

## Intelligent Trade Surveillance: Anomaly Detection in Ecuadorian Imports Using Data Mining

### Vigilancia Inteligente del Comercio Exterior: Detección de Anomalías en las Importaciones del Ecuador con Minería de Datos

Pablo Xavier Molina Narváez<sup>1</sup> <https://orcid.org/0009-0006-6506-1340>, Marcos Orellana<sup>1</sup> <https://orcid.org/0000-0002-3671-9362>, Juan-Fernando Lima<sup>1</sup> <https://orcid.org/0000-0003-3500-3968>, Jorge Luis Zambrano-Martínez<sup>1</sup> <https://orcid.org/0000-0002-5339-7860>

<sup>1</sup>Laboratorio de Investigación y Desarrollo en Informática (LIDI), Cuenca, Ecuador  
[pmolinamsn@es.uazuay.edu.ec](mailto:pmolinamsn@es.uazuay.edu.ec), [marore@uazuay.edu.ec](mailto:marore@uazuay.edu.ec),  
[flima@uazuay.edu.ec](mailto:flima@uazuay.edu.ec), [jorge.zambrano@uazuay.edu.ec](mailto:jorge.zambrano@uazuay.edu.ec)



Esta obra está bajo una licencia internacional  
Creative Commons Atribución-NoComercial 4.0

Sent: 2024/07/12

Accepted: 2024/10/02

Published: 2024/10/15

#### Abstract

This study employs advanced data mining techniques to comprehensively analyze Ecuador's import data spanning three decades, from 1990 to 2021. By leveraging the k-means clustering algorithm, we meticulously identified anomalies within tariff items. The study utilized a comprehensive dataset that encompassed tariff items, years, and critical variables such as import volume, CIF and FOB values, import cost, and trade attractiveness. The data mining model generated insightful reports that accurately pinpointed anomalous patterns in tariff items. These reports empower experts to delve into the underlying causes of these anomalies, enabling them to make well-informed decisions to optimize Ecuador's import strategies. Our research underscores the transformative potential of data mining in detecting import anomalies, providing valuable intelligence for the strategic management of Ecuador's foreign trade. The findings contribute significantly to the prevention of customs fraud and unfair trade practices, ultimately enhancing the competitiveness of the country's import sector.

**Keywords:** Anomalies, Clustering, Data mining, Foreign trade, K-Means.

**Summary:** Introduction, Related Works, Materials and Methods, Results, Discussion, Conclusions.

**How to cite:** Molina, P., Orellana, M., Lima, J.-F., Zambrano-Martínez, J. L. (2024). Intelligent Trade Surveillance: Anomaly Detection in Ecuadorian Imports Using Data Mining. *Revista Tecnológica - Espol*, 36(2), 12-24. Recuperado a partir de <https://rte.espol.edu.ec/index.php/tecnologica/articulo/view/1208>

## Resumen

Este estudio emplea técnicas avanzadas de minería de datos para realizar un análisis exhaustivo de los datos de importación de Ecuador que abarcan tres décadas, de 1990 a 2021. Al aprovechar el algoritmo de agrupamiento k-means, identificamos meticulosamente anomalías dentro de las partidas arancelarias. Se utilizó un conjunto de datos completo que abarca partidas arancelarias, años y variables clave como el volumen de importación, los valores CIF y FOB, el costo de importación y el atractivo comercial. El modelo de minería de datos generó informes reveladores que identificaron con precisión patrones anómalos en las partidas arancelarias. Estos informes permiten a los expertos profundizar en las causas subyacentes de estas anomalías, lo que les permite tomar decisiones bien informadas para optimizar las estrategias de importación de Ecuador. Nuestra investigación subraya el potencial transformador de la minería de datos para detectar anomalías en las importaciones, proporcionando inteligencia valiosa para la gestión estratégica del comercio exterior de Ecuador. Los hallazgos contribuyen significativamente a la prevención del fraude aduanero y las prácticas comerciales desleales, mejorando en última instancia la competitividad del sector importador del país.

**Palabras clave:** Anomalías, Clusterización, Comercio Exterior, K-Means, Minería de datos.

## Introduction

Many countries, such as Ecuador, have meticulously collected data on their export and import activities for years. These valuable resources, available on the Central Bank of Ecuador website, hold enormous potential for generating productive knowledge. However, their use has been limited. Regarding imports, which range from raw materials to technology and vehicles to supply national industries, decision-making has been based mainly on empirical approaches and isolated initiatives. This aspect has impeded more significant economic growth. (Quintana et al., 2021).

Imports play a crucial role in the Ecuadorian economy, as the country relies heavily on external products to meet domestic demand and support various industrial sectors. Analyzing these imports using data mining techniques allows us to identify patterns, trends, and relationships invaluable in strategic decision-making.

Knowledge and training are essential in a globalized world where competition intensifies. Innovative strategies and techniques are required to access and strengthen new markets. When applied to imports, data mining provides behavioral patterns that can serve as a basis for defining solid strategies. However, it is not enough to master economic theory or replicate successful techniques from other countries. The key is to develop original strategies deeply rooted in Ecuador's historical import data. Morales Zurita (2023) and Suárez y Amador (2009) state that information is critical to understanding the past and the present and, in a certain way, predicting the future.

A data mining process is essential for transforming a dataset into a complete and relevant one. This set must include detailed information about the imported products, their origin, quantities, prices, means of transport, ports of entry, and other relevant aspects.

In the information age, where the amount of available data grows exponentially, data mining has become an indispensable tool for its analysis and strategic use. This discipline ranges from simple graphical techniques to complex statistical methods, complemented by artificial

intelligence and machine learning algorithms. These allow for solving problems of automatic grouping, classification, value prediction, pattern detection, and attribute association, providing invaluable information for decision-making (López et al., 2023; Suárez y Amador, 2009).

The analysis of imports with data mining must require knowledge of statistics, machine learning algorithms, and data analysis tools (Al Ayub Ahmed et al., 2023). In addition, the quality and reliability of the data used are crucial to obtaining accurate results and conclusions. Ultimately, data mining offers unprecedented potential to analyze imports in Ecuador. By applying appropriate techniques and tools, companies and decision-makers can use this information to optimize their business strategies, strengthen their import processes, and ultimately boost the country's economic development.

This study, which focuses specifically on Ecuadorian imports, their granular analysis at the tariff item level, detecting anomalies in multiple variables, and using the gravity model to support decision-making, provides practical insights directly applicable to your work. Under this context, the document has been structured as follows: Section 2 presents the related works; Section 3 sets out the methodology and methods used; Section 4 details the results obtained; Section 5 presents the discussions; and finally, in the last section, conclusions and future work are offered.

### **Related Works**

Previous studies on anomaly detection using data mining are extensive and applied to different areas but not specifically to countries' imports and exports. Regarding the economy, data mining studies have focused more on electronic commerce and the other economic variables that have affected trade relations between countries. These studies have been based on the gravitational model.

The foreign trade study by González Argote y Ticona González (2019) analyzes a landlocked country's effects on foreign trade. They applied data mining, especially the clustering technique with the k-means algorithm and Partitioning Around Medoids (PAM), with information on international trade indicators from 188 countries over ten years. This aspect was done to detect whether the condition of being landlocked is a limiting factor in the commercial dynamics of countries. The study considers variables such as per capita income, institutional quality, economic integration, latitude, variable indicating Mediterranean identity, economically active population, land area, a natural resources indicator, export cost (USD per container), import cost (USD per container), time to export (days), time to import (days) (Djayeola y Fujs, 2018; The World Bank Group, 2024).

The results showed that in recent years, a small subset of landlocked countries, including Bolivia and Paraguay, have eased the restrictions landlocked countries impose on the costs and times of export and import. Countries were also identified as having other types of economic barriers despite having access to the sea. Therefore, they present international trade characteristics similar to, or even less favorable than, landlocked countries, such as Venezuela. These results are consistent with the relationships of trade variables proposed by the gravity model of international trade, where countries were identified that do not have access to the sea but improved trade relations with other countries thanks to the improvement in economic terms such as the Gross Domestic Product (GDP) of these countries. (González Argote y Ticona González, 2019).

Similarly, the study on anomaly detection has been applied in electronic commerce, as in a survey performed at York University, Canada, where the k-means cluster segmentation technique is used to investigate anomalies in electronic commerce transactions. Applying this technique generated several clusters, each including records of similar transactions. Using a mathematical method, they defined a threshold, and if the number of records in that group is greater than the threshold, that group is called normal; otherwise, it is labeled as an anomaly cluster. Then, they applied different classification algorithms to investigate each new group, such as Logistic Regression, Naive Bayes, Nbtree, and Radial Basis Functions (RBF) to evaluate each cluster. With this model, they identified approximately 2.2% of the transactions analyzed as anomalies (Tan et al., 2020).

In the study by Wohl and Kennedy (2018), they performed a Neural Network analysis of International Trade using a data set assembled for a gravity model of international trade. The objective was to predict trade relations between countries, so they used the following variables: bilateral trade, the distance between countries, the GDP of the exporter, and the GDP of the importer; dummy variables, indicating whether the countries share a language, a border, a colonial relationship, or a trade agreement; and country or country-year fixed effects as independent variables. They randomly divided the data into a training set and a test set, then used the data from the training set to create an Ordinary Least Squares (OLS) linear regression, a Poisson pseudo-maximum likelihood estimator, and a neural network, and then used the test data to measure how the different methods generalize to new data.

Furthermore, they compared the results with a reference model, a model with country-fixed effects, and a model with country-year fixed effects. They then compared the neural network predictions to actual trade between the United States and its major trading partners outside the sample period to examine whether trade between countries can be accurately predicted with limited data. Finally, they verified that neural networks can be efficiently used with a few economic, geographic, and historical variables to generate reasonably accurate predictions about international trade (Wohl & Kennedy, 2018).

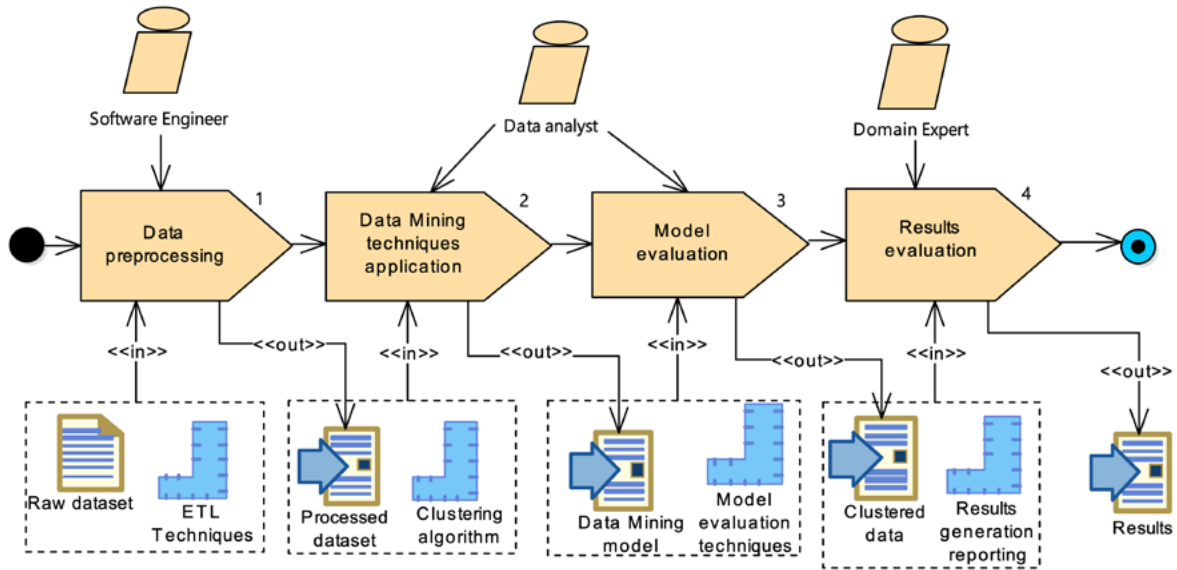
## Materials and Methods

This study presents a detailed analysis of imports made by Ecuador during the period from 1990 to 2021. To detect anomalies in import patterns, the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology has been applied in conjunction with the Software Process Engineering Metamodel (SPEM) scheme (Plotnikova et al., 2022). The data clustering technique has been used with the k-means algorithm on a data set of Ecuadorian imports obtained from the website of the Central Bank of Ecuador (Banco Central del Ecuador, 2023). The data set used in this study contains detailed information on Ecuadorian imports by tariff heading, year, and country. The variables included in the data set are imported tons, import, cost, insurance and freight (CIF) value, and accessible on board (FOB) value.

### The methodology used in the research

This research considers the following stages: i) data preprocessing, ii) data mining technique application, iii) model evaluation, and iv) results evaluation. Figure 1 shows the stages of the methodology.

**Figure 1**  
*SPEM Methodology*



**Data preprocessing**

The data set used to detect anomalies in Ecuadorian imports contains detailed data by tariff heading, year, and country with which Ecuador has maintained trade relations from 1990 to 2021. This data set contains 1,471,125 records and 43 variables.

Imported tons and the number of imports are mainly analyzed. Figure 2 shows the sustained growth over the years of imports in the number of tons, with only one drop recorded in 2016 and another in 2020, then a rebound in 2021.

**Figure 2**  
*Tons imported per year*

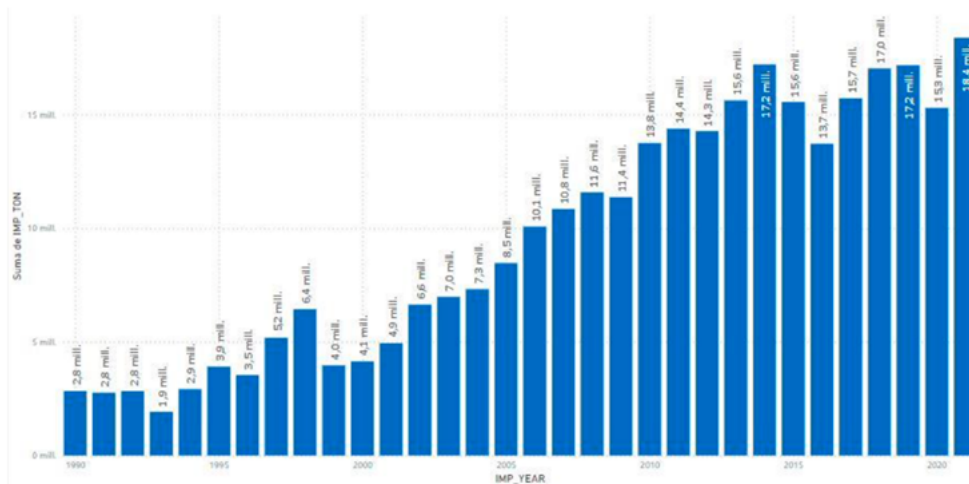
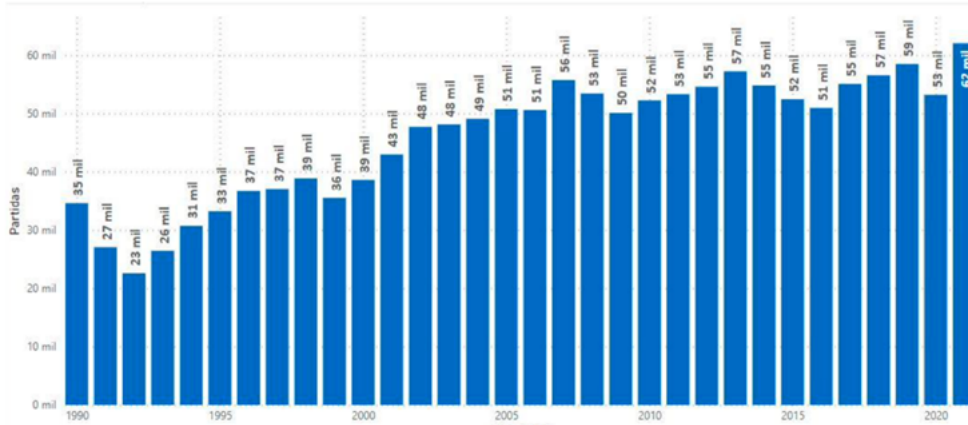


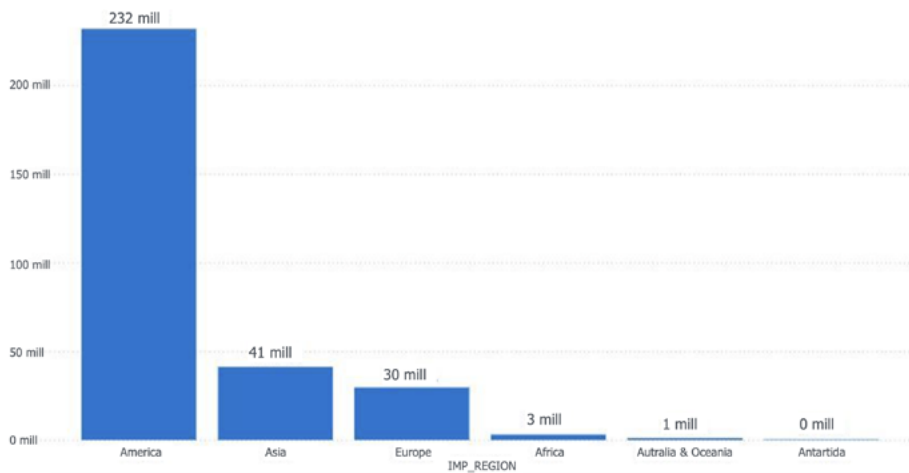
Figure 3 shows an increase in the number of imported items as the years go by and reflects a drop in the years 2015, 2016, and 2020, reaching a rebound in 2021, the year with the highest number of imports.

**Figure 3**  
*Number of imported items per year*



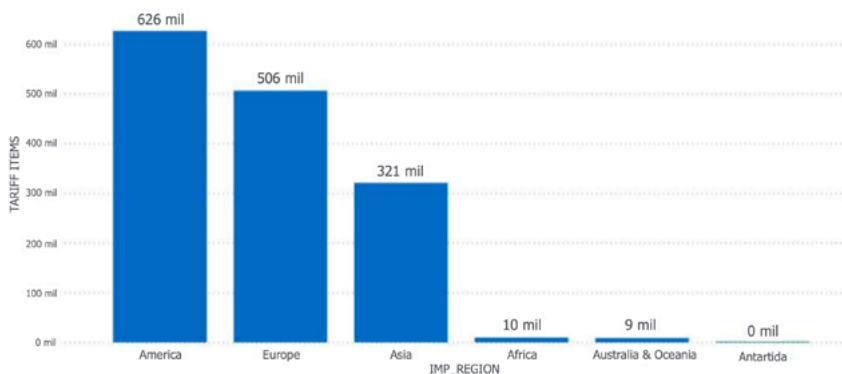
In Figure 4, in the countries of the American continent, we can observe that Ecuador imports a large number of tons, marking a significant difference from the rest of the continents.

**Figure 4**  
*Tons per region*



Unlike Figure 4, where there is a significant difference in the number of tons imported between the American continent and the rest of the regions, Figure 5 does not show a substantial difference in the analysis of the number of imported items by region.

**Figure 5**  
*Imported tariff items by region*





In addition, a statistical analysis was performed on each of the variables used in this study to verify their behavior, as seen in Table 1. For this, the variables were discretized by frequencies to obtain a more significant number of records in each range. Without this discretization operation, 99.99% of the data are in the first range. The leading cause that influences this behavior is abnormal values (outliers).

**Table 1**  
*Statistical values of the variables used*

MEASURES	VARIABLES					
	IMP_TON	IMP_FOB	IMP_CIF	IMP_FREIGHT_COST	TRADE FORCE ATTRACTION	COSTO_X_TONELADA
Count	1,471,123	1,471,123	1,471,123	1,471,123	1.31E+06	1,471,123
Mean	208.55	258,86	275.92	17.06	3.19E+19	26,155.89
STD	7,706.42	5,147.03	5,373.74	278.90	7.79E+19	15,425,580
Min	0.0000001	0.00	0.00	0.00	1.22E+14	0.00
25%	0.05	0.87	1.06	0.09	2.02E+18	0.25
50%	0.61	7.32	8.22	0.61	6.27E+18	0.72
75%	7.29	50.08	54.36	3.48	2.52E+19	3.07
Max	2,701,835	2,300,336	2,370,111	71,167.97	4.96E+20	17,964,870,000.00

The original dataset for this study was provided by the Laboratory for Research and Development in Informatics (LIDI) and obtained from the website of the Central Bank of Ecuador, which contained null data in the numerical fields, which were attributed using the Multiple Imputation by Chained Equations (MICE) algorithm (Wulff & Jeppesen, 2017). In addition, the normalization technique called Z value was implemented on the numerical data to apply the clustering technique.

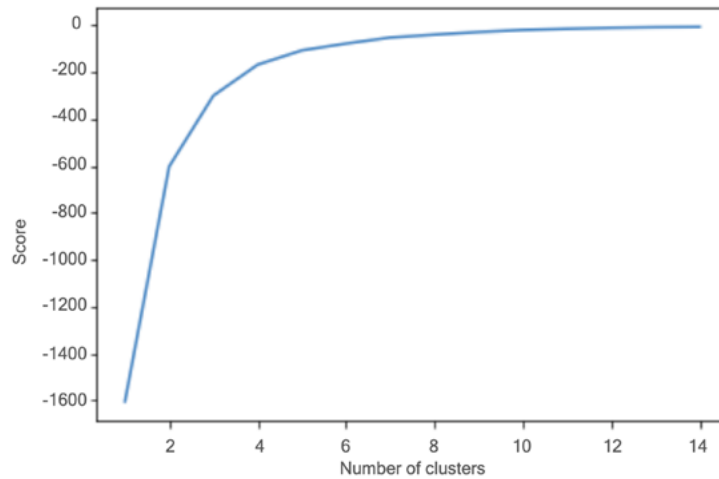
### *Apply data mining techniques*

In this project, the clustering technique was used to detect anomalies in imports from Ecuador. Clustering is an unsupervised machine-learning task. These algorithms involve using many unlabeled variables, and this algorithm groups records that contain similar values to each other (Orellana et al., 2022).

There are several clustering algorithms; the most well-known and used is k-means, based on centroids. It is a simple and efficient algorithm, so it is necessary first to define how many groups (k) the records will be divided into. The elbow method was applied to determine the number of clusters, as seen in Figure 6. The ideal values to define the number of groups are 3, 4, and 5.

The variables that have been used in this model to identify anomalies in Ecuadorian imports are IMP\_YEAR (import year), IMP\_TON (imported tons), IMP\_FOB (FOB value of the import), IMP\_CIF (CIF value of the import), IMP\_FREIGHT\_COST (import cost value), TRADE\_FORCE\_ATTRACTION (trade attraction force, calculated with the gravity equation), and COSTO\_X\_TONELADA. As a parameter, the system requests a tariff item level; when running the program, the model detects the anomalies of all tariff items per year at the required level. For the data preparation, they were normalized with the z-value algorithm and credited with the MICE algorithm to treat missing data.

**Figure 6**  
*Applied elbow method*



Once the data was ready, the k-means algorithm, a popular clustering algorithm, was applied. This algorithm groups the data into a specified number of clusters, in this case, three. The standard deviation value of all records was then calculated for all variables in the clustering algorithm. This step helps to identify the variability of the data within each cluster. Then, the average of the standard deviations of each variable was calculated, grouping them by year and by item code.

As for the averages of the standard deviations by year and by item, the values above and below three standard deviations of the centroids of each cluster were detected. This approach is based on statistical principles and helps identify extreme values significantly different from the average, labeling those records as anomalies for  $k = 3$ .

This procedure was repeated for 4 and 5 clusters, and anomalies were obtained for  $k = 3$ ,  $k = 4$ , and  $k = 5$ . Finally, it was possible to identify the anomalies in Ecuadorian imports, selecting only the records detected as anomalies in the three cases. The model generates tables, allowing the user to perform a complete analysis of imports at the tariff item level, year of import, and all variables with outliers.

### **Model Evaluation**

Validation metrics were used to evaluate the clustering result. The goal of clustering is to group similar objects in the same cluster and different objects in different clusters. On the other hand, the validation metrics are based on the criteria of cohesion, where objects in the same cluster are located as close to each other as possible; and separation, where clusters should be widely separated from each other, to measure the distances of the nearest clusters and the distance between the most distant clusters or the distance between centroids.

The Silhouette index is commonly used to evaluate the quality of the formed clusters, which quantifies how well the data has been grouped compared to how disjoint the clusters are from each other. In other words, the Silhouette index considers two main aspects of the clusters: cohesion and separation. A high and positive Silhouette index close to one indicates that the clusters are cohesive and well separated. A Silhouette index close to zero suggests overlap or that points are at or near the boundaries between clusters, and a negative Silhouette index close to -1 indicates that a point might be assigned to the wrong cluster.



Table 2 shows the values of the Silhouette index for  $k$  equal to 3, 4, and 5. 'k' represents the number of clusters. As can be seen, these are positive values close to 1, indicating that the clusters are cohesive and the groups are well separated.

**Table 2**  
*Statistical values of the variables used*

K	SILHOUETTE INDEX
3	0.987
4	0.974
5	0.964

### **Results Evaluation**

This data mining model generated five comprehensive reports that allowed for in-depth analysis of the results to be verified by experts. Each report provides valuable information for understanding import behavior and detecting potential anomalies.

The standard deviation values report details the year, tariff item, and standard deviation values for each variable used in the k-means algorithm. Additionally, identifying the cluster to which each record belongs is included. The outlier report presents the year, tariff item, and variables used in the k-means algorithm for each record, indicating whether the record is classified as typical or atypical. The label report is a copy of the original dataset but with an additional column displaying the cluster to which each record was assigned. This report allows you to analyze how the k-means algorithm grouped the imports, which is based on minimizing the within-cluster sum of squares and facilitates the identification of abnormal values in each dataset variable.

The final report presents the year, the tariff heading, and an indicator of whether the record was detected as an anomaly by the algorithm with three, four, and five clusters. In addition, a final column is included, showing the number of matches as an abnormal value. Records with the highest number of matches are those that require further analysis.

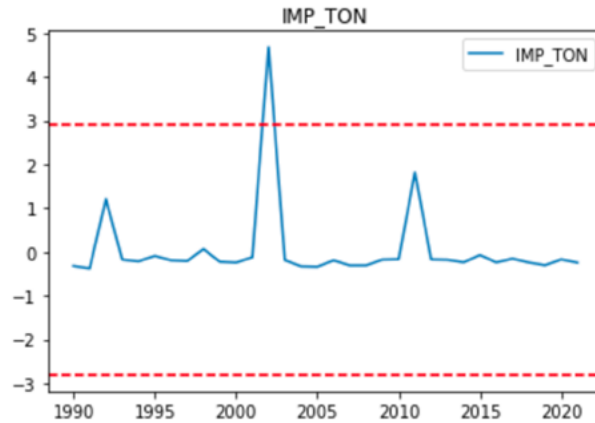
These generated reports offer a complete and detailed view of the behavior of Ecuadorian imports. Identifying outliers and anomalies allows competent authorities and entities to take timely measures, such as conducting further investigations, imposing penalties, or implementing stricter monitoring to investigate possible customs fraud, unfair trade practices, or other irregularities.

### **Results**

For the analysis of results, only the information obtained with the model applied to the level of one of the tariff items was considered, and the following results were obtained.

Figure 7 graphically displays the abnormal values of imports in tons of item 1, "Live animals," over time; those shown outside the red lines represent the centroid's third positive and negative standard deviation. In this example, imports in tons of live animals in 2002 contain an anomaly, and indeed, after reviewing the data set, it was verified that there was an import of 240.72 tons to New Zealand. The anomaly can be verified concerning the average tons of live animals in 2002, which is 7.15 tons.

**Figure 7**  
Abnormal values of imports in tons of heading 1 “Live animals,” three clusters



In Table 3, the item code and the year that was detected as anomalies for  $k=3$ ,  $k=4$ , and  $k=5$  are displayed; that is, they have three matches. From this table, item 87 “Motor vehicles, tractors, bicycles, and other land vehicles, their parts and accessories” was taken as an example to analyze it and find the anomalies in the original dataset. As can be seen, this item has anomalies in the years 1998 and 2021. The analysis was verified in the spreadsheet check tables generated by the model to confirm that they were indeed anomalies in these years.

**Table 3**  
Tariff items per year were detected as anomalies in the three cases.

TARIFF ITEM	YEAR	3 CLUSTERS	4 CLUSTERS	5 CLUSTERS	COINCIDENCES
6	2018	1	1	1	3
9	1996	1	1	1	3
15	1990	1	1	1	3
27	1991	1	1	1	3
32	1996	1	1	1	3
48	1998	1	1	1	3
61	2007	1	1	1	3
61	2008	1	1	1	3
63	2017	1	1	1	3
67	2007	1	1	1	3
79	1990	1	1	1	3
80	2005	1	1	1	3
82	2016	1	1	1	3
87	1998	1	1	1	3
87	2021	1	1	1	3
88	2008	1	1	1	3
90	2015	1	1	1	3

There are outliers in item 87 of Table 4 for the variables tons, FOB, and CIF between 1988 and 2021. For this study, only the variable of imported tons was analyzed. In this example, the results generated by the model with three clusters were verified. However, it should be considered that, in 1998, the clustering algorithm placed the anomalous record in cluster 3, and in 2021, it placed that record in cluster 2.

**Table 4**  
*Outliers by variable of item 87 between the years 1998 and 2021*

YEAR	IMP_TON	IMP_FOB	IMP_CIF	IMP_FREIGHT_COST	TRADE_FORCE_ATTRACTION	COSTO_X_TONELADA	CLUSTER	TARIFF ITEM
2021	Atypical	Atypical	Atypical	Atypical	Normal	Normal	Cluster 2	87
1998	Atypical	Atypical	Atypical	Atypical	Normal	Normal	Cluster 3	87

When analyzing the Ecuadorian import data set, it was found that in 1998, Ecuador imported 8,126.78 tons from Japan, when in that year, the average import of this item was 48.80 tons. In 2021, Ecuador imported 29,460.65 tons from China, while the average number of tons imported that year was 114.05 tons. With these analyzed examples, it was randomly verified that the model for detecting anomalies in Ecuadorian imports generates correct information. However, a more systematic verification method can be developed in future work.

For an import to be labeled as abnormal with greater certainty, it is when it has been detected as an anomaly for  $k = 3$ ,  $k = 4$ , and  $k = 5$ . Table 5 shows the number of cases detected by the model with one, two, and three matches.

**Table 5**  
*Anomalies by number of matches*

NUMBER OF MATCHES	NUMBER OF CASES	PERCENTAGE
1	478	83.71
2	76	13.31
3	17	2.98
Total	571	100

Similarly to the variables used for this study, it is observed that the variable with the most atypical values is the cost per ton, followed by the number of tons, as seen in Table 6.

**Table 6**  
*Number of atypical records per variable*

VARIABLE	NORMAL	ATYPICAL
COSTO_X_TONELADA	9,017.00	187.00
IMP_TON	9,030.00	175.00
IMP_FREIGHT_COST	9,040.00	164.00
IMP_CIF	9,046.00	158.00

VARIABLE	NORMAL	ATYPICAL
IMP_FOB	9,047.00	157.00
TRADE_FORCE_ ATTRACTION	9,101.00	103
Total	45,180.00	840.00

### Discussion

This study on detecting anomalies in Ecuadorian imports is particularly relevant to international trade. It distinguishes itself from previous research in data mining by its specific approach and detailed scope. Unlike other works that analyze general economic factors or e-commerce transactions, this study focuses on Ecuadorian imports over 32 years (1990-2021). This allows experts in the field to analyze the factors that influenced these changes, making it a valuable resource for future negotiations.

González Argote and Ticona González (2019) use the data mining clustering technique but with a different purpose: to identify the trade situation of landlocked countries, including various economic variables. This study only analyzes Ecuadorian imports with varying countries by tariff heading and year, detecting anomalies in variables such as imported tons, cost per ton, import cost, and CIF and FOB value.

The study by Wohl and Kennedy (2018) uses the gravity model to detect anomalies and support decision-making in future trade negotiations. The trade attraction force calculated with this model becomes a relevant indicator to assess the impact of trade relations between Ecuador and its trading partners.

The results of this study offer valuable information for Ecuadorian authorities responsible for trade policies. Identifying anomalies and understanding import trends at a granular level can contribute to i) investigating potential customs fraud or unfair trade practices, ii) optimizing trade negotiation strategies, iii) making informed decisions about supplier diversification and target market selection, and iv) strengthening the competitiveness of the Ecuadorian import sector.

### Conclusions

In the results of the anomaly detection, it is verified that the model shows precisely the item and the year in which the values of the variables are atypical due to the difference they have concerning the average of these variables in the same year of importation, fulfilling the main objective of this work. The present data mining model applies the clustering technique for 3, 4, and 5 clusters. At least in one of the cases, the variables exceed three standard deviations of their centroids; that item is labeled as an anomaly, and the atypical values are shown in a report. With this process, the tariff item and the year of importation are indicated precisely. The variables that present the most anomalies in this research are the cost per ton and the number of tons imported.

Regarding the analysis of statistical values, it is observed that atypical values exist in all variables. For example, analyzing the variable of imported tons, the minimum value is 0.0000001 tons, the maximum is 2,701,835.00 tons, and the average is 208.55 tons. This shows that 99.99% of the data is in the first range, and the remaining 0.01% is distributed in the other five remaining ranges.

This study demonstrates that the data mining clustering technique is efficient for detecting anomalies in a dataset that experts in various fields of study can use. As future work and as a

complement to this data mining model, this process can be further systematized, applying the same methodology but giving the user the option of searching for anomalies only in the required items and displaying the results on the screen. This enhancement could significantly reduce processing and search times while providing direct results on the screen without based search and the need for more specifics.

### Acknowledgments

This work was partially supported by the vice rectorate of Research at Universidad del Azuay for their financial and academic support and the entire staff in the Computer Science Research & Development Laboratory (LIDI).

### References

- Al Ayub Ahmed, A., Rajesh, S., Lohana, S., Ray, S., Maroor, J. P., & Naved, M. (2023). *Using Machine Learning and Data Mining to Evaluate Modern Financial Management Techniques* (pp. 249-257). [https://doi.org/10.1007/978-981-19-0108-9\\_26](https://doi.org/10.1007/978-981-19-0108-9_26)
- Banco Central del Ecuador. (2023). *Estadísticas de Comercio Exterior*.
- Djayeola, B. M., & Fujs, T. (2018). Policies, Technology, and Quality Returns from the World Development Indicators. *Statistika: Statistics & Economy Journal*, 98(4).
- González Argote, H. R., & Ticona Gonzáles, U. A. (2019). Clustering, mediterraneidad y comercio internacional: aplicación empírica de los algoritmos Partitioning Around Medoids y K-means. *Revista Latinoamericana de desarrollo económico*, 32, 95-129.
- López, D. S. M., Orellana, M., Tonon Ordóñez, L. B., & Zambrano-Martinez, J. L. (2023). Modelo Visual del Comercio Externo en Exportaciones Ecuatorianas. *Revista Tecnológica-ESPOL*, 35(2), 143-156.
- Morales Zurita, G. B. (2023). *La inflación y el comercio exterior agropecuario en el Ecuador*. Universidad Técnica de Ambato. Facultad de Contabilidad y Auditoría. Carrera~....
- Orellana, M., Acosta-Urigüen, M.-I., & García, R. R. (2022). *Implementation of Clustering Techniques to Data Obtained from a Memory Match Game Oriented to the Cognitive Function of Attention* (pp. 201-216). [https://doi.org/10.1007/978-3-031-18272-3\\_14](https://doi.org/10.1007/978-3-031-18272-3_14)
- Plotnikova, V., Dumas, M., & Milani, F. P. (2022). Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements. *Data & Knowledge Engineering*, 139, 102013. <https://doi.org/10.1016/j.datak.2022.102013>
- Quintana, R. A., Donoso, M. R., Kusactay, V., Chagerben, W. M., & Espinoza, J. B. (2021). *Introducción al Comercio Exterior*. Liveworkingeditorial.
- Suárez, Y. R., & Amador, A. D. (2009). Herramientas de minería de datos. *Revista Cubana de Ciencias Informáticas*, 3(3-4), 73-80.
- Tan, X. S., Yang, Z., Benlimane, Y., & Liu, E. (2020). Using Classification with K-means Clustering to Investigate Transaction Anomaly. *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 171-174.
- The World Bank Group. (2024). *World Development Indicators*.
- Wohl, I., & Kennedy, J. (2018). *Neural Network Analysis of International Trade*.
- Wulff, J. N., & Jeppesen, L. E. (2017). Multiple imputation by chained equations in praxis: guidelines and review. *Electronic Journal of Business Research Methods*, 15(1), 41-56.