Jurnal BPPK Volume 16 Nomor 1 Tahun 2023 Halaman 132-145

BADAN PENDIDIKAN DAN PELATIHAN KEUANGAN KEMENTERIAN KEUANGAN REPUBLIK INDONESIA

JURNAL BPPK

DETECTING CROSS-BORDER TRANSACTION PATTERNS USING MACHINE LEARNING: THE CASE OF INDONESIA

Gitarani Prastuti[a], Indah Permatasari[b], Lasmin[c], Ag. Sigit Adi Satmoko[d], Arman Imran[e]

[a] Directorate General of Taxes, Blavatnik School of Government, Oxford University, Oxford, United Kingdom. Email: gitarani.prastuti@pajak.go.id, gitarani.prastuti@gmail.com (penulis berkorespondensi)

[b] Directorate General of Taxes, The University of Melbourne, Melbourne, Australia. Email: indah.permatasari@pajak.go.id

[c] Directorate General of Taxes, Indonesia. Email: lasmin@pajak.go.id

[d] Directorate General of Taxes, Indonesia. Email: ag.satmoko@pajak.go.id

[e] Directorate General of Taxes, Indonesia. Email: arman.imran@pajak.go.id

ARTICLE INFORMATION

ARTICLE HISTORY
Received
12 September 2023

Accepted to be published
24 November 2023

KEYWORDS:
Machine Learning
Cross-Border Trade
Intercompany Transaction
International Trade
Customs and Excise
Indonesia

ABSTRACT

Penelitian ini mengeksplorasi perdagangan lintas batas antar perusahaan afiliasi yang melibatkan entitas Indonesia antara tahun 2018 dan 2022 untuk menilai risiko ketidakwajaran transaksi *(undervalued)* ekspor/impor barang dengan algoritma *machine learning*. Dengan menggunakan data mikro dan makro, penelitian ini mengungkap pola lintas batas antar-perusahaan sehingga menunjukkan risiko undervalued yang lebih tinggi dalam perdagangan lintas batas antara entitas di Indonesia dan di luar negeri. Temuan penelitian ini antara lain: 1) Perusahaan penanaman modal asing (PMA) lebih cenderung melakukan perdagangan lintas batas jika dibandingkan dengan perusahaan penanaman modal dalam negeri (PMDN), sehingga memiliki risiko yang lebih tinggi, 2) Perdagangan lintas batas antar perusahaan afiliasi selaras dengan teori *extended gravity model*, 3) Beberapa negara tertentu menunjukkan pola serupa pada transaksi afiliasi dan setiap pola memiliki profil risiko yang berbeda, dan 4) Risiko yang lebih tinggi tecermin dari volume transaksi perdagangan dengan negara yang memiliki tarif pajak yang lebih rendah. Temuan pada penelitian ini dapat memberikan nilai tambah untuk mengidentifikasi potensi risiko ketidakwajaran transaksi pada perdagangan lintas batas. Sehingga, Otoritas Bea dan Cukai dengan menggunakan temuan ini sebagai indikator risiko untuk meningkatkan pengawasan perdagangan lintas batas negara, mendorong keadilan perdagangan, dan mengoptimalkan pendapatan negara.

This study explores the intercompany cross-border trade involving Indonesian entities between 2018 and 2022 to identify to assess the risk of undervaluing in the export/import of goods with machine learning algorithms. Utilizing both micro and macro-level data, this study uncovers unique intercompany cross-border patterns, exposing a higher risk of undervaluing in export/import activities between Indonesian entities and their counterparts overseas. Our findings indicate the following: 1) Foreign investment-based companies are more inclined to engage in cross-border trade compared to domestic-based companies, posing higher risks, 2) Intercompany cross-border trade aligns with the extended gravity model, 3) Certain countries exhibit similar patterns in affiliation transactions, each pattern potentially associated with specific risk profiles and 4) Higher risk is evident in the number of trade transactions with low-tax jurisdictions. These findings offer valuable insights for identifying the potential risk of undervaluing trade. Consequently, this can benefit Customs by using the findings as risk indicators to improve its cross-border management, promote trade fairness, and optimize state revenue.

1. INTRODUCTION

1.1. Background of Study

The world trade has experienced a slowdown in growth during 2022 and early 2023. In 2022, there was a deceleration in world merchandise trade growth, and this sluggish performance persisted during the first quarter of 2023, which plummeted by 1.0% when compared to the same period in 2022 (WTO, 2023). Nevertheless, merchandise trade in 2022, as reported by the World Bank, accounted for 50.54% of GDP, underscoring the continued significant role of international trade as a driving force for economic recovery and resilience.

Over time, the trade of goods has consistently held a dominant position within global trade flows, contributing to 78% (US$ 24.2 trillion) of the total world trade in 2022 (WTO, 2023). Further, the largest share of exports, accounting for 63%, was attributed to fuels and mining products. Manufactured goods accounted for 21% of exports, and agricultural products made up 10%. The remaining 6% of exports comprised other goods.

Looking at regional trends, Asia experienced a substantial 8.4% increase in merchandise trade value. Moreover, during the initial quarter of 2023, Asia's exports demonstrated an impressive rise of 13.3% from 2019.

In 2022, Indonesia held the 27th position among leading merchandise exporters, with a total value of 291,979 million dollars, and the 28th position among leading merchandise importers, with a total value of 237,447 million dollars. One contributing factor to these figures is the substantial volume of cross-border trade, which undoubtedly increases the trade flow among affiliated parties across countries.

As noted by Squirrell (2020), approximately 30% of global trade originates from intercompany transactions. These transactions should adhere to the 'Arm's Length Principle,' which requires that the terms reflect those in transactions between independent parties and are determined by market forces (OECD, 2022). Failure to adhere to this principle could result in distortions in duties and taxes that should have been paid, thus eroding the state's revenue.

Intercompany transactions are of significant importance to customs authorities, primarily because they contribute substantially to state revenue. Moreover, these transactions play a crucial role in ensuring that the declared values align with the actual amounts exchanged. Exploration of intercompany cross-border transactions therefore is of paramount importance. It aids in the design of more effective cross-border trade management strategies, which, in turn, help prevent potential revenue loss and promote fair contributions to state revenue

The utilization of big data analytics from extensive cross-border data empowers customs authorities to enhance the efficiency and effectiveness of managing import and export activities. In recent times, customs authorities worldwide have increasingly adopted big data analytics. According to a survey by the WCO (2021), 44% of customs authorities currently employ big data analytics, artificial intelligence (AI), and machine learning for several purposes, including enhanced risk assessment and profiling as well as improved fraud detection and compliance.

1.2. Problem Formulation

The field of international economics has provided a plethora of comprehension and empirical findings on factors that drive trade and its outcomes. However, in recent years, intercompany trade involving cross- border transactions has experienced rapid growth, raising questions about how custom and excise authorities enrich its monitoring on cross-border trade.

To address these challenges, this paper aims to provide insights that can enhance existing cross-border trade monitoring by employing machine learning (ML) techniques to illuminate patterns in intercompany cross-border trade. This analysis aims to identify potential intercompany cross-border risks, thereby benefiting Customs in improving its management of cross-border operations.

The research hypothesis for this study posits that cross-border intercompany trade is associated with an increased risk of undervalued export and import transactions, potentially resulting in revenue loss. In this context, ML algorithms can uncover trade patterns and behaviors in cross-border trade involving affiliated entities. Furthermore, the risk of undervalued transactions is further elucidated through patterns in affiliated transactions.

Insights generated from the analysis could serve as risk indicators, enriching existing risk management by helping customs authorities differentiate between monitoring and treatment for cross-border trade conducted by independent parties and affiliated entities. Consequently, this would lead to improvements in cross-border management strategies, enabling authorities to make informed decisions and develop comprehensive policies. These aspects are integral to the work of customs and excise authorities.

Within the context of international trade, while many studies have focused on utilizing data analytics to predict and finding the most accurate algorithms for trade prediction (Wohl & Kennedy, 2018; Dumor & Yao, 2019), as well as trends for particular commodities (Nummelin & Hänninen, 2016), smuggling (Kuo & Chou, 2021; Medel et al., 2015), and e-commerce (Chen, 2022; Zhuang et al., 2021; Alrumiah & Hadwan, 2021; Rukanova et al., 2019), there remains a research gap in exploring trade behavior and patterns among affiliated entities, particularly within the context of Indonesia.

Similarly, past research involving intercompany trade were focused on the performance comparison with independent firms (Guest & Sutherland, 2010), potential fraudulent affiliated party transactions (Ruan, et al., 2019), evasion behaviors of cross-border trade actors (Ferrantino et al. (2012), using lens from taxation.

To the best of our knowledge, there is currently no specific data analytics utilization to reveal insights from cross- border intercompany trade patterns in Indonesia and other developing countries. This study is, therefore, the first to delve into intercompany cross-border trade in Indonesia, a developing country, using ML algorithms.

Given this novelty, our study is poised to enrich findings and unveil new insights and risk posed on intercompany cross- border trade patterns in Indonesia. By shedding light on patterns and structures of intercompany trade, this paper enriches the knowledge base and paves the way for further exploration, fostering informed decision-making and facilitating advancements within customs and excise monitoring of cross-border trade in Indonesia. This holds particularly true to enrich current cross-border trade management conducted by Customs.

This manuscript begins by introducing the research context, followed by the problem formulation, where the study delves into the theoretical underpinnings of international trade and data analytics algorithms theory. The third section of the paper delineates the methodology and data sources used in this study.

Moving on to section 4, the outcomes of this study are presented, along with the insights discovered. These insights can potentially enrich the compliance monitoring conducted by customs authorities in Indonesia and other countries with similar economic and trade characteristics. Finally, the last part encapsulates the conclusions drawn, acknowledges limitations, discusses implications and recommendations, and outlines potential avenues for future research.

2. LITERATURE REVIEW

2.1. The Gravity Model of World Trade

The gravity equation is an economic model that estimates the volume of trade between two countries, where countries trade more when their Gross Domestic Products are larger and they are geographically closer, similar to how planets are attracted based on their size and distance (WTO (2012), Krugman et al. (2023)). The gravity equation is presented below.

$$T_{ij} = A \times Y_i \times Y_j / D_{ij}$$

Where A is the constant term, T_{ij} represents the trade flow between country i and country j, Y_i is country i's GDP, Y_j is country j's GDP, and D_{ij} is the geographical distance separating the two countries.

Nonetheless, according to Bikker (2009), the model has two significant imperfections: it lacks a clear theoretical derivation based on economic principles and does not account for the possibility of substitution between trade flows, particularly in the context of economic integration. These limitations suggest that the model, in its original formulation, may not fully capture the complexities of real- world trade dynamics.

Derived from such empirical findings, the extended gravity model incorporates a wider range of macro-economic variables, such as trade tariff, common language and cultural ties. The extended gravity model could be used as the baseline theory underlying the basic pattern between international trades.

2.2. Intercompany Trade

Intercompany trade occurs when there is exchange of goods and services between different units or subsidiaries of the same multinational corporation or enterprise (Bonturi & Fukasaku (1993). This type of trade involves transactions within the same corporate entity but across different geographical locations or national borders. In cases where transactions are not conducted at arm's length, it can result in undervaluing export/import goods, eroding the amount of state revenue.

Ample of past research has been conducted to explore intra-company trade. One of the studies was conducted by Folfas (2009) indicated that (1) change in trade structure and volume of intercompany trade resemble shifts in global trade patterns, and (2) while the proportion of service-related and non-trade transactions has been on the rise, trade of physical goods between related entities continues to constitute the largest portion of overall transaction volumes. One way to efficiently explore intercompany trade is through affiliated transactions.

2.3. Machine Learning Algorithms

2.3.1. Clustering

Clustering refers to unsupervised classification techniques which categorize input-only samples by grouping them together according to their similarities. One of the clustering methods is k-means clustering. The k-means clustering algorithm refers to a non-probabilistic technique in identifying clusters within a dataset, aiming to minimize within-cluster variance by assigning cluster labels to data points ((Lloyd (1982), Bonturi & Fukasaku (1993), Sugiyama (2016)).

The k-means algorithm remains one of the most popular and straightforward clustering algorithms in the field of data analysis due to its applicability for effective utilization in various clustering scenarios, its simple implementation, and minimal computational demands (Ikotun et al., 2023). Thus, it is effective to get initial patterns in cross border transactions.

2.3.2. Anomaly Detection

Detecting unique patterns can be done by utilizing isolation forest, that is, a novel anomaly detection algorithm based on the assumption that outliers are always rare and far from the center of the main cluster (Liu, et al., 2018). Isolation forest has more efficient and effective performance, in comparison to other outlier detection algorithms.

In general, the objective of anomaly detection is twofold. First, to detect and separate anomalies within a sample from the normal data. Second, to characterize an existing clean sample and utilize it to identify new patterns that appear from it. In this paper, the focus is on the latter scenario, assuming the presence of a distinct pattern among affiliated party cross-border transactions.

2.3.3. Association Rule Learning

Association rule learning is a rule-based machine learning technique that sheds light on a unique relation between variables in extensive databases (Han, 2023). The purpose of this unsupervised learning is to discover robust rules within databases by utilizing metrics of significance.

Association rules are typically used for market basket analysis, which involves identifying patterns in customer purchasing behavior. Items that are frequently bought together are considered to have an association.

This logic can also be applied to understanding the relationship between countries based on affiliation patterns. Indonesian entities with affiliations to multiple countries may share similar patterns in knowing which countries are most likely to appear together as counterparts for Indonesian companies.

2.3.4. Decision Tree Algorithm

A decision tree is a machine learning algorithm that structures decisions as a tree- like hierarchy. Each node represents a decision or test on a feature, each branch signifies an outcome of the test, and the tree's leaves represent predicted class labels (in classification) or numerical values (in regression) (Bishop, 2016).

The tree is structured sequentially by most important features to distinguish the label. Thus, apart from predictive purpose, the decision tree could also be used to define feature importance underlying the relationship between Indonesian taxpayers who have affiliations to cross border entities. Compared to other supervised machine learning methods, decision tree algorithms are popular due to their simplicity, interpretability, and ability to handle both categorical and numerical data.

3. RESEARCH METHODOLOGY

3.1. Research Design

This study employs a quantitative research design to address the research objectives. The authors used ML algorithms to explore the dataset and unveil insights to identify the risk of cross-border trade transactions conducted by affiliated entities between 2018 and 2022.

We utilized unsupervised learning algorithms to illuminate patterns from cross-border transactions to identify potential cross-border risk. Specifically, we applied clustering algorithms, anomaly detection, and association rules to unearth patterns, groups, and relationships within the dataset. Every discovered pattern could be associated with a profile that may vary in terms of potential risk.

We employed the K-Means Clustering algorithm to classify entities into predefined clusters (k), determined by the similarity of features to the centroids of these clusters. We used the elbow method to determine the optimal number of clusters created from the dataset. Subsequently, we conducted profiling of each cluster based on the general characteristics observed within it. We chose the K-Means Clustering algorithm over other clustering techniques due to its efficiency, simplicity, ease of implementation, interpretability, and scalability.

We then employed the isolation forest algorithm to identify anomalies within the dataset, thereby highlighting potential unusual risks and emerging threats that might arise in intercompany cross-border trade.

Meanwhile, we utilized association rules to reveal correlations. We applied the Apriori algorithm for several reasons. First, our dataset is available in a horizontal data format. Second, the Apriori algorithm is more straightforward, which facilitates the derivation of meaningful insights from the results (Krishnan et al., 2021). Third, it exhibits scalability and can efficiently handle large datasets, rendering it suitable for a variety of data sizes. Additionally, Apriori can work with various types of data, making it versatile for different domains.

3.2. Population and Sample

We utilized data from Indonesian corporate taxpayers who have affiliated parties, both domestically and overseas, sourced from the Indonesian Tax Authority. The dataset includes transactional details, trade values, partner countries, and entities' profiles.

We chose to use tax authority data for our dataset with the belief that it would allow for a more in-depth exploration of the complexities within the broader landscape of cross-border trade conducted by affiliated parties. As a result, the insights derived from this dataset could enhance cross-border trade risk management.

3.3. Data Collection

To gain insights into economic performance and trade activities, this study utilized a combination of macro and micro data. Macro-level data included various macroeconomic indicators such as the distance between countries, Gross Domestic Product figures, and Export and Import data.

At the micro level, data was sourced from the Indonesian Tax Authority, encompassing information about Indonesian entities that have reported affiliations. For overseas affiliations, the data is organized at the country level, with countries determined from reported addresses using Natural Language Processing techniques.

To certify the security and privacy of the collected data, we implemented stringent measures to ensure data confidentiality and sensitivity. The taxpayer data provided was in the form of country-level aggregates. We conducted the analysis using the Python programming language. However, due to Indonesian privacy and tax laws, sharing the raw data is not feasible.

4. RESULTS AND FINDINGS

4.1. Data Description

Figures 1 and 2 depict the proportions of cross-border trade conducted by Indonesian entities with their affiliated partners and the proportions of intercompany cross-border trade involving Indonesian entities from 2018 to 2022. Both figures exhibit a consistent pattern over the years, with a slight decrease in 2022.

Figure 1. Proportion of Cross- Border Intercompany Trade from 2018 to 2022

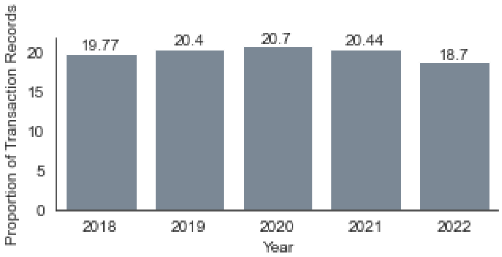


Figure 2. Proportion of Intercompany Trade in Indonesia from 2018 to 2022

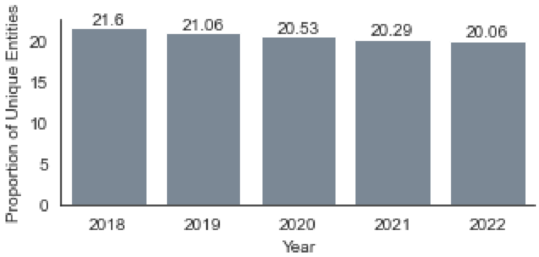
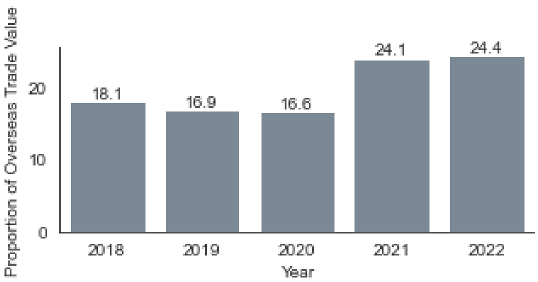


Figure 3 illustrates a significant increase in the total value of merchandise trade transactions in 2021 and 2022. This reinforces Indonesia's position as a leading country in merchandise trade.

Figure 3. Proportion of Overseas Trade Transaction Value



Before applying any profiling methods, we analyze the baseline trends among countries affiliated with Indonesia (Figure 4). It's worth noting that over half of these affiliations (55.85%) are primarily with Singapore, Japan, China, Malaysia, and Thailand.

The first profiling method is conducted descriptively as shown in Figure 5. The highest volume of cross-border trade was carried out by foreign investment companies in Indonesia, identified as both exporters and importers. They traded with their affiliates in overseas countries (FI:IMP:EXP:LN), accounting for 10%, and with low-tax jurisdictions as their counterparts (FI:IMP:EXP:LN LTR), amounting to 6%. The description of each attribute is presented in Appendix 1.

Figure 4. Proportion of Affiliated Transaction Based on Countries

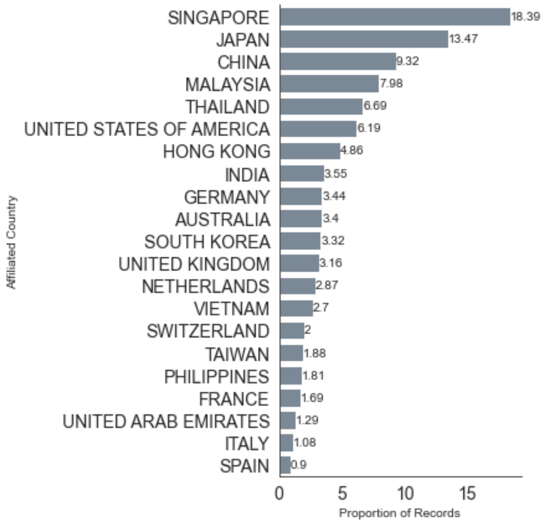
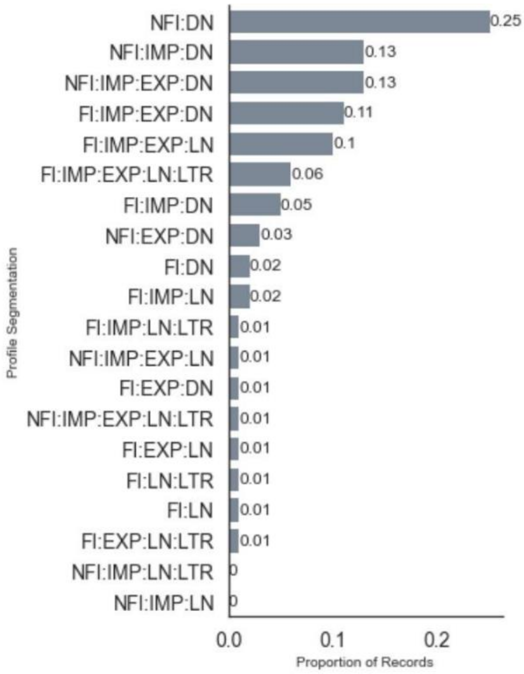


Figure 5. Trade Proportion Based on Related Entities Profile



We also identified whether any intercompany cross-border trade conducted by entities that were not indicated as exporters or importers. Our analysis revealed 2,647 transactions involving Indonesian entities that did not fall into either the exporter or importer category (Figure 6). The majority of these transactions comprised foreign investment- based companies with low-tax jurisdictions as their counterparts (FI:LN:LTR), totaling 1,157 trade records. The details of each attribute are presented in Appendix 1.

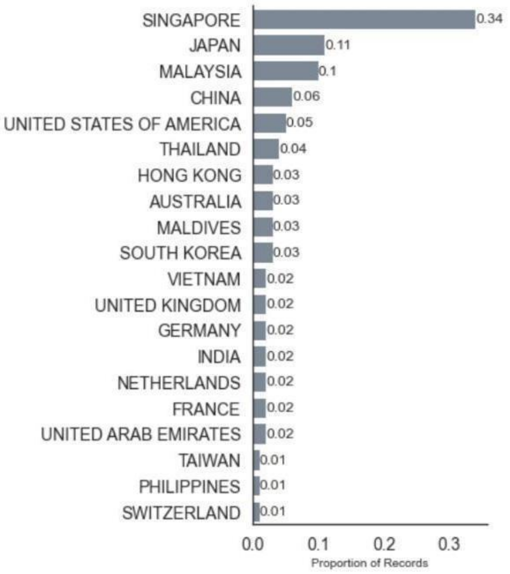
Figure 6. Cross-Border Transactions Records Conducted by Non-Exporters and Non-Importers

Profile_Segmentation	Total_Records
FI:LN:LTR	1157
FI:LN	861
NFI:LN:LTR	406
NFI:LN	223

Following this, we also analyzed the country partners of cross-border trade conducted by unidentified exporters and importers (Figure 7). It appears that the majority of unidentified exporters and importers engage in cross-border trades with Singapore (34%) and Japan (11%).

These discoveries shed light on the risks associated with illegal trading, in which Indonesian entities indicated as non- exporters and non-importers participate in export and import activities related to high- demand commodities. This discovery can serve as a valuable risk indicator for countries involved in cross-border trade, both as the source and destination, where such activities are conducted by non- exporters and non-importers.

Figure 7. Source/Destination Countries of Cross-Border Trade Conducted by Non-Exporters and Non-Importers



4.2. Correlation Analysis

We used correlation analysis to identify the factors that strengthen relationships between Indonesia and its overseas affiliated transactions. Figure 8 displays the total number of affiliated transaction records, including domestic and the most frequent overseas countries, revealing moderate correlations between America and Australia, China and Thailand, China and Vietnam, Vietnam and Taiwan.

Meanwhile, Figure 9 shows the result of similar correlation but for the top 20 overseas countries which have lower-tax- rates compared to Indonesia. It reveals a weak-to-moderate correlation between Vietnam and Taiwan, followed by Thailand and Vietnam, as well as Sweden and Denmark.

Figure 8. Correlation Analysis of the Top 20 Countries with the Largest Cross-Border Trade Records

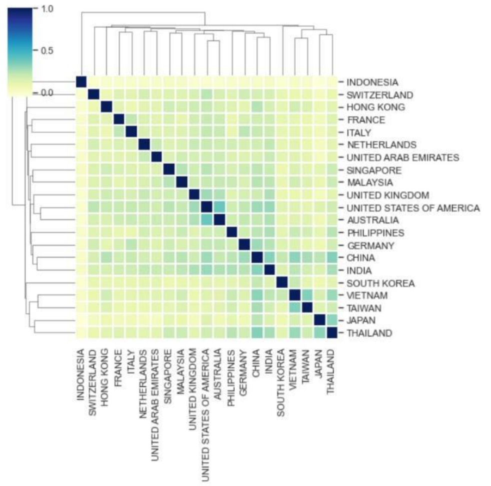
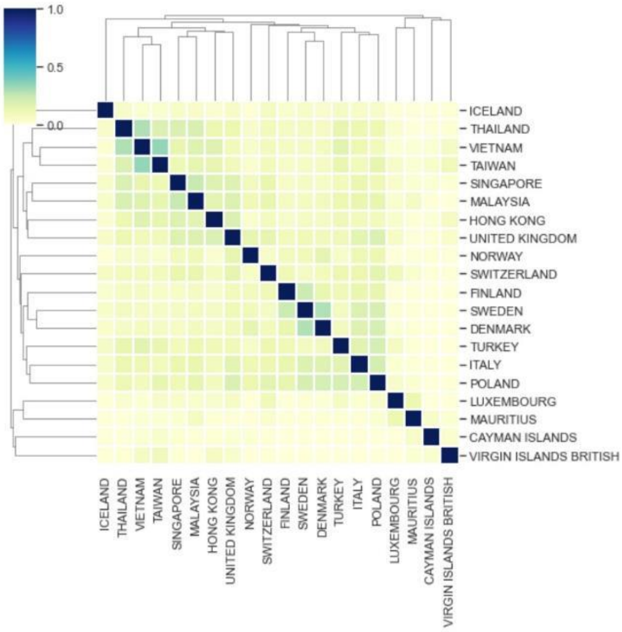


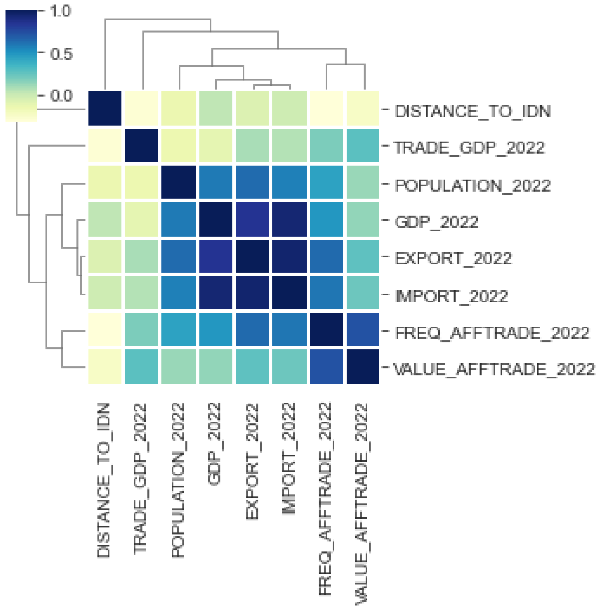
Figure 9. Correlation Analysis of the Top 20 Low Income Tax Rate Countries with Indonesia



This finding suggests that even though affiliated transactions are reported by entities assumed to be subject-wise, noticeable patterns emerge among the reported affiliate countries. These patterns may arise for various reasons, such as destination countries adhering to the gravity theorem or supporting transfer-pricing strategies.

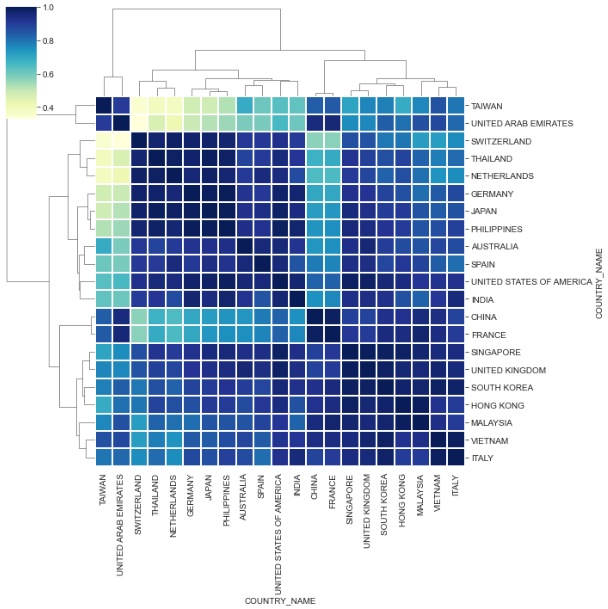
We therefore incorporated trade-related macroeconomic indicators to delve deeper into the relationship. As demonstrated in Figure 10, the extended gravity model is validated by the presence of a weak correlation with distance and strong correlations with Population, GDP, Export, Import, and Frequency of Affiliation Transactions.

Figure 10. Correlation Analysis with Macroeconomics Indicators



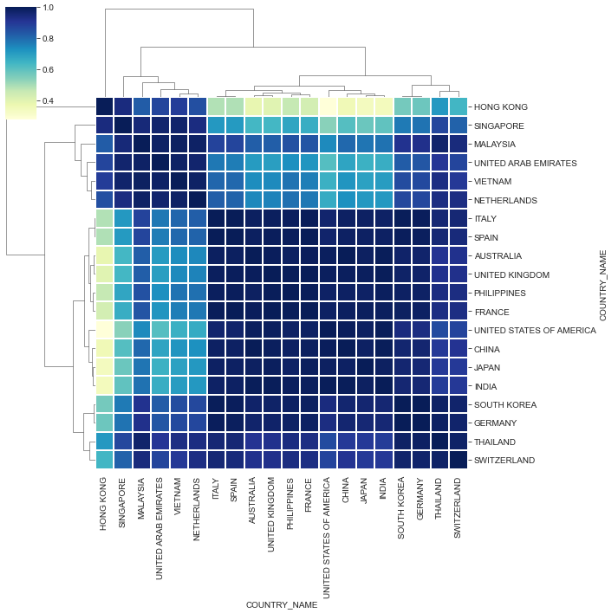
The marginal effect of affiliation transactions, including both the frequency of records and total transaction values, is evident when comparing Figure 11 and Figure 12. This comparison allows us to conclude that aggregate trade affiliation indicators strengthen the relationships between countries.

Figure 11. Correlation Analysis using Trade Affiliation Indicators 2018 - 2022



The macroeconomic indicators were unable to reveal the strong relationships between Hong Kong, Singapore, Malaysia, Vietnam, the Netherlands, and several other countries, as detected through the correlation analysis using aggregate trade affiliation indicators.

Figure 12. Correlation Analysis using Macroeconomic Indicators 2018 - 2022



The finding suggests that the volume of cross-border trade transactions is influenced not only by economic factors but also by affiliated transactions. Affiliated transactions may represent a new form of the extended gravity model. Using affiliation aggregate indicators can strengthen the relationships between countries compared to using transaction records directly.

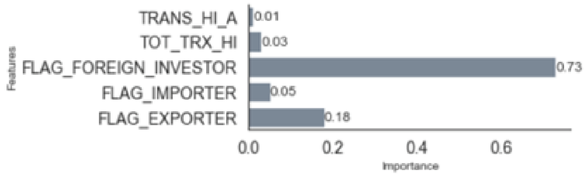
Countries that show stronger correlations in affiliation indicators, as opposed to macroeconomic indicators, are suspected to carry a higher risk. This suspicion arises from the fact that these affiliations themselves may signify the risk of transfer pricing, a form of tax evasion. This, in turn, could indicate the presence of undervalued trade goods in transactions between Indonesia and those countries.

4.3. Decision Tree Algorithm

The risk of tax evasion could be further revealed through affiliation transactions between Indonesia and countries with lower tax rates. Therefore, we utilized a decision tree algorithm to classify whether a transaction belongs to countries with lower tax rates or not. The selected attributes are specifically related to trade transactions, overall affiliation transactions, and entity profiles concerning cross-border trade. The focus here is more on feature importance than on prediction, aiming to discover which factors have the greatest contributions to the risk associated with cross-border transactions.

The model returns precision, recall, and F1 scores of 91%, 68%, and 75%, respectively. These metrics suggest that the model is reliable for drawing insights. As shown in Figure 13, our findings indicate that the most influential factors in determining cross-border transactions to low-tax jurisdictions are the entity's status as foreign investment-based (FLAG_FOREIGN_INVESTOR), followed by its status as an exporter (FLAG_EXPORTER).

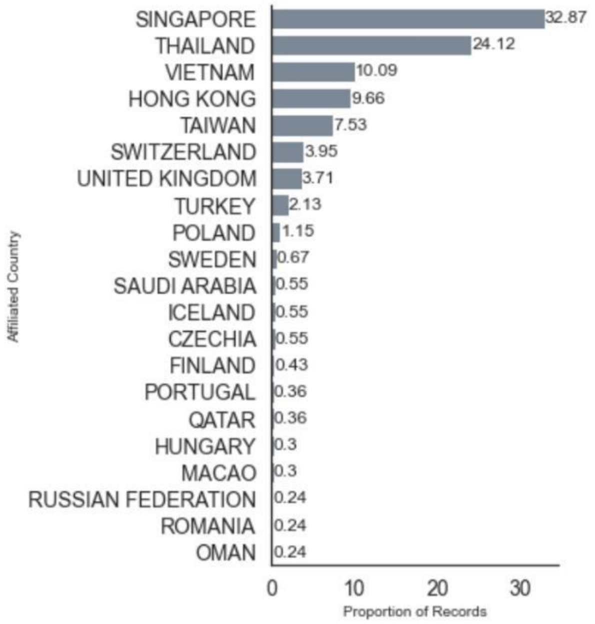
Figure 13. Feature Importance in Predicting Cross-Border Transaction between Countries



One scenario illustrating the risk associated with these two significant features is when an exporter attempts to declare a lower selling price of goods in the export declaration to reduce the export duty in the origin country and the import tax and export duty paid by its affiliate in the destination country.

Figure 14 lists the top 21 countries for which the model accurately predicts transactions. Transactions between Indonesia and these countries, mainly Singapore and Thailand, exhibit clear patterns associated with potential cross- border risks.

Figure 14. Correctly Predicted Transaction Based on Countries



4.4. K-Means Clustering

We conducted k-means clustering using the total transaction records for each entity based on affiliated countries. The elbow method indicated that the optimal number of clusters is four. Subsequently, we applied k-means clustering to the dataset, which resulted in the identification of four clusters with profiles displayed in Figure 15. The details of the attributes are provided in Appendix 2. The bar charts are proportional for each attribute based on clusters.

Figure 15. Cluster Profiles

ATTRIBUTES	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4
TOTAL_RECORDS	48%	14%	35%	3%
TOTAL_TAXPAYERS	81%	17%	2%	0%
TOTAL_FOREIGN_INVESTOR	88%	11%	1%	0%
TOTAL_IMPORTER	78%	19%	3%	0%
TOTAL_EXPORTER	77%	20%	3%	0%
TOTAL_DN	39%	11%	46%	4%
TOTAL_LN	75%	22%	3%	0%
TOTAL_LOWER_TAXRATE	73%	23%	4%	0%
PROPORTION_LTR_LN	24%	25%	27%	24%
PROPORTION_LTR_TOTAL	45%	49%	3%	3%
PROPORTION_LN_TOTAL	46%	48%	3%	3%
TOTAL_COUNTRIES	39%	31%	19%	10%
TOTAL_COUNTRIES_LTR	39%	32%	19%	9%
PROPORTION_COUNTRY_LTR	26%	27%	26%	22%
TOTAL_TRANS_A	3%	4%	2%	91%
TOTAL_TRANS_A_LN	52%	27%	17%	4%
TOTAL_TRANS_A_LTR	43%	28%	23%	6%
PROPORTION_TRANS_A_LN_TOTAL	50%	23%	27%	0%
PROPORTION_TRANS_A_LTR_LN	18%	23%	30%	29%
PROPORTION_TRANS_A_LTR_TOTAL	40%	23%	37%	0%

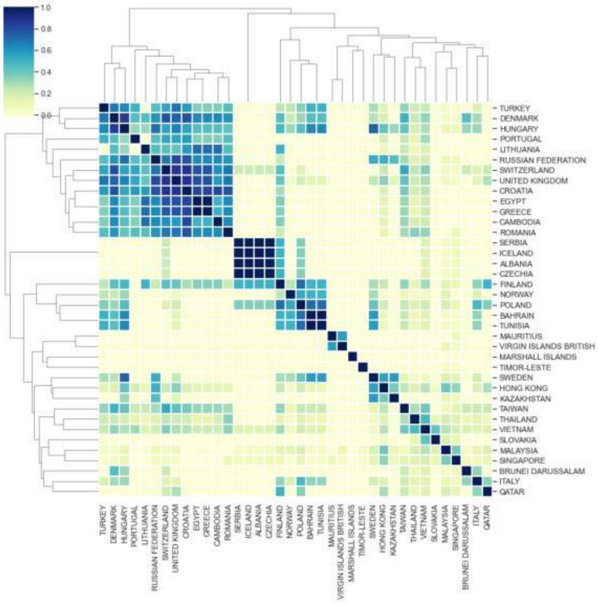
We interpret the profile for each cluster based on the attribute distributions as follow:

1) Cluster 1 primarily consists of medium- sized businesses with a significant number of entities. This cluster includes foreign investment-based companies, exporters, and importers engaged in frequent international trade, including transactions with low-tax-rate countries. Given that most transactions fall under this cluster, it is assumed to have a low potential risk.
2) Cluster 2 has an almost equal number of domestic and overseas transactions. Despite having a moderate number of entities, it engages in high-value overseas transactions, primarily with trade partners located in low-tax jurisdictions. This cluster is considered to have high potential risk.
3) Cluster 3 comprises large businesses with their primary market in Indonesia and significant transaction frequency for each entity. They engage in substantial cross-border trade, particularly to low- tax jurisdictions, with relatively high values. However, this cluster has fewer cross-border players in terms of entity profiles. It is considered to have medium risk.
4) Cluster 4 is composed of a small group of entities, mostly domestic investment- based, with Indonesia as their main market, and they engage in a massive amount of domestic trade. This cluster is considered to have low risk.

From the k-means clustering results, it can be inferred that Cluster 2 shows the highest risk compared to the rest. The risk is defined based on overseas transaction patterns, mainly to low-tax jurisdiction. Countries within this cluster and their relationships could be seen through correlation in Figure 16.

Countries that exhibit strong correlations, such as Turkey, Denmark, Hungary, Portugal, Lithuania, Russia, Switzerland, the United Kingdom, Croatia, Egypt, Greece, Cambodia, and Romania as well as tax haven countries such as Mauritius and the British Virgin Islands should be subject to closer monitoring due to the suspicious patterns they exhibit in terms of affiliation composition.

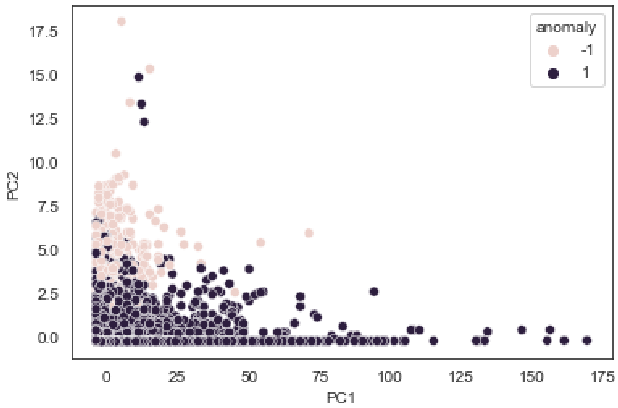
Figure 16. Correlation by Countries on Cluster 2



4.5. Isolation Forest

Anomaly detection using isolation forest equips clustering results with a closer look to abnormal patterns. Figure 17 visualizes how the distribution of anomalies (flagged as -1) differ from the normal ones through the plot of Principal Components 1 (PC1) and 2 (PC2).

Figure 17. Total Anomaly Data on Each Cluster



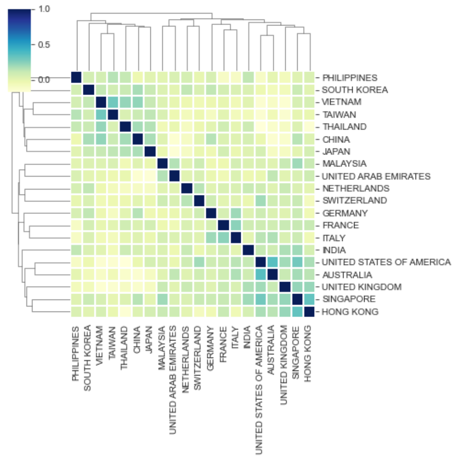
We compared the number of anomalies for each cluster. According to the result (Figure 17), the proportion of anomaly data on each cluster is relatively small, ranging between 1% and 3%, indicating the number of transactions within each cluster that are considered to have higher potential risk than others.

Figure 18. Total Anomaly Data on Each Cluster

CLUSTER	ANOMALY	NOT_ANOMALY	PROPORTION_ANOMALY
CLUSTER_1	508	57585	0.01
CLUSTER_2	14	1576	0.01
CLUSTER_3	184	10970	0.02
CLUSTER_4	4	136	0.03

There is no specific exploration to explain the anomalies profile, however, transaction patterns classified as anomalies can be observed through the correlations presented in Figure 19. Two groups of countries exhibit similar trends in their transaction records: the first group includes the United States, Singapore, and Australia, while the second group comprises China, Vietnam, Taiwan, and Thailand. It is suspected that transactions to these countries exhibit further potential cross- border risk.

Figure 19. Correlation by Countries on Anomaly Records



4.6. Association Rule Learning

Previous findings show how countries are correlated based on affiliation transactions from different perspectives. Association rule detects the co-occurrences between countries, specifically to distinguish which countries are most likely to appear together as affiliate counterparts for Indonesian multinational enterprises. We are using all transaction records including domestic affiliations.

Based on our findings, as illustrated in Figure 20, we can infer that the occurrence of affiliations between business groups with entities in Indonesia is highly correlated with overseas affiliations in Australia and the United States. A substantial number of entities with affiliations in the United States also reported affiliations with Singapore and China. Anomaly detection reveals some anomalies in the related patterns between the United States, Australia, and Singapore.

Figure 20. Association Analysis of Affiliation Transactions on the Top 20 Trade Countries including Indonesia

NO	ANTECEDENTS	CONSEQUENTS	LIFT
1	AUSTRALIA, INDONESIA	UNITED STATES OF AMERICA	3.569
2	UNITED STATES OF AMERICA	AUSTRALIA, INDONESIA	3.569
3	AUSTRALIA	UNITED STATES OF AMERICA, INDONESIA	3.548
4	UNITED STATES OF AMERICA, INDONESIA	AUSTRALIA	3.548
5	SINGAPORE, CHINA	UNITED STATES OF AMERICA	3.345
6	UNITED STATES OF AMERICA	SINGAPORE, CHINA	3.345
7	UNITED STATES OF AMERICA	AUSTRALIA	3.255
8	AUSTRALIA	UNITED STATES OF AMERICA	3.255
9	UNITED STATES OF AMERICA, SINGAPORE	CHINA	3.144
10	CHINA	UNITED STATES OF AMERICA, SINGAPORE	3.144
11	CHINA	JAPAN, SINGAPORE, INDONESIA	3.118
12	JAPAN, SINGAPORE, INDONESIA	CHINA	3.118
13	CHINA, INDONESIA	JAPAN, SINGAPORE	3.079
14	JAPAN, SINGAPORE	CHINA, INDONESIA	3.079
15	CHINA	UNITED STATES OF AMERICA, INDONESIA	2.840
16	UNITED STATES OF AMERICA, INDONESIA	CHINA	2.840
17	JAPAN, SINGAPORE	CHINA	2.781
18	CHINA	JAPAN, SINGAPORE	2.781
19	JAPAN, SINGAPORE	THAILAND	2.772
20	THAILAND	JAPAN, SINGAPORE	2.772

Such findings suggest that intercompany trades conducted by affiliated parties in Indonesia will show a higher frequency of trading with the aforementioned countries. Furthermore, these insights also corroborate the empirical findings of the extended gravity model, indicating that contemporary international trade is not solely dependent on the economic size and distance between trading partners.

Since corporate groups are under common administrative or financial control, the results from the association rules reveal the risks associated with undervaluing export or import transactions conducted by intercompany entities located in other jurisdictions. Indonesian companies engaging in cross-border trade with companies in the above-mentioned countries face a higher risk. In other words, transactions occurring between high-risk countries are more likely to involve undervaluing.

5. CONCLUSIONS

This paper has examined intercompany cross-border transaction patterns in Indonesia. Using machine learning algorithms, this study has uncovered insights that could serve as valuable risk indicators to enhance the current cross- border risk management strategies developed by Customs Authorities.

Our findings show that: 1) Foreign investment-based companies are more likely to conduct cross-border trade compared to domestic-based investment, indicating higher risk, 2) Intercompany cross-border trade follows the extended gravity model rule, 3) Some countries exhibit similar patterns in affiliation transactions, and each pattern could be associated with specific risk profiles, 4) Higher risk entity is indicated by the number of trade transactions with low-tax jurisdictions.

Leading countries in affiliation transactions include Singapore, Japan, China, Malaysia and Thailand. Singapore stands out in several ways: (i) It's a country with high overseas transactions conducted by non-foreign investor, non-exporter, and non-importer entities, followed by Japan, (ii) It's a low-tax-rate country with evident patterns that can be correctly predicted, followed by Thailand, (iii) Its correlations with other countries are stronger from a trade affiliation perspective compared to macroeconomics, along with Hong Kong, Malaysia, Vietnam, and the Netherlands, (iv) It contributes to anomalies, along with the United States and Australia and (v) It frequently appears in affiliated groups with Japan and China. Given its significant presence in various patterns along with other countries, Singapore may serve as the focal point of cross-border trade. Therefore, it could be a key focus for customs strategy.

In addition, Customs should also pay particular attention to countries with a potentially high risk of underreporting, especially those closely associated with Singapore, as well as countries classified as low-tax jurisdictions or tax havens

6. IMPLICATION AND LIMITATION

6.1. Limitation

Given its novelty, however, much remains to be done considering the study's limitations. Firstly, although this study used a mixture of macro-data and micro-level data, the micro-level dataset utilized is drawn solely from the Directorate General of Taxes (DGT). To deliver a more comprehensive view on the cross-border pattern and trade behavior, future research should consider incorporating micro data from Customs. Additionally, the scope of this study is confined to companies with foreign affiliations, thereby excluding a holistic representation of cross-border trade patterns in Indonesia. Consequently, the findings may not fully encompass the entire diversity of trade behaviors within the country.

Finally, this research solely focuses on the Indonesian context and hence could not be generalized to different economic context. These limitations highlight the potential for future research to contribute to the existing literature, thus providing a more comprehensive understanding of cross-border trade dynamics and behaviors.

6.2. Recommendation

Based on our findings, we recommend that policymakers should consider enriching the existing cross-border trade management both for risk assessment and operational strategy through data analytics techniques using macro and micro indicators, from a broader perspective.

6.3. Policy Implication

Policymakers should consider incorporating advanced technology to improve their risk management in cross- border trade, regardless of their current level of analytics maturity. The findings presented in this paper also emphasize the necessity of inter-agency collaboration, both within and across countries, to enhance the effectiveness of cross-border trade monitoring. Furthermore, these findings may also be relevant to other countries with similar economic characteristics to Indonesia.

Acknowledgement

We extend our sincere gratitude to the Compliance Risk Management and Data Science Unit in the Directorate of Tax Data and Information for their invaluable support throughout this research endeavor. We are particularly grateful to the support provided by the Data Science Unit, which greatly enriched our study.

REFERENCES

Kotter, J.P., (1997). *Leading Change: Menjadi Pionir Perubahan*, (Alih Bahasa, Joseph Bambang MS). PT Gramedia Pustaka Utama

Alrumiah, S. S., & Hadwan, M. (2021). Implementing Big Data Analytics in E-Commerce: Vendor and Customer View. IEEE Access, 9, 37281-37286. https://doi.org/10.1109/ACCESS.2021.3063615
In-text reference: (Alrumiah & Hadwan, 2021)

Bikker, Jacob. (2009). An extended gravity model with substitution applied to international trade. The Gravity Model in International Trade: Advances and Applications. 10.1017/CBO9780511762109.005.
In-text reference: (Bikker, 2009)

Bishop, C. M. (2016). Pattern recognition and machine learning. Springer New York.
In-text reference: (Bishop, 2016)

Bonturi, M., & Fukasaku, K. (1993). Globalisation and Intra-Firm Trade: An Empirical Note. OECD Economic Studies No. 20, Spring 1993.
In-text reference: (Bonturi & Fukasaku, 1993)

Chen, Q. (2022). Analysis and Prediction of Cross-Border E-Commerce Scale of China Based on the Machine Learning Model. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/7906135

In-text reference: (Chen, 2022)

Dumor, K., & Yao, L. (2019). *Estimating China's trade with its partner countries within the Belt and Road Initiative using neural network analysis*. Sustainability, 11(5), 1449.

In-text reference: (Dumor & Yao, 2019)

Ferrantino, M. J., Liu, X., & Wang, Z. (2012). Evasion behaviors of exporters and importers: Evidence from the U.S.–China trade data discrepancy. Journal of International Economics, 86(1), 141-157. https://doi.org/10.1016/j.jinteco.2011.08.006
In-text reference: (Ferrantino et al., 2012)

Fernández, Á., Bella, J., & Dorronsoro, J. R. (2022). *Supervised outlier detection for classification and regression*. Neurocomputing, 486, 77–92. https://doi.org/10.1016/j.neucom.2022.02.047 https://doi.org/10.1016/j.neucom.2022.02.047
In-text reference: (Fernández et al., 2022)

Folfas, P. (2009). *Intra-firm trade and non- trade intercompany transactions: Changes in volume and structure during 1990-2007.*
In-text reference: (Folfas, 2009)

Guest, P., & Sutherland, D. (2010). T*he impact of business group affiliation on performance: Evidence from China's 'national champions'*. Cambridge Journal of Economics, 34, 617-631. https://doi.org/10.1093/cje/bep017
In-text reference: (Guest & Sutherland, 2010)

Han, J. (2023). Data Mining Concepts and Techniques. Elsevier.
In-text reference: (Han, 2023)

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). *K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data*. Information Sciences, 622, 178-210. https://doi.org/10.1016/j.ins.2022.11.139
In-text reference: (Ikotun et al., 2023)

Kuo, Y.-H., & Chou, S.-C. (2021). *Manifest Monitoring Model as Support for Customs Risk Management: Evidence from Taiwan*. World Customs Journal, 15(2), 73-81.
In-text reference: (Kuo and Chou, 2021)

Krishnan, M. S., Nair, A. S., & Sebastian, J. (2021). *Comparative analysis of Apriori and Eclat algorithm for Frequent Itemset Data Mining.* Smart Innovation, Systems and Technologies, 489–497. https://doi.org/10.1007/978-981-16-3675-2_37
In-text reference: (Krishnan et. al.,2021)

Krugman, P. R., Obstfeld, M., & Melitz, M. J. (2023). *International Economics: Theory and policy*. Pearson.
In-text reference: (Krugman et al., 2023)

Liu F.T., Ting K.M., and Zhou Z. H. (2008). *Isolation forests*. In Proceedings of International Conference on Data Mining, 413–422.

In-text reference: (Liu et al., 2008)

Lloyd, S. P. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137.

In-text reference: (Lloyd, 1982)

Medel, M., Lu, Y., & Chow, E. (2015). Mexico's drug networks: Modeling the smuggling routes towards the United States. *Applied Geography*, *60*, 240–247. https://doi.org/10.1016/j.apgeog.2014.10.018
In-text reference: (Medel et al., 2015)

Nummelin, T., & Hänninen, R. (2016). Model for International Trade of Sawnwood Using Machine Learning Models. In Natural Resources Institute Finland (Luke) (Vol. 74). Natural Resources Institute Finland (Luke), Helsinki 2016.
In-text reference: (Nummelin & Hänninen, 2016)

OECD. (2022). *OECD Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations 2022*, OECD Publishing, Paris, https://doi.org/10.1787/0e655865-en.

In-text reference: (OECD, 2022)

Ruan, J., Yan, Z., Dong, B., Zheng, Q., & Qian, B. (2019). *Identifying suspicious groups of affiliated-transaction-based tax evasion in big data*. Information Sciences, 477, 508-532. https://doi.org/10.1016/j.ins.2018.11.008
In-text reference: (Ruan et al., 2019)

Rukanova, B. et al. (2019). *Value of Big Data Analytics for Customs Supervision in e-Commerce.* In: Lindgren, I., et al. Electronic Government. EGOV 2019. Lecture Notes in Computer Science(), vol 11685. Springer, Cham. https://doi.org/10.1007/978-3-030-27325-5_22
In-text reference: (Rukanova et al., 2019)

Squirrell, M. (2020). *How machine learning can automate the determination of the valuation of goods*. WCO News. https://mag.wcoomd.org/magazine/wco-news-91-february-2020/how-machine-learning-can-automate-the-determination-of-the-valuation-of- goods/
In-text reference: (Squirrell, 2020)

Sugiyama, M. (2016). Clustering. In M. Sugiyama (Ed.), Introduction to Statistical Machine Learning (pp. 447-456). Morgan Kaufmann. ISBN 9780128021217. https://doi.org/10.1016/B978-0-12-802121-7.00048-0.
In-text reference: (Sugiyama, 2016)

Wohl, I., & Kennedy, J. (2018). Neural network analysis of international trade. US International Trade Commission: Washington, DC, USA.
In-text reference: (Wohl & Kennedy, 2018)

WCO. (2021). The role of advanced technologies in cross-border trade: A customs perspective.
In-text reference: (WCO, 2021)

WTO, (2012). A Practical Guide to Trade Policy Analysis. https://www.wto.org/english/res_e/publications_e/wto_un ctad12_e.pdf
In-text reference: (WTO, 2012)

Zhuang, W., Wang, M., Nakamoto, I. & Jiang, M. (2021). Big Data Analytics in E-commerce for the U.S. and China Through Literature Reviewing. Journal of Systems Science and Information, 9(1), 16-44. https://doi.org/10.21078/JSSI-2021-016-29
In-text reference: (Zhuang et al.,2021)

APPENDIX 1

Attributes	Description
NFI	Non-Foreign Direct Investment-based company
FI	Foreign Direct Investment-based company
EXP	Export Activities
IMP	Import Activities
DN	Trade Origin/Destination: Domestic
LN	Trade Origin/Destination: Overseas
LTR	Trade Origin/Destination: Overseas, Low-tax Jurisdiction

APPENDIX 2

No	Attribute	Description
1	TOTAL_RECORDS	Total cross-border affiliated trade between 2018 and 2022
2	TOTAL_TAXPAYERS	Total Indonesian entities engaged in cross-border affiliated trade between 2018 and 2022
3	TOTAL_FOREIGN _INVESTOR	Total foreign investment company
4	TOTAL_IMPORTER	Total Indonesian entities identified as importers
5	TOTAL_EXPORTER	Total Indonesian entities identified as exporters
6	TOTAL_DN	Total intercompany transactions in Indonesia
7	TOTAL_LN	Total intercompany transactions overseas
8	TOTAL_LOWER _TAXRATE	Total cross-border intercompany transactions from or to lower tax rate jurisdiction
9	PROPORTION_LTR_LN	Percentage of cross-border intercompany transactions from or to lower tax rate jurisdictions
10	PROPORTION_LTR_TOTAL	Percentage of cross-border trade to low-tax jurisdictions as a proportion of total trade
11	PROPORTION_LN_TOTAL	Percentage of cross-border trade to total trade
12	TOTAL_COUNTRIES	Total partner countries
13	TOTAL_COUNTRIES_LTR	Total partner countries categorized as low tax jurisdictions
14	PROPORTION_COUNTRY_LTR	Proportion of cross-border intercompany trade to partner countries categorized as low tax jurisdictions
15	TOTAL_TRANS_A_LN(B)	Total export-import transactions
16	TOTAL_TRANS_A_LTR(B)	Total export-import transactions to low tax jurisdiction
17	PROPORTION_TRANS_A_LTR_LN	Percentage of cross-border intercompany trade to and from lower tax rate jurisdictions out of cross-border transactions
18	PROPORTION_TRANS_A_LTR_TOTAL	Percentage of cross-border intercompany trade to and from lower tax rate jurisdictions out of all transactions