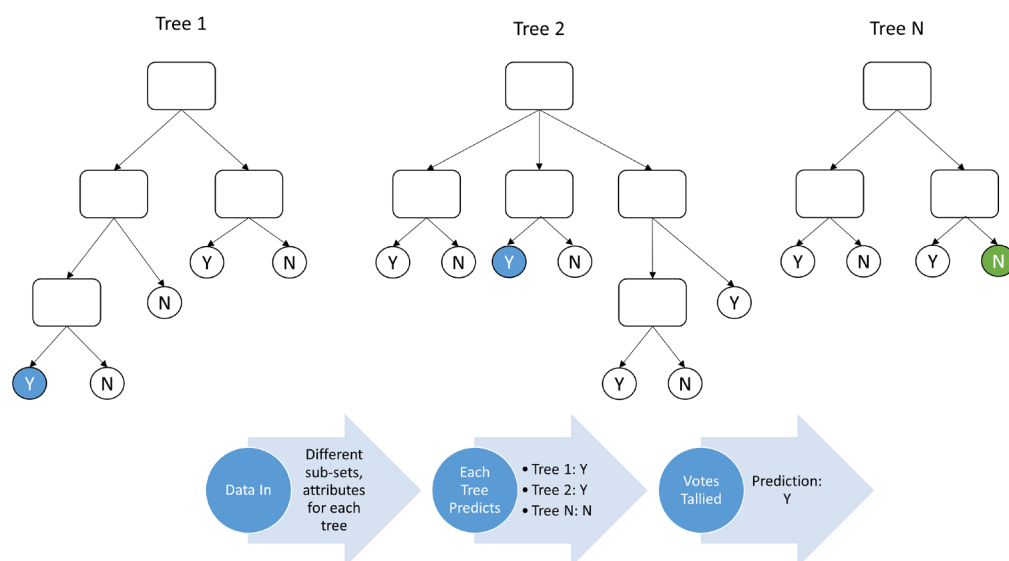
This algorithm has many advantages. By utilising random samples of features and data over many trees, it can help prevent overfitting the model. Individual decision trees, unless pruned or controlled, can continue splitting the dataset in increasingly specific ways, creating a very specific model that cannot generalise to new data. In addition, random forest is very versatile, handling categorical and numeric data. Many machine learning algorithms require workarounds for categorical data. Applications of this algorithm also provide an ability to identify the importance of individual features to the final classification. This would allow the identification of which aspects of transactions, such as destination, HS code or value, are most valuable in predicting whether a transaction involves a strategic good. By utilising a multitude of trees, no individual test of the data will predominate, increasing confidence in the ultimate classification. Random forest can be computationally expensive; however, parameters of the algorithm may need to be adjusted for the amount of data involved.

The algorithm would be trained on the data and performance would be measured against the reserved test set, which has a known classification. Based on the test, parameters or features could be changed to increase performance. This model would apply to a particular strategic good. After the approach has been tested, it can be used iteratively to create models for a broad portfolio of strategic goods based on a state-level risk or priority assessment. Also, once these models are created, they could be applied as new data arrives. Pre-shipment data or shippers' export declarations would be inputted into the models to determine if the transaction is likely to contain a particular strategic good. The creation of multiple models has the advantage of taking into account the characteristics of particular commodities in addition to providing a prediction of which particular strategic good may be involved. A general model identifying if any strategic good was likely involved in a transaction would mix a multitude of very disparate commodities that have varying levels of risk or priority.

8. Benefits, potential applications and next steps

Given the data-dense nature of international trade transactions, customs authorities around the world are in an excellent situation to exploit advances in machine learning to improve risk analysis, enforcement and outreach. As more transactions are recorded every day, the models created to classify strategic goods

can improve, be adjusted and reworked under the same methodological construct. In addition, since this approach proposes the use of state-centric data, the models will inherently be designed to identify strategic goods in the context of that state, taking into account geography, trading partners and industrial capabilities. The recent expansion of distributed computing and cloud-based services allows for state authorities to analyse and create models for a much larger portion of data that could be handled even five to ten years ago.

Beyond better handling of data, the ongoing advances in machine learning algorithms and techniques continues to push the boundaries of classification predictions. The random forest algorithm was first introduced in 2001 and further refinements to it and other models continue constantly (Breiman, 2001). Random forest was chosen for this proposal because it can handle a wide range of data types and can also identify the most important features to the classification of a transaction as involving or not involving a strategic good. Practical testing may find other methods to provide better performance or efficiencies, but this is exploration and could be very fruitful in many customs analyses.

The classification of transactions involving strategic goods has a wide variety of useful applications for states. From an enforcement perspective, this approach would allow for better profiling of transactions involving strategic goods using real-world data. Reliance on the accuracy of licence applications and proper HS code declarations is often inaccurate for a variety of reasons. Training a random forest model on transactions declared to be a strategic good and those that are not can allow for identification of patterns that delineate the transaction types. Once trained and refined, these models could be applied to incoming transactions, thereby enhancing risk profiling and potential documentation or end-use checks. Modelling based off a select set of high-priority strategic goods could enhance resource allocation and provide data-based justifications for inspections. This approach would also allow for states to better understand common trade flows for strategic goods and identify potential transshipment points. The random forest model can consider the origination and destination points of the transactions. If a particular trading point appears as a key node in the model, it may assist in enforcement targeting or outreach efforts.

In addition to enforcement, this approach could be used to design outreach efforts that would cyclically improve overall customs efficiencies. First, based on the models trained through existing data, it would be possible to identify transactions that fit the profile of a strategic good, but were not licensed. The entities involved in these transactions could be identified for training on export control regulations and future end-use checks. Since this approach uses a basket of HS codes to identify how entities are shipping strategic goods, it could also be used to improve the way states and entities classify which HS codes are used in practice and which should be used in the trade of a strategic good. This could provide a baseline for how entities in a state are operating and how customs authorities might work with them and their international partners to make more effective use of the HS. If this approach proved effective, over time the data collected by customs authorities would be more comprehensive and accurate, thus improving the modelling over time. In addition, the exchange of strategic good transaction models with other trading partners could improve detection of relevant imports and sharing of best practices.

There are many potential next steps to test and improve this approach. This study was limited to a theoretical exercise based on the author's experience. The logical next step would be to apply this machine learning technique to state-level data. Based on findings from such an experiment, many refinements could be made, and the model could be tested against real-world data through a multitude of classification metrics. There are also many other cutting-edge machine learning techniques that could potentially increase performance and the range of applications. Advances in natural language processing could allow for models that include a breakdown of text-based fields on transaction descriptions, flagging for key words or terms related to strategic goods. The expansion of deep learning and neural networks as a machine learning approach has provided increased performance and flexibility in model building in many other fields of research. Neural network algorithms are more complicated than decision trees but can handle a wide range of input variables and have the ability to create non-linear and complex

relationships between variables. This approach could be explored to train more accurate models in predicting strategic trade transactions (Mahanta, 2017).

From bank fraud to disease detection to Netflix recommendations, machine learning has vastly expanded our ability to draw insights from large datasets. Detection of trade in strategic goods is a necessary but challenging task for customs authorities. The expansion of data processing abilities and machine learning techniques with subject matter expertise sets the groundwork for the future of Customs in this area. As proliferation and advancement in manufacturing and technology in nuclear, biological, chemical and weaponry continues, it is important to leverage advances in data-driven decision-making to meet the challenge.

Disclaimer

This article was prepared by Christopher Nelson in his personal capacity. The views expressed in this article are the author's own and do not reflect the view of the New York State Office of the Attorney General or any other agency of the United States Government.

References

Breiman, L. (2001). *Random forests*. University of California. https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf

Cariolle, J., Chalendard, C., Geourjon, A. M., & Laporte, B. (2018). Measuring and improving the performance of customs valuation controls: An illustration with Gabon. *The World Economy*, *42*(6), 1850–1972.

Chalendard, C. (2017). Using internal and external sources of information to reduce customs evasion. *International Centre for Tax and Development.* ICTD Working Paper 62.

Chatelus, R., & Heine, P. (2016). Rating correlations between customs codes and export control lists: Assessing the needs and challenges. *Strategic Trade Review*, *2*(3), 34–67.

Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Chawla, S., & Chandola, V. (2011). *Anomaly detection: A tutorial. Theory and applications*. University of Sydney & Oak Ridge National Laboratory. https://webdocs.cs.ualberta.ca/~icdm2011/downloads/ICDM2011_anomaly_detection_tutorial.pdf

Chermiti, B. (2019). Establishing risk and targeting profiles using data mining: Decision trees. *World Customs Journal*, *13*(2), 39–57.

Dictionary of International Trade. (n.d.). *Pre-shipment Inspection* (PSI). https://www.globalnegotiator.com/international-trade/dictionary/pre-shipment-inspection-psi/

Digiampietri, L., Roman, M., Meira, L., Filho, J., Ferreira, C., & Kondo, A. (2008). Uses of artificial intelligence in the Brazilian customs fraud detection system. *Proceedings of the 9th Annual International Conference on Digital Government Research*, pp. 181–187.

European Commission (EC). (2020). *Correlation list between TARIC and the Dual-use Annex of the Regulation 428/2009*, EC, Brussels. https://trade.ec.europa.eu/doclib/html/155445.htm

Kim, H. (2017). Enhancing the interface between the harmonized system and strategic trade controls. *Strategic Trade Review*, *2*(3), 69–84.

Laporte, B. (2011). Risk management systems: Using data mining in developing countries' customs administrations. *World Customs Journal*, *5*(1), 17–28.

Liu, F., Ting, K., & Zhou, Z. (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data*, *6*(1), 1–39.

Mahanta, J. (2017). Introduction to neural networks, advantages and applications. *Towards Data Science*. https://towardsdatascience.com/introduction-to-neural-networks-advantages-and-applications-96851bd1a207

Okazaki, Y. (2017). Implications of big data for Customs – How it can support risk management capabilities. *Worlds Customs Organization*. WCO Research Paper No. 39.

Perry, T. (2019). Reducing proliferation risk through export control outreach: Assistance providers' use of maturity model-based approaches. *Strategic Trade Review*, *5*(7), 5–25.

Singh, K., & Upadhyaya, S. (2012). Outlier detection: Applications and techniques. *International Journal of Computer Science Issues*, *9*(1), no. 3, 307–323.

US Department of Commerce and Bureau of Industry and Security. (2018). *Statistical analysis of US trade with the world*. https://www.bis.doc.gov/index.php/country-papers/2439-2018-statistical-analysis-of-u-s-trade-with-the-world-pdf/file

World Customs Organization (WCO). (2019). *Strategic trade control enforcement: Implementation guide*. http://www.wcoomd.org/en/topics/enforcement-and-compliance/instruments-and-tools/~/media/7A05799E8D3A46C8B8355175EEBA4322.ashx

Zhou, X. (2019). Data mining in customs risk detection with cost-sensitive classification. *World Customs Journal*, *13*(2), 115–130.

Notes

1   Commodities not subject to export licensing requirements are those commodities that do not meet the technical parameter thresholds that would typically require an export license under international export control agreements, such as the Wassenaar Agreement, Nuclear Suppliers Group, Australia Group, and others. For the purpose of this research, strategic goods or those subject to export control requirements do not include those subject to national-level controls or sanctions that are applied solely on the basis of destination or export controls for other reasons than the technical specifications of the commodity in question.

Christopher Nelson

Christopher Nelson is a senior data analyst at the New York State Office of the Attorney General. He previously served as a safeguards information analyst for nuclear trade at the International Atomic Energy Agency and a trade and industry analyst at the US Department of Commerce, Bureau of Industry and Security. He holds a Bachelor's degree in International Studies and Mathematics from American University and a Master's degree in War Studies from King's College London.