



Facultad de Ciencias de la Administración

**Carrera de Ingeniería de Sistemas y
Telemática**

**CREACIÓN DE UN MODELO DE
VISUALIZACIÓN PARA EL ANÁLISIS DE LAS
IMPORTACIONES DEL ECUADOR A NIVEL
MUNDIAL, ENTRE LOS AÑOS 2008 Y 2018**

**Trabajo de titulación previo a la obtención del
grado de Ingeniero de Sistemas y Telemática**

Autor:

Pablo Santiago Carrasco Crespo.

Director:

Marcos Orellana Cordero.

Cuenca – Ecuador

2023

Índice de Contenidos, Figuras y Tablas

Índice de Contenidos

| | |
|--|-----|
| Índice de Contenidos, Figuras y Tablas | i |
| Índice de Contenidos | i |
| Índice de Figuras | ii |
| Índice de Tablas | iii |
| RESUMEN | iv |
| ABSTRACT | iv |
| 1. Introducción | 1 |
| 1.1 Objetivos | 1 |
| 1.1.1 Objetivo General | 1 |
| 1.1.2 Objetivos Específicos | 1 |
| 1.2 Marco teórico | 1 |
| 2. Revisión de literatura | 3 |
| 3. Métodos | 5 |
| 3.1 Comprensión del negocio | 6 |
| 3.2 Comprensión de datos | 6 |
| 3.3 Extracción de información | 7 |
| 3.4 Preprocesamiento de información | 8 |
| 3.5 Clasificación | 8 |
| 3.6 Creación de modelo de visualización de datos | 11 |
| 4. Resultados | 15 |
| 4.1 Modelo Visual | 15 |
| 5. Evaluación | 21 |
| 6. Discusión | 23 |
| 7. Conclusión | 24 |
| 8. Referencias | 25 |

Índice de Figuras

| | |
|---|----|
| Figura 1. Metodología CRISP - DM | 5 |
| Figura 2. Diagrama SPEM | 6 |
| Figura 3. Proceso Principal RapidMiner | 8 |
| Figura 4. Proceso Loop (2) RapidMiner | 10 |
| Figura 5. Proceso Loop RapidMiner | 10 |
| Figura 6. Proceso Branch RapidMiner | 11 |
| Figura 7. Pipeline de Visualización | 11 |
| Figura 8. Mapamundi | 12 |
| Figura 9. Radar | 13 |
| Figura 10. Mockup del modelo visual | 14 |
| Figura 11. Tablero 1 – Modelo Visual Países | 16 |
| Figura 12. Gráfica de barras horizontales - Países | 16 |
| Figura 13. Gráfica de líneas - Países | 17 |
| Figura 14. Tablero 2 – Modelo Visual Continentes | 17 |
| Figura 15. Mapa Mundial – Continentes | 18 |
| Figura 16. Gráfica de barras horizontales – Continentes | 18 |
| Figura 17. Gráfica de Líneas - Continentes | 18 |
| Figura 18. Tablero 3 – Modelo Visual | 19 |
| Figura 19. Gráfica de tipo árbol - Países | 20 |
| Figura 20. Gráfica de radar - Clústers | 20 |
| Figura 21. Resultados – Modelo TAM | 23 |

Índice de Tablas

| | |
|---|----|
| Tabla 1. - <i>Valores Seleccionados</i> | 9 |
| Tabla 2. - <i>Diccionario de datos</i> | 7 |
| Tabla 3. - <i>Tabla de Constructos</i> | 21 |

RESUMEN

Las transacciones de comercio exterior entre países se generan diariamente a una gran velocidad y con una gran cantidad de datos. Por ello, es importante contar con herramientas que sirvan de acompañamiento al experto en la toma de decisiones. El presente estudio genera un modelo visual para las importaciones del Ecuador desde los años 2008 y 2018. El modelo visual fue creado en base a la metodología CRISP-DM. Los resultados generarón una herramienta de visualización para las distintas áreas de interés vinculadas a la importación de mecacaderias, se consideraron variables como: toneladas importadas, valor FOB, valor PIB del país importador, entre otras. A través de la aplicación de un modelo TAM-BDA, se obtuvieron resultados que representan una valiosa aportación al área de estudio basado en la evaluación del modelo por parte de expertos.

Palabras clave: comercio exterior, importaciones, minería de datos, modelo visual, visualización de datos

ABSTRACT

Foreign trade transactions between countries are generated daily at high speeds and with a large amount of data. Therefore, it is important to have tools that help experts make strategic decisions. This study generated a visual model for imports from Ecuador from 2008 to 2018. The visual model was created based on the CRISP-DM methodology. The results generated are a visualization tool for different areas of interest related to import variables such as: imported tons, FOB value, GDP value of the importing country, among others. The results obtained through the application of a TAM-BDA model represent a valuable contribution to the study area based on the evaluation of the model by experts.

Keywords: data mining, data visualization, foreign trade, imports, visual model



Este certificado se encuentra en el repositorio digital de la Universidad del Azuay, para verificar su autenticidad escanee el código QR

1. Introducción

La digitalización de información genera nuevos datos, de gran variedad, a una gran velocidad. Esto se debe a que la mayoría de datos se almacenan en diferentes aplicaciones utilizando tecnologías modernas, tales como la nube (Jiang et al., 2018). La generación de la misma tiene varios orígenes, como redes sociales, Internet de las cosas (IoT), archivos ejecutivos, entre otros. Dicho esto, existe mucha información proveniente de varias áreas que se encuentra cargada en internet, la cual es accesible por la mayoría de la población. Un área de interés que diariamente genera información, de manera exponencial, es el comercio exterior de un país. Todas las transacciones comerciales que se realizan a nivel mundial son registradas en archivos para su respaldo. Existen varias aplicaciones que almacenan los mencionados datos del comercio exterior del Ecuador, por ejemplo: el Banco Mundial y Banco Central del Ecuador. Los mismos se encuentran clasificados, sea por país o producto. Aunque las mencionadas aplicaciones almacenan cada transacción realizada, la información no se encuentra agrupada de tal manera que se pueda crear un modelo visual, con el fin de generar conocimiento.

Debido al formato en que se almacena la información, los mismos contienen errores y pueden perjudicar los resultados finales. Para evitar esto, es importante aplicar técnicas de preprocesamiento de datos y minería de datos. Si bien, se han implementado distintas técnicas de minería de datos, la aplicación de clusterización genera una clasificación de la información previo al análisis con el fin de identificar patrones de comportamiento en los datos evaluados. La creación de un modelo visual, utilizando datos normalizados, facilita la construcción de visualizaciones intuitivas para el usuario final y el análisis de las mismas, con el objetivo de aportar información valiosa acerca del comercio exterior del Ecuador a nivel mundial entre los años 2008 y 2018. Un modelo visual brinda, a las diferentes áreas de interés, una nueva manera de evaluar el comportamiento del comercio exterior del Ecuador y tomar decisiones en base a datos reales e históricos, analizando los diferentes patrones que pueden existir (Wei & Liran, 2017).

El modelo visual generado en el presente estudio abarca todas las importaciones del Ecuador, para identificar diferentes patrones de comportamiento de las mismas durante el período establecido. Además, el mencionado modelo visual consolida la información que facilita la navegación de la misma a través del uso de esquemas de visualizaciones de datos simples e intuitivos. A diferencia de otros modelos creados como el de los autores Mao & Cheng, (2019), los usuarios cuentan con la posibilidad de manipular los filtros del modelo para depurar la información que desean visualizar. Los objetivos de este estudio son: identificar los distintos modelos de visualización existentes en la literatura actual aplicado al comercio exterior, con el fin de aportar una nueva perspectiva desde el punto de vista de las importaciones del Ecuador a nivel mundial, entre el período mencionado; la creación de un modelo que incluye los procesos aplicados de minería de datos a información acerca de las importaciones del Ecuador con el fin de extraer patrones en la misma. y la validación de los resultados obtenidos, en base a los criterios y opiniones de tres expertos en el área de estudio, aplicado a un modelo de aceptación tecnológica.

1.1 Objetivos

1.1.1 Objetivo General

Construir un modelo de visualizaciones de datos para reconocer los patrones de comportamiento de las importaciones del Ecuador a nivel mundial.

1.1.2 Objetivos Específicos

- Identificar los modelos de visualización existentes en la literatura actual aplicados a comercio exterior.
- Crear un modelo de visualización de datos, utilizando técnicas de minería de datos, aplicadas en base a información del comercio exterior del Ecuador.
- Validar los resultados obtenidos, por medio de la presentación del modelo creado.

1.2 Marco teórico

En el presente contexto, se analiza la rama de importaciones del Ecuador a nivel mundial, dentro de un rango de 10 años. Existen varios indicadores que reflejan el desempeño comercial de un país. Se podrían descomponer en tres indicadores principales que se analizarán en el sector de importaciones, las cuales son: un perfil general del comportamiento de comercio exterior del país, el desempeño histórico del comercio exterior del país, y la relación comercial que existe entre los países a analizar (International Trade Center,

2007). Es indispensable conocer los términos que son los pilares de la investigación. Algunos de estos términos son *Free on Board* (FOB), *Cost, Insurance, Freight* (CIF), Producto Interno Bruto (PIB), entre otros. Para comprender de mejor manera, primero se debe conocer sobre el comercio exterior.

El comercio exterior de un país es un indicador fundamental que forma parte del análisis el crecimiento y desarrollo económico del mismo. Además de demostrar su crecimiento económico o globalización, también es un indicador de la dependencia que existe entre un país y otro (Jiang et al., 2018). El comercio exterior abarca dos aspectos principales, las exportaciones o venta de productos y las importaciones o compra de bienes y servicios que realiza un país con otro. Dentro de las importaciones, existen algunos costos que son importantes de mencionar, tales como el FOB, CIF y PIB.

El FOB hace referencia a todos los costos que conlleva la ubicación de los productos en la nave para ser trasladados al país exportador. Por otro lado, el CIF abarca los costos que tienen el envío de los productos/servicios de un país exportador al país importador. Es decir, son los valores de transacción de bienes, costo de entregar los mismos en el país exportador y los costos de entregar los mismos desde la frontera del país exportador hasta la frontera del país importador. Así mismo, el PIB es el valor que mide la productividad de un país al evaluar todos los bienes y servicios generados por el mismo, dentro de un período de tiempo (Banco Central del Ecuador, 2017).

El comercio exterior se fundamenta en lo que se denomina un modelo económico, el cual tiene como objetivo principal, tomar información real acerca de la economía de un país, y posteriormente generar hipótesis acerca de su comportamiento. Existen diferentes modelos económicos que analizan el mencionado comportamiento económico de un país utilizando las mencionadas variables. Para este estudio, se utiliza el modelo de gravedad. El mismo es un modelo que al aplicarlo, determina el flujo de comercio dentro de un sector basado en el Producto Interno Bruto (PIB) del país exportador, PIB del país importador, una variable que afecte este comercio y la distancia entre los países que se están analizando (Zhou et al., 2020).

El flujo comercial de un país cuantifica la participación del mismo, en el comercio existente a nivel mundial. El mencionado flujo se representa a través de la recopilación de información, basado en las transacciones comerciales que realiza un país, con respecto a las ventas y compras de bienes o servicios a otros países. El modelo de gravedad toma esta información en el cual se aplica en la siguiente fórmula (Isard, W.1954):

$$T_{ij} = \frac{A * Y_i * Y_j}{D_{ij}}$$

T_{ij} = Flujo comercial entre país i y país j
 A = Constante
 Y_i = PIB del país i Y_j = PIB del país j
 D_{ij} = Distancia entre el país i y el país j

Al aplicar el modelo de gravedad, se estima el flujo comercial entre dos países, tomando a una variable que puede influir en la comercialización entre estos países, tales como el crecimiento de población, el crecimiento del PIB del país, entre otros. (Vazquez & Tonon, 2021)

Al momento de recopilar una gran cantidad de datos, es necesario facilitar a los usuarios finales la visualización de los patrones encontrados, para ello, es necesario estructurarlos a través de un modelo visual que presente los resultados a distintas áreas de interés. Un modelo visual consiste en la aplicación de diferentes procesos, como la extracción y recopilación de datos, utilización de técnicas de preprocesamiento para normalizar los mismos, entre otros, con el fin de obtener una visualización de datos que refleje los datos seleccionados (Islam & Jin, 2019). La visualización de datos crea visualizaciones intuitivas y fáciles de comprender por usuarios finales, a partir de datos coherentes, lo cual conlleva a que personas interesadas tomen decisiones en base a las visualizaciones presentadas. Una correcta presentación visual de datos tiene varios pasos a seguir para lograr lo mencionado. Dentro de los pasos está la colección de datos, regularización de los datos, filtración de datos y técnicas de *data mining*, para finalmente culminar con la representación visual de los datos (Shreyans Pathak, 2019).

El primer paso mencionado, la colección de datos, consiste en identificar las diferentes fuentes de información de las cuales se extraerán los datos a utilizar. El siguiente paso es la regularización de los datos extraídos. Para esto, se deben emplear procesos que eliminen valores o registros incoherentes ya que es posible que perjudiquen los resultados al momento de analizar los mismos. Algunos de los mencionados procesos es la limpieza de datos, que consiste en analizar los registros extraídos y encontrar registros

incompletos o incoherentes. Luego, se evalúa la manera de la cual se enfrentan a estos casos. Para registros incompletos, existen varias alternativas. Algunas de estas son, ignorar el registro en su totalidad, completar la información manualmente, utilizar una constante global que sustituya el dato incompleto, utilizar el valor promedio medida en base a los otros datos extraídos, entre otros. En el presente contexto, se utiliza una constante global para reemplazar a los registros que no se encuentran completos (Alasadi & Bhaya, 2017).

La recolección y regularización de datos debe ser la pista para encontrar patrones que determinan comportamientos futuros de las importaciones de datos de un país. Para ello, es necesario utilizar técnicas denominadas Data Mining. Algunas de estas técnicas incluyen la clasificación de información que categorizan los datos, con el fin de estructurar de mejor manera la misma, y la creación de clústeres que ayudan a encontrar similitudes entre datos (Osman, 2019). La creación de clústeres tiene como objetivo principal, dividir la información procesada en una cantidad específica de clases, con el fin de encontrar patrones de comportamiento. Por consiguiente, se consigue una clasificación de datos que tienen patrones similares, lo cual sirve como un pre análisis de la información utilizada. (Shen & Duan, 2020)

La aplicación de los mencionados pasos forma parte de una metodología que se sigue para alcanzar el objetivo propuesto. La metodología *Cross-Industry Standard Process for Data Mining* (CRISP-DM) consiste en seis pasos fundamentales para alcanzar el objetivo principal, que en este caso es la creación del modelo visual. La misma fue creada en 1996 y con la evolución tecnológica, se han realizado mejoras en los procesos, pero se mantienen seis pasos indispensables: comprensión del negocio, comprensión de datos, preparación de datos, creación del modelo, evaluación y despliegue, con el fin de cumplir un proceso de ciencia de los datos. (Shearer, 2000).

La etapa de evaluación de la metodología CRISP-DM fue implementado a través del uso de un modelo TAM. El modelo TAM fue creado para modelar la aceptación del usuario en la implementación de sistemas de información, en base a las intenciones de utilizar el sistema evaluado y al comportamiento del usuario (Verma et al., 2018). Sin embargo, el modelo TAM no incluye ciertos aspectos que son importantes en el presente estudio, como la calidad de información, entre otros. Debido a esto, se implementa el modelo TAM extendido (TAM-BDA), para evaluar la aceptación del modelo de visualización de datos, en base a distintos constructos adicionales al TAM, como la calidad del modelo, calidad de información, creencia en la implementación de un modelo innovador, confiabilidad de datos, entre otros. La aplicación del TAM-BDA aumenta las medidas que se evalúan por parte de los expertos del área, con el fin de comprobar la validez de la implementación del modelo de visualización de datos creado.

2. Revisión de literatura

El análisis del comercio exterior de un país es un tema que tiene mucho potencial en el aspecto de la generación de información valiosa para los gestores de comercio exterior, ya que se registra cada transacción comercial, es decir la compra o venta de bienes o servicios, que realiza el país con otro. Un correcto análisis e interpretación de esta información lleva a la creación de conocimiento. A continuación, se detallan algunas investigaciones que tratan sobre el comercio exterior, las cuales son coherentes para el desarrollo del presente trabajo.

Como lo menciona en el estudio llevado a cabo por Jiang et al. (2018), el análisis de comercio exterior que realizan los autores, es sobre la dependencia que puede existir en el comercio exterior de la provincia Yunnan, China. Dentro de la investigación mencionada, se tomaron datos del Banco Mundial, desde el año 2000 hasta el 2015. Jiang y los demás autores, aplican otro análisis que aporta un punto de vista distinto acerca del comercio exterior. En este caso, los autores analizan solamente la dependencia del comercio exterior de la provincia China, utilizando solamente datos de las exportaciones e importaciones que ha realizado la provincia. No obstante, la mencionada investigación visualiza de una manera básica el flujo de comercio exterior de un país basado en datos que fueron tomados del Banco Mundial. Aunque los mencionados autores analizan el comercio exterior de China, no se observa la aplicación de un modelo de gravedad.

También se evaluó la investigación realizada por Zhou et al., (2020), en la cual se implementó un modelo de gravedad para analizar la influencia que tienen las leyes conocidas como antiglobalización, utilizando datos acerca de las exportaciones de vegetales que realiza China, a nivel mundial, desde el año 2000 hasta 2019. Aunque Jiang y los demás autores utilizaron el mismo repositorio de información, Zhou et al. aplicaron el modelo de gravedad utilizando las diferentes barreras políticas implementadas como la constante dentro del modelo. A pesar de las diferentes variables que se utilizan y los datos recopilados, se

aplica el modelo de gravedad para evaluar la eficiencia del mismo, empleando datos reales acerca del comercio exterior de China. Aunque se hayan utilizado técnicas similares a la presente investigación, se diferencian por las fuentes de datos de las cuales se extrajeron los datos utilizados. Adicionalmente, se aplicó el análisis en otro contexto de información y no se implementó un modelo visual que represente la información acerca del comercio exterior de un país.

A diferencia de la anterior investigación, Boar et al., (2017) realizaron un estudio en el cual se recopiló información acerca del comercio exterior que existe en Europa entre los años 2008 a 2016, con el objetivo de aplicar técnicas complejas de análisis de redes para visualizar la mencionada información, de una mejor manera. Aunque no se implementó un modelo de gravedad como en la investigación llevada a cabo por Zhou y los demás autores, Boar y los demás crearon un modelo visual que ejemplifica las relaciones que existen dentro del comercio exterior en la Unión Europea. Además, se implementó el uso de una técnica de minería de datos llamada *clustering* para crear un análisis previo al resultado final, así, identificar patrones de comportamiento dentro del comercio exterior. Aunque se han aplicado varias técnicas en la presente investigación, la información recopilada no incluye información a nivel mundial. Sin embargo, las visualizaciones empleadas interpretan intuitivamente las importaciones que existen en el continente europeo. Adicionalmente, a la utilización de una mayor cantidad de información sobre el comercio exterior, se plantea crear un modelo visual que genere conocimiento a las distintas áreas de interés, con el fin de interpretar la información recopilada.

En la investigación llevada a cabo por Ye et al. (2020), se presentan algunas visualizaciones de datos que demuestran la relación de comercio que existe entre dos países principales, Estados Unidos y China. Los autores recopilaron datos de la base de datos de comercio exterior de la Unión Europea, desde el año 1992 hasta el 2018. Para ello, utilizaron una visualización de datos espacial, que informa al usuario sobre la historia del comercio entre los dos países mencionados, sobre varios productos en específico, a lo largo del rango de años establecidos. La investigación mencionada aporta información valiosa al crear visualizaciones intuitivas e innovadoras acerca del comercio exterior bilateral entre Estados Unidos y China. A diferencia de la mencionada investigación, el presente estudio crea un modelo visual que representa, además del comercio exterior de Estados Unidos y China, el comercio exterior a nivel mundial, tomando a Ecuador como su foco de interés principal. Además, se aplicará el método de gravedad para analizar patrones de comportamiento entre las importaciones del Ecuador a nivel mundial.

Otra investigación que aplicó un análisis similar fue la de Bernal et al. (2021), en la cual se realiza un análisis del posicionamiento de Brasil en el mercado agrónomo, en la sección de venta de frutas. Se utilizaron también datos del Banco Mundial. Además, varias técnicas de visualizaciones de datos fueron empleadas para representar esta información. Sin embargo, la cantidad de información que se analizó se limita a solamente una sección del comercio exterior, la compra y venta de frutas. Aporta información valiosa al demostrar cómo relacionar la información recopilada y presentarla en diferentes visualizaciones de datos que son de gran utilidad para áreas de interés. De manera similar, este estudio abarca la importación de productos, pero a diferencia Bernal, se enfoca la evaluación de todo tipo de productos con respecto al Ecuador, a nivel mundial.

En la investigación realizada por Huang (2014), se analizan las diferentes variables conocidas como factores de influencia en la aplicación del modelo de gravedad. En este caso, el autor toma la información de comercio exterior de China para aplicar el modelo de gravedad y evaluar los efectos que tienen otras variables que se pueden emplear, como la población del país, en el flujo comercial entre dos países. Es importante destacar la mencionada investigación, debido a que se observa la aplicación de diferentes variables, utilizando el modelo de gravedad. Además, el autor llega a la conclusión sobre la efectividad de la aplicación del modelo de gravedad para medir el flujo comercial entre países, lo cual se aplicará de la misma manera en el presente trabajo. Sin embargo, los datos que se analizarán son acerca del comercio exterior del Ecuador a nivel mundial, específicamente en las importaciones desde el 2008 hasta el 2018.

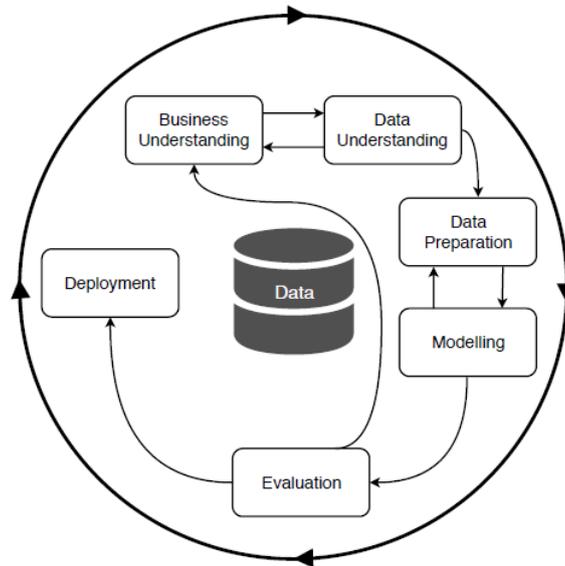
Existen algunas investigaciones con patrones similares en cuanto a la metodología y conjunto de datos, pero no existe ninguna investigación que utilice información que abarque las importaciones del Ecuador, el uso del modelo de gravedad y la aplicación de varios productos conjuntamente. En base a esta información recopilada, se crea un modelo visual que genere conocimiento valioso a varias áreas de interés como gobiernos, empresarios, entre otros. Además, se emplearon técnicas de minería de datos que normalizan la información que se extrajo del Banco Mundial y Banco Central del Ecuador.

3. Métodos

El presente trabajo utilizará la metodología *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Esta metodología se puede observar en Figura 1. La misma consiste en seis pasos que se deben seguir para procesar información recopilada y generar conocimiento (Martinez-Plumed et al., 2021):

Figura 1.

Metodología CRISP - DM



Nota: Esta figura muestra la metodología CRISP-DM.

Fuente: Basado en el autor Martinez-Plumed, 2021.

La metodología CRISP-DM se basa en una metodología llamada *Knowledge Discovery in Databases* (KDD). Como su nombre lo indica, KDD es un conjunto de pasos secuenciales que tiene como objetivo principal, el descubrir información valiosa a partir de grandes cantidades de datos, almacenadas en bases de datos. Estos pasos abarcan la selección de datos, preprocesamiento de los mismos, transformación de datos, minería de datos y finalmente la interpretación y evaluación, en este caso a través del modelo TAM (Gupta et al., 2020).

Inicialmente, se debe comprender el área de negocio que se está estudiando. En este caso, el primer paso es entender cómo funciona el comercio exterior de un país y tener claro el objetivo que se desea cumplir, el cual es la creación del modelo visual que abarque las importaciones del Ecuador entre el 2008 y 2018. Dentro del contexto de la presente investigación, se debe comprender en qué consisten las importaciones de un país y los diferentes factores que intervienen en el mismo, como el PIB, CIF y FOB. En segundo lugar, se procede a recopilar la información necesaria para evaluar en el estudio. Además, se analiza la información a utilizar y se verifica la calidad de la misma. En este paso, se lleva a cabo la extracción de datos de importaciones de diferentes repositorios de información como el Banco Mundial y Banco Central del Ecuador, en archivos planos. Los mismos cuentan con información sobre las importaciones del Ecuador a nivel mundial, durante el período de 2008 hasta el 2018. Los datos tomados consisten en la cantidad en toneladas de la importación, el valor del FOB de la importación, el valor CIF de la importación, el número de la partida arancelaria y descripción del producto o servicio que se está importando, la distancia a la que se encuentra el país del cual se importa y el PIB del país con el cual se está comercializando. El siguiente paso consiste en preparar la información recolectada. Es decir, aplicar técnicas de minería de datos que facilite la selección de información coherente e integrar la misma para formatear y crear un conjunto de datos relacionados a partir de los archivos planos mencionados. Para esto, se aplicaron técnicas de preprocesamiento de datos con el objetivo de eliminar valores nulos o incoherentes que puedan interferir con el análisis de la información. La siguiente fase es la creación del modelo. En este caso, se trata sobre la creación del modelo de visualizaciones de datos, tomando en cuenta los datos que se han seleccionado para ser evaluados. En esta fase, es importante probar diferentes modelos hasta llegar a una conclusión exitosa.

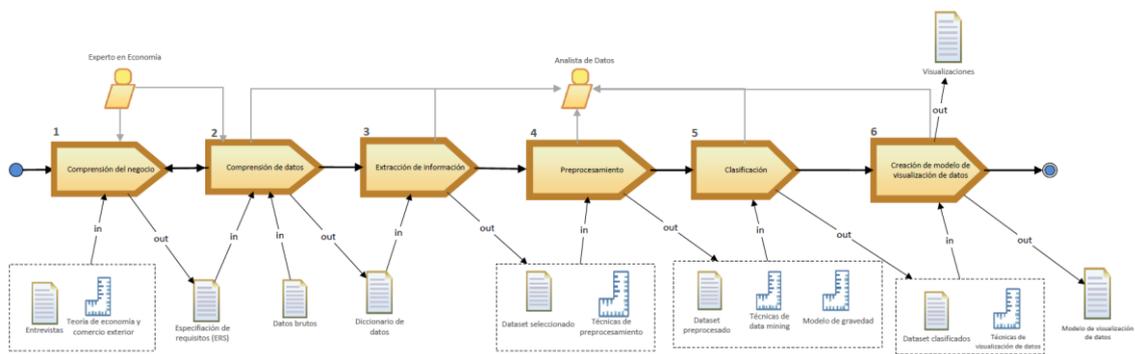
Dentro del modelo creado, se aplica el modelo de gravedad para evaluar el flujo comercial entre el Ecuador y los diferentes países a nivel mundial. Tras aplicar el modelo de gravedad, se procede a crear una visualización de datos que ejemplifique los datos y análisis creados. Además, se crearon diferentes objetos gráficos para generar conocimiento valioso. Finalmente, para la presente investigación, se procede a la última fase de la metodología, la cual es la validación y evaluación de los resultados obtenidos. Como la Figura 1. lo indica, se validan los resultados obtenidos a partir del modelo creado, en base a los datos recopilados. No obstante, existe la última fase que es la de despliegue. En la misma se comparte el conocimiento generado para aplicar a diferentes áreas de estudio. Sin embargo, no se llegará a esta fase dentro de la presente investigación.

El enfoque usado en la presente investigación es de carácter mixto, es decir cuantitativo y cualitativo, ya que se utilizan datos brutos acerca del comercio exterior para posteriormente analizar los mismos a través del modelo de visualización creado. Los datos manipulados en la presente investigación fueron extraídos del Banco Central y Banco Mundial, en formato de archivos planos. Los mismos consisten en datos históricos, entre los años 2008 hasta el 2018, acerca del comercio exterior del Ecuador a nivel mundial, tales como el PIB de los países comerciados, entre otros.

Para cumplir con la metodología establecida, se siguió un conjunto de pasos que ayudaron a obtener el resultado final. Estos pasos fueron representados a través de un diagrama SPEM, como se observa en la Figura 2. El mismo se utiliza para definir, de manera abstracta, procesos de desarrollo de software y de sistemas, y sus respectivos componentes que intervienen durante el desarrollo. (OMG, 2008)

Figura 2.

Diagrama SPEM



Nota: Esta figura detalla el proceso, representado a través de un diagrama SPEM.

3.1 Comprensión del negocio

El primer paso que se siguió es la comprensión del negocio. Dentro de esta etapa, a través de la participación de un experto en economía, se realizaron capacitaciones para alcanzar una mayor comprensión del negocio. Se trataron varios temas basados en teoría de economía, tales como el flujo en que consiste el comercio exterior y los diferentes factores que intervienen en el mismo, haciendo énfasis en las importaciones del Ecuador. Algunos temas que fueron tratados que son importantes dentro del análisis de las importaciones son el FOB, CIF, modelo de gravedad y PIB. Además, el experto indicó la importancia y el rol que tienen las partidas arancelarias dentro del comercio exterior entre países y la estructura de las transacciones, es decir en toneladas y por millones de dólares. Como resultado se obtuvo conocimiento acerca del área de estudio, lo cual es esencial para proceder con la creación de los distintos documentos de especificación de requisitos (ERS), para así continuar con el siguiente paso.

3.2 Comprensión de datos

El siguiente paso consiste en la comprensión de los datos. Con el conocimiento adquirido y los requisitos claros, se facilita la comprensión de los datos a analizar. Aquí se tomaron los datos crudos que fueron analizados. Los datos crudos hacen referencia a los conceptos de los datos que son relevantes para el caso de estudio, como los que se mencionaron anteriormente. Con la ayuda del experto de economía, se generó

el diccionario de datos, el cual se observa en la Tabla 2. Los primeros dos pasos trabajan en conjunto, debido a que a medida que se comprende el negocio, los datos que se deben analizar aumentan.

Tabla 1.

Diccionario de datos

| Campo | Tamaño | Tipo de Dato | Descripción |
|--------------------------------------|---------------|---------------------|--|
| Producto Interno Bruto (PIB) | 12 | Numérico | Valor del PIB del país del cual se importa. |
| <i>Free On Board</i> (FOB) | 12 | Numérico | Valor FOB de la importación registrada. |
| <i>Cost, Insurance Freight</i> (CIF) | 12 | Numérico | Valor de seguro, flete y demás costos que intervienen en la importación. |
| Distancia | 12 | Numérico | Distancia del país del cual se importa con respecto al Ecuador. |
| Año | 12 | Numérico | Valor que representa el año de la importación. |
| Toneladas | 12 | Numérico | Cantidad en miles de toneladas de la importación. |
| País | 50 | Caracter | País del cual se importa. |
| Nivel de partida | 12 | Numérico | Identificador que representa el nivel de partida de la importación. |
| Partida Arancelaria | 16 | Caracter | Campo que representa el número de partida arancelaria del producto/servicio importado. |

Nota: Esta tabla muestra el diccionario de datos generado.

Fuente: Elaborado por el autor

3.3 Extracción de información

Luego de comprender el negocio y los datos del mismo, se procede al tercer paso. Este consiste en la extracción de la información necesaria, utilizando el diccionario de datos como entrada. La misma se realizó en base a datos brutos cargados en el Banco Mundial y Banco Central del Ecuador, entre los años 2008 y 2018. Estos datos se originan al momento de que se realiza una transacción de comercio con otro país. En otras palabras, son registrados y cargados al sistema del Banco Mundial y Banco Central del Ecuador en archivos planos, respectivamente. Después de extraer la información, es necesario relacionar la misma y generar una estructura que facilite su análisis. Para esto, se cargó la información en una base de datos relacionada en distintas tablas. La base de datos se descompone en una tabla de continentes, una tabla de países, una tabla del valor PIB de los países con los cuales comercia el Ecuador, una tabla de las partidas arancelarias, una tabla de las importaciones y una tabla que contiene información acerca de la población de los diferentes países. La extracción de los datos se ejecutó en torno a los temas tratados anteriormente. Tras recuperar estos datos, se obtuvo el dataset de información no estructurada. Es decir, se recopilieron los datos de los archivos planos cargados en los repositorios de información para obtener como resultado el dataset con la información seleccionada.

3.4 Preprocesamiento de información

El dataset no estructurado se utilizó como entrada para el siguiente paso que es la aplicación de técnicas de preprocesamiento de información. Se aplicaron diferentes técnicas de preprocesamiento con el objetivo de normalizar los datos a analizar. Algunas técnicas implementadas de preprocesamiento fueron la limpieza de datos, reemplazo de información nula y reducción de dimensionalidad. La limpieza de datos se llevó a cabo al evaluar el dataset creado, revisando los registros irrelevantes que se extrajeron. Esto se realiza con el fin de identificar los mismos y tomar acción sobre ellos, sea la eliminación o modificación del mismo, para que no perjudique el resultado final. Dentro del dataset que se generó, existen varios casos en donde el Ecuador no comercializó con otro país dentro de cierto período. Estos datos fueron cargados sin valor y es necesario reemplazar los valores con cero, para que no alteren el resultado final. En otras ocasiones, se detectaron registros (19,481) en los cuales no existía el valor PIB del país del cual se importa. Dada esta condición, se reemplazaron los valores nulos por el valor 0.

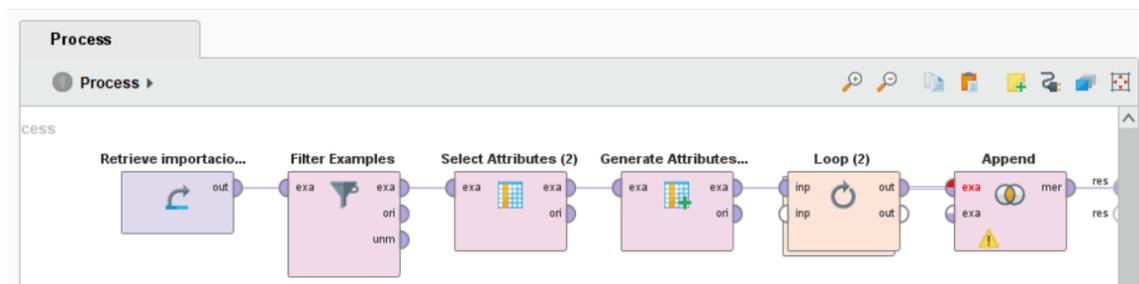
Otra técnica implementada fue la reducción de dimensionalidad de los datos. De igual manera, se filtró para casos en los cuales existan un mínimo de cincuenta registros de importaciones creadas, en base al nivel de la partida de la importación. Esto asegura que los resultados conseguidos posteriormente serán relevantes y más precisos. Además, debido a la gran cantidad de información acerca del comercio exterior archivada en las diferentes fuentes mencionadas, es indispensable reducir la cantidad de columnas que serán evaluadas y seleccionar solamente los valores que son de relevancia para la presente investigación. Para esto, se seleccionaron las siguientes columnas: el nombre del país del cual se importa, las toneladas de importación, el valor CIF de la importación, el valor FOB de la importación, el año de la importación, la partida arancelaria de la importación, el valor PIB del país del cual se importó, la descripción de la partida arancelaria, costos de importación, valor PIB del Ecuador, el modelo de gravedad, el precio por tonelada de importación, el costo por unidad de importación, el nivel de la partida arancelaria y la distancia a la cual se encuentra el país del cual se importa con respecto al Ecuador. Como salida del mencionado paso, se logró estructurar el dataset y normalizar, el cual fue utilizado como entrada para el siguiente paso a seguir.

3.5 Clasificación

El quinto paso a seguir fue la aplicación de minería de datos. El dataset creado se descompone en diferentes tablas. Las mismas son: una tabla que detalla la distancia que existe entre el país de referencia, Ecuador, y el país con el cual se comercializa, una tabla de todos los países con el que el Ecuador tiene registros de comercio dentro del período establecido. Una tabla de todos los productos que importa el Ecuador, dentro del rango de años establecidos y una tabla de las importaciones donde se detalla en diferentes columnas la partida arancelaria del producto que se transaccionó, el peso en toneladas de la cantidad de la transacción y distintos valores como el CIF, FOB, costo de flete, entre otros. Para la consolidación de la información recopilada, se utilizó el programa *RapidMiner*. Dentro del mencionado programa, se creó un flujo de datos para crear clústeres, según las variables importantes, tales como el precio por tonelada de importación, costos por unidad de importación y modelo de gravedad.

Figura 3.

Proceso Principal RapidMiner



Nota: Esta figura muestra los procesos principales utilizados en RapidMiner.

La Figura 3 muestra los pasos iniciales que se llevaron a cabo para la creación del flujo de datos mencionado. El flujo de datos creado consiste en la extracción de las importaciones, y posteriormente el filtrado de la misma, utilizando el proceso *Filter Examples*. Luego, a través del proceso

Select Attributes, se seleccionaron las variables que se utilizaron, como lo demuestra la Tabla 1. Además, se creó un código, usando el proceso *Generate Attributes*, en base al año que se está analizando, para identificar el clúster creado en base al año evaluado. Después, utilizando el proceso *Loop*, se repite el flujo dentro del proceso mencionado por 11 veces, debido a que son 11 años de comercio que se analizaron.

Tabla 2

Valores Seleccionados

| Campo | Tamaño | Tipo de Dato | Descripción |
|--------------------------------------|---------------|---------------------|--|
| Producto Interno Bruto (PIB) | 12 | Numérico | Valor del PIB del país del cual se importa. |
| Producto Interno Bruto (PIB) Ecuador | 12 | Numérico | Valor del PIB del Ecuador. |
| <i>Free On Board</i> (FOB) | 12 | Numérico | Valor FOB de la importación registrada. |
| Costos de importación | 12 | Numérico | Valor que representa los costos de la importación. |
| <i>Cost, Insurance Freight</i> (CIF) | 12 | Numérico | Valor de seguro, flete y demás costos que intervienen en la importación. |
| Distancia | 12 | Numérico | Distancia del país del cual se importa con respecto al Ecuador. |
| Año | 12 | Numérico | Valor que representa el año de la importación. |
| Toneladas | 12 | Numérico | Cantidad en miles de toneladas de la importación. |
| País | 50 | Carácter | País del cual se importa. |
| Nivel de partida | 12 | Numérico | Identificador que representa el nivel de partida de la importación. |
| Partida Arancelaria | 16 | Carácter | Campo que representa el número de partida arancelaria del producto/servicio importado. |
| Descripción de Partida Arancelaria | 200 | Caracter | Campo que hace referencia a la descripción completa de la partida arancelaria del producto o servicio importado. |
| Modelo de Gravedad | 50 | Numérico | Campo que consiste en el resultado de aplicar la ecuación de modelo de gravedad. |

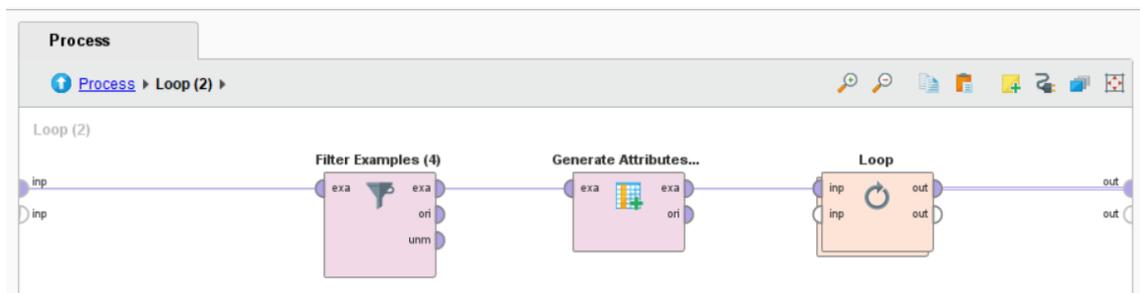
| | | | |
|---------------------|----|----------|---|
| Costos por unidad | 50 | Numérico | Valor de los costos de importación por unidad de cada importación. |
| Precio por tonelada | 50 | Numérico | Valor que hace referencia al precio por cada tonelada de importación. |

Nota: Esta tabla muestra los datos seleccionados para el análisis.

Fuente: Elaborado por el autor

Figura 4.

Proceso Loop (2) Rapidminer

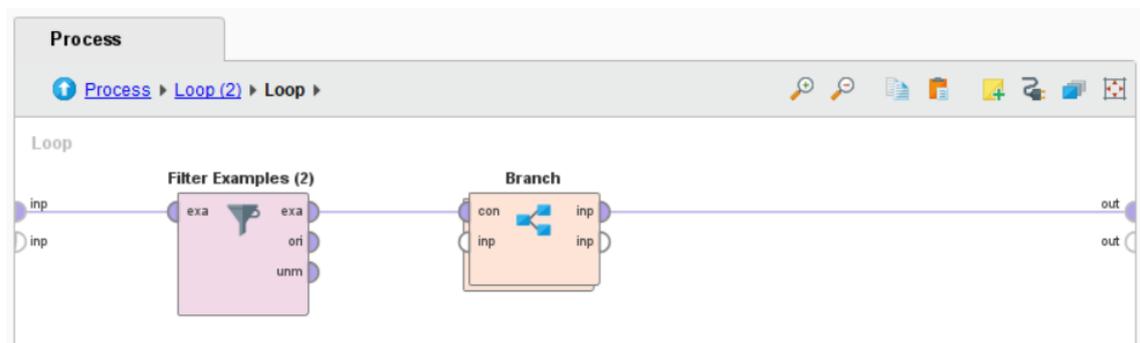


Nota: Esta figura muestra los procesos que componen el primer Loop implementado.

La Figura 4 muestra los procesos que se realizan dentro del proceso *Loop (2)*. En primer lugar, se filtró el año evaluado utilizando la variable de iteración, con la utilización del proceso *Filter Examples*. De igual manera, se generó un campo como identificador único, el cual está conformado por el identificador del país de la importación, el año de la importación y la partida arancelaria de la importación. Con los campos mencionados anteriormente, se iteran nuevamente a través de un proceso *Loop*, durante ocho veces. El mismo valor hace referencia a la cantidad de niveles de la partida que existe en las transacciones registradas.

Figura 5.

Proceso Loop RapidMiner

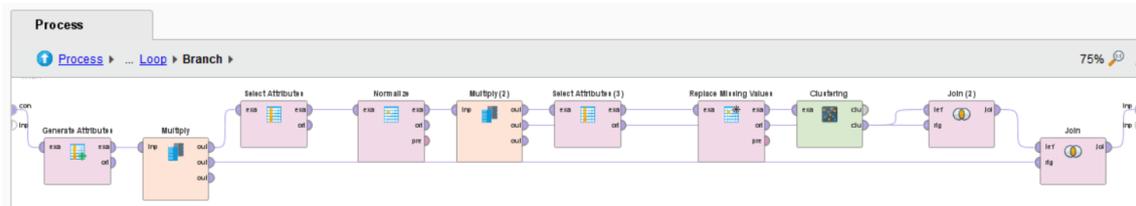


Nota: Esta figura los procesos que componen el segundo Loop implementado.

Como se observa en la Figura 5, se filtran los registros del dataset que cumplan la condición donde el nivel de la partida, es equivalente a la variable de iteración. Por medio del uso del proceso *Branch*, se crea una condición, la cual se debe cumplir para proceder con el análisis. La misma consiste en seleccionar los registros que contengan al menos 50 registros para el nivel evaluado. En el caso que se cumple, se procede a realizar los procesos dentro del *Branch*. Por otro lado, se descartan los registros que no cumplen esta condición, ya que pueden generar resultados incoherentes.

Figura 6.

Proceso Branch RapidMiner



Nota: Esta figura muestra los procesos implementados dentro del proceso Branch.

En la Figura 6 se observan los procesos que forman parte de la validación del proceso *Branch*. Con los atributos seleccionados y, en base al año y nivel de partida, se duplicó esta información para usarlo posteriormente. Después, se seleccionaron los valores que serán normalizados a través del proceso *Select Attributes*. Se normalizaron los valores numéricos: precio por tonelada, modelo de gravedad y costos por unidad, para crear clústeres en base a los mismos, con un rango establecido, siendo cero el valor mínimo y uno el valor máximo. Pevio a la creación de clústeres, se implementó el proceso *Replace Missing Values*, para evitar que se generen los clústeres con valores nulos, ya que no es posible crear clústers con valores nulos. Los valores nulos fueron reemplazados en este caso con un valor cero. A continuación, se procedió a generar los clústeres, utilizando los valores normalizados como entrada, mediante el proceso *Clustering*. Como salida, se obtienen los centroides de cada clúster y se unen a la información duplicada al inicio para así obtener el dataset completo, con el uso del proceso *Join*. Finalmente, como lo muestra la Figura 6, los resultados obtenidos se concatenaron con la creación de cada *subset* de datos generados para obtener el dataset clasificado, mediante el uso del proceso *Append*.

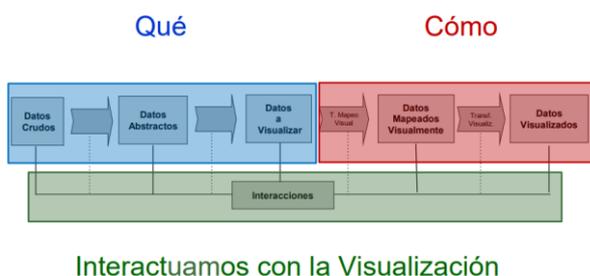
La creación de clústeres es un gran indicador como pre análisis de la información para observar los datos de una manera distinta en base a las diferentes agrupaciones de clústeres aplicadas, según el criterio implementado. En este caso, se analizó en base a las variables mencionadas, previa a la creación del modelo visual y se obtuvo el dataset clasificado.

3.6 Creación de modelo de visualización de datos

Finalmente, en base a los resultados obtenidos en el paso anterior, se creó el modelo de visualización de datos. La creación del modelo de visualización de datos tiene su inicio desde la extracción de información. La extracción de datos crudos es el inicio de un conjunto de pasos que siguieron para llegar a los resultados deseados. A continuación, se observa en la Figura 7, los pasos mencionados que forman parte del *pipeline* de visualización que se siguió (Castro, 2021).

Figura 7.

Pipeline de Visualización



Nota: Esta figura muestra el pipeline del modelo de visualización de datos.

Fuente: Basado en Silvia Castro, 2021

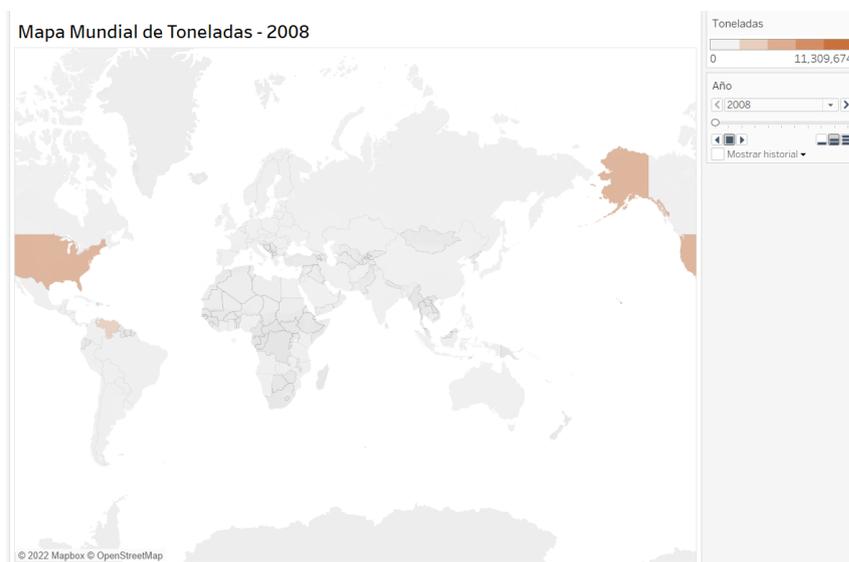
Tras obtener los datos crudos, se establecen los datos abstractos y se seleccionan los datos a visualizar. Las primeras tres interacciones mencionadas forman parte del *Qué* se desea lograr y en base a los datos seleccionados. Dentro de las primeras interacciones, es indispensable establecer el dominio y la audiencia

a la cual se desea llegar. En base a la audiencia, se establecen los datos que se visualizarán y las tareas que se llevarán a cabo para cumplir lo mencionado. En el presente caso, la audiencia son las diferentes áreas de interés, tales como empresarios, gobiernos, expertos de economía, entre otros. Para esto, se extrajeron los datos crudos de los dos repositorios mencionados anteriormente. Junto al experto de economía, se comprendieron los datos abstractos que se desean analizar a través de entrevistas y capacitaciones. Para conseguir los datos a visualizar, se realizaron tareas de preprocesamiento para normalizar los datos crudos. Con el dataset normalizado, se aplicaron técnicas de minería de datos como clusterización. Como resultado, se obtuvo el dataset clasificado, lo que se puede hacer referencia a los datos a visualizar. Las siguientes dos interacciones consisten en el *Cómo* se conseguirá el resultado final. Con los datos a analizar, se crea un mapeo visual de los mismos. Esto se realiza en base a la selección de la manera en la cual se codificarán los datos. En este caso, se realizó a través del uso de software como *Tableau* para representar la información de manera gráfica. Mediante el uso de la herramienta mencionada, se tomó el dataset clasificado y se implementó el esquema de visualización de datos geo-espaciales y espacio-temporales. Este esquema de visualización toma valores y los analiza a nivel geográfico. Además, se implementó el uso de coropletas en un mapa mundial, aplicando una paleta de colores únicos para facilitar la identificación de clústeres en el mapa. La aplicación de coropletas en dos dimensiones se basa en utilizar la misma gama cromática de color, donde la intensidad de color indica que el área coloreada hace referencia a un valor que se está analizando. Este puede ser el valor FOB, valor CIF, valor de la importación en toneladas, entre otras variables. Asimismo, con el uso de un esquema espacio-temporal, se observa el comportamiento del comercio de los distintos países de los cuales el Ecuador importa bienes y servicios, a lo largo de los años establecidos. También se implementaron visualizaciones de gráficas bidimensionales de líneas con el esquema espacio-temporal, para evaluar el comportamiento de las importaciones del Ecuador con respecto al período seleccionado. Adicionalmente, se crearon visualizaciones de tipo radar, basado en coordenadas radiales y de tipo árbol, basado en un orden jerárquico para representar las importaciones,

Las Figuras 8 y 9 muestran algunos de los elementos visuales creados para apoyar las tareas planteadas en un inicio.

Figura 8.

Mapamundi

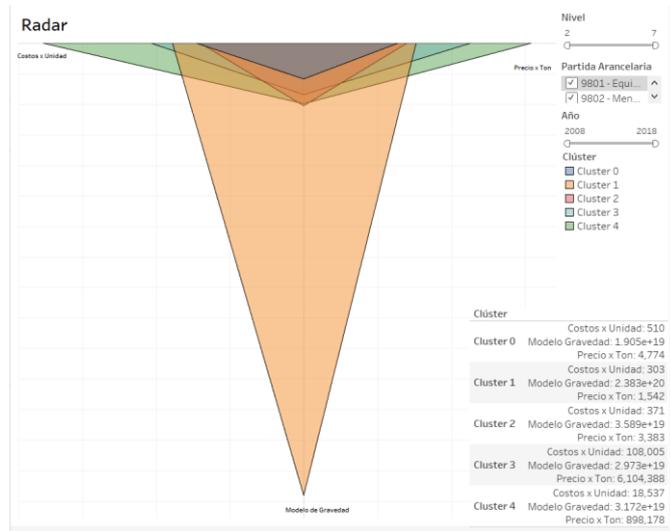


Nota: Esta figura muestra el ejemplo de la implementación de un mapamundi.

La Figura 8 muestra un ejemplo de una visualización en torno a un mapa geográfico, en el cual se observa los países de los cuales importa el Ecuador, evaluando la cantidad de toneladas importaciones para un año y una partida arancelaria en específico.

Figura 9.

Radar



Nota: Esta grafica muestra la implementación de una gráfica de radar.

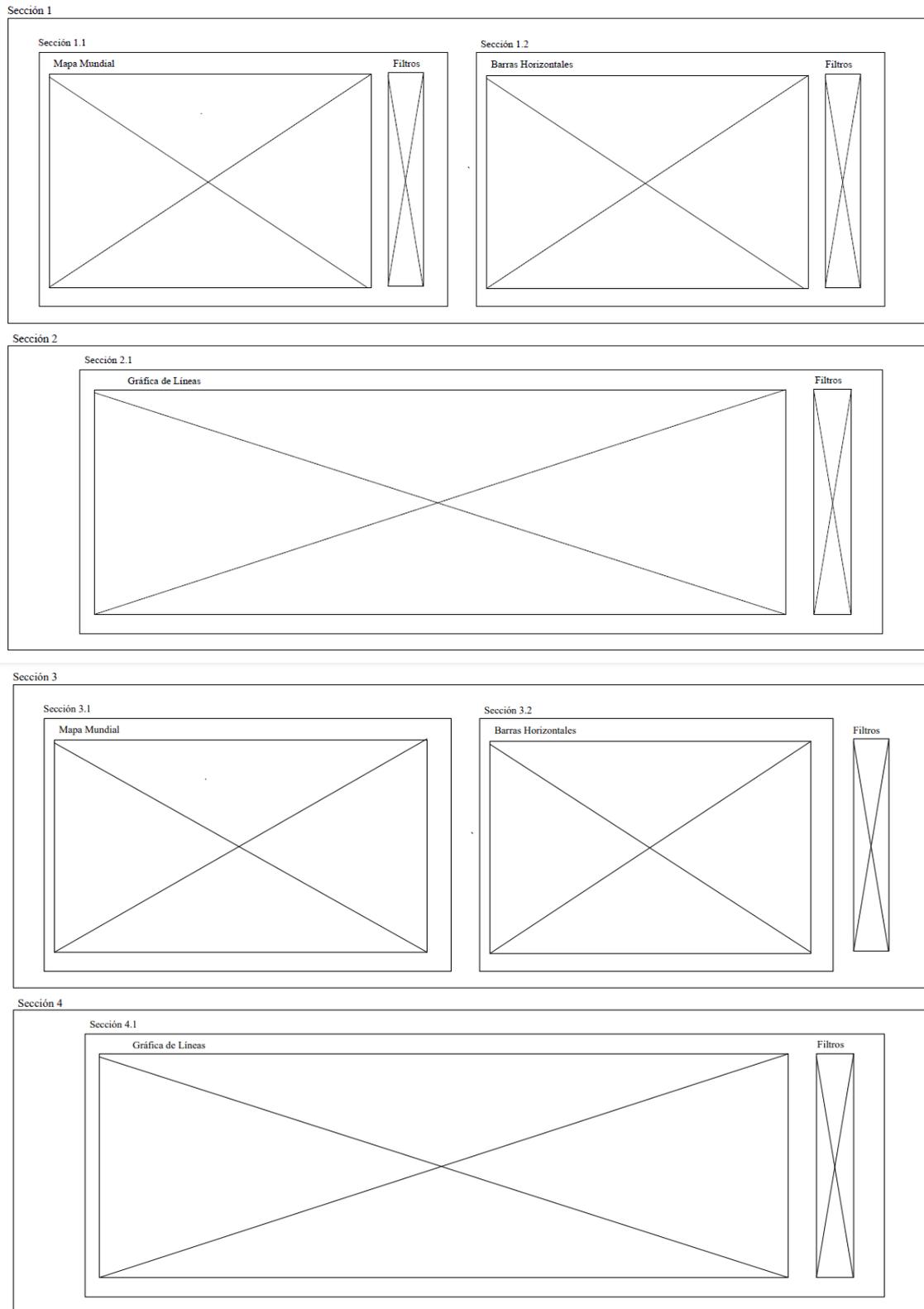
Fuente: Elaborado por el autor en *Tableau*

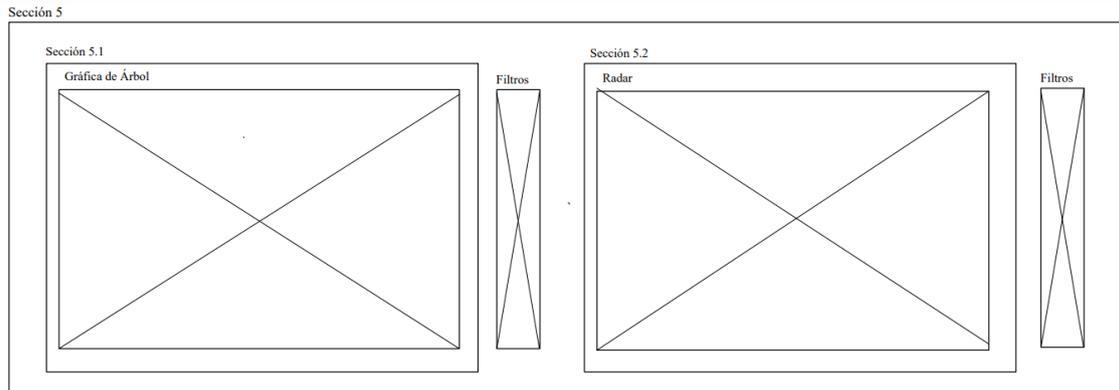
La Figura 9 representa un gráfico de tipo araña o radar el cual demuestra, a través de la implementación de los clústeres obtenidos mediante el uso de minería de datos, patrones en base a las variables analizadas (Costos por unidad, modelo de gravedad y precio por tonelada). La gráfica de tipo araña es una representación basada en coordenadas radiales, tomando como referencia los distintos valores que son evaluados.

Para identificar de mejor manera la estructura del modelo visual, se creó un *mockup* en base a las necesidades de la audiencia del proyecto. El *mockup* es una representación visual de la distribución de elementos del objetivo final, de manera simplificada. A continuación, se observa el *mockup* creado:

Figura 10.

Mockup del modelo visual





Nota: Esta figura muestra el mockup del modelo visual creado.

En la Figura 10, se observan las diferentes plantillas creadas para representar de manera gráfica la información analizada. Como se puede observar, existen cinco secciones principales que componen las visualizaciones.

La primera sección consiste en dos subsecciones. La sección 1.1 hace referencia a la visualización geo-espacial que se utilizará para representar la información de manera global, mediante el uso de un mapa mundial. Existen filtros a nivel de la sección 1 que afectarán a las subsecciones 1.1 y 1.2, al momento de personalizar y evaluar el modelo. En la sección 1.2, existe la gráfica de barras horizontales. La misma se basa en una visualización bidimensional de dos ejes en el cual se evalúan las importaciones.

En la parte inferior se encuentra la sección 2. La misma cuenta con dos subsecciones. La sección 2.1 hace referencia a una gráfica de línea espacio-temporal para analizar el comportamiento de la importación de una partida arancelaria a través del tiempo. La sección 2.2 representa una gráfica de barras verticales que refleja, en base a los costos de importación, los diferentes valores que existen para una partida arancelaria en específico dentro del período seleccionado. Para la gráfica lineal espacio-temporal (sección 2.1), existe la posibilidad de filtrar la información según las necesidades del usuario. Las dos gráficas en la sección inferior son representaciones visuales basadas en las técnicas de ejes, de dos dimensiones.

Seguido de esto, se encuentran las siguientes dos secciones. La sección 3 está formada por dos gráficas principales, un mapa mundial basado en el esquema geoespacial y una gráfica de barras horizontales, basada en ejes. Como se puede observar en la Figura 10, los filtros dentro de la sección 3 afectan a ambas subsecciones, 3.1 y 3.2. Por otro lado, la sección 4 engloba a la subsección 4.1, la cual hace referencia a una gráfica de línea espacio-temporal donde se observa el cambio de la cantidad de toneladas importadas a lo largo de los años evaluados. Al igual que en la sección 2.1, la gráfica de línea de la sección 4.1 cuenta con parámetros para evaluar el comportamiento de una o varias partidas arancelarias. También se implementó una línea de tendencia en ambas gráficas lineales para analizar patrones en las importaciones.

El último componente, sección 5, consiste en dos subsecciones principales. La sección 5.1 es una gráfica de tipo árbol donde se evalúan las importaciones de una o varias partidas arancelarias, establecidas por el parámetro, en base a las toneladas de importación de la misma. Esta información se representa mediante el uso de rectángulos anidados para demostrar una jerarquía. La sección 5.2 hace referencia a una gráfica de tipo radar que, utilizando coordenadas radiales, evalúa los clústers generados para crear un pre análisis a la información recopilada.

4. Resultados

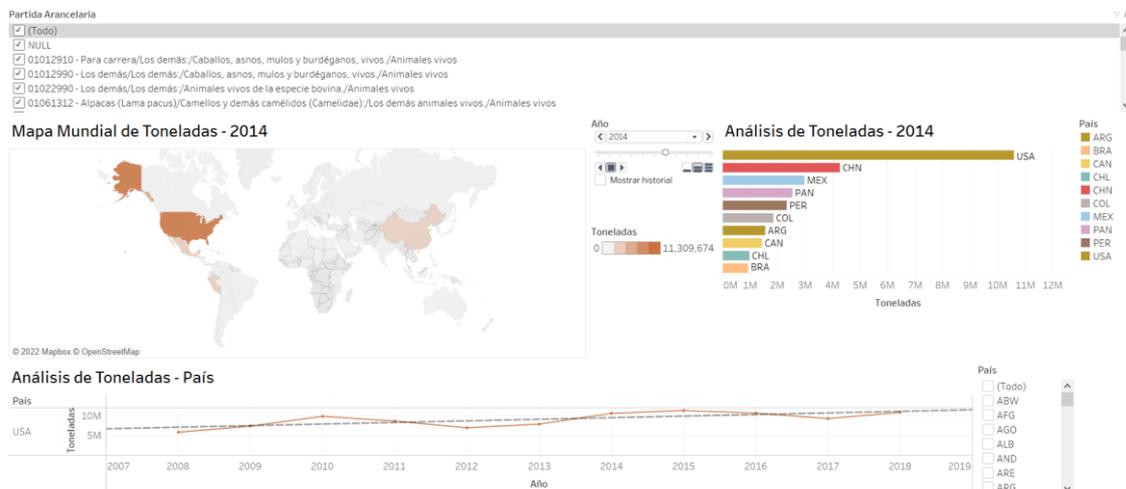
4.1 Modelo Visual

El espacio creado para el modelo visual se define en tres tableros de información, utilizando espacio bidimensional, como se puede observar en las siguientes figuras. Dentro de cada tablero, existen varias visualizaciones que ayudan al usuario a evaluar la información y, a través del uso de parámetros, el usuario tiene la posibilidad de personalizar los resultados de las visualizaciones, dependiendo de sus necesidades.

El primer tablero se compone de tres gráficas, una visualización del mapa mundial, una gráfica de barras horizontales y una gráfica de líneas, como lo muestra la Figura 11.

Figura 11.

Tablero 1 – Modelo Visual Países

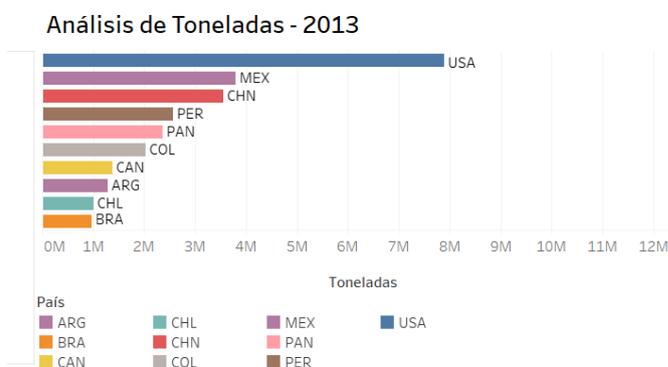


Nota: Esta gráfica muestra las visualizaciones incluidas en el primer tablero creado.

El mapa mundial es una visualización geoespacial, ya que toma en cuenta la ubicación real de los países de los cuales importa el Ecuador. Mediante la implementación de coropletas, se observa la cantidad en toneladas que importa el Ecuador de los distintos países. La intensidad del color representa una mayor cantidad de toneladas de importación. La visualización tiene como parámetro el año de comercio exterior. Al utilizar un filtro de tipo página, es posible crear una animación en la cual se observa el comportamiento de las importaciones a lo largo de los años evaluados. Es decir, entre los años 2008 y 2018. Además, el usuario tiene la posibilidad de observar distintos datos relevantes acerca de cada país, tales como la distancia a la cual se encuentra el país con respecto al Ecuador, el valor PIB del país del cual se importa, el valor CIF de las importaciones del año seleccionado, los costos de importación por unidad del año de importación, las toneladas importadas en el año, el valor FOB de las importaciones del año, los costos de importación del año y el precio por tonelada del año seleccionado.

Figura 12.

Gráfica de barras horizontales - Países



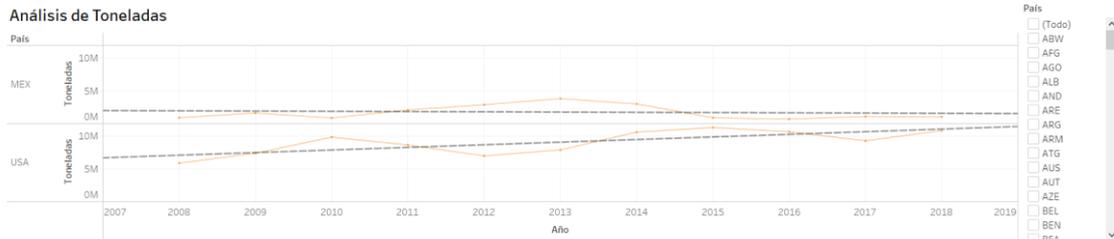
Nota: Esta gráfica muestra la visualización de barras horizontales creada

Como muestra la Figura 12, la gráfica de barras horizontales está basada en ejes, con las toneladas en el eje horizontal y los países en el eje vertical. Con esta visualización, se puede observar, ordenado de mayor a menor, el comportamiento de las toneladas de importación de los 10 países de los cuales se importa más.

La visualización de barras horizontales tiene como parámetro, al igual que el mapa mundial mencionado anteriormente, la partida arancelaria y año de importación. Además, se aplicó una paleta de colores para identificar a cada país.

Figura 13.

Gráfica de líneas - Países

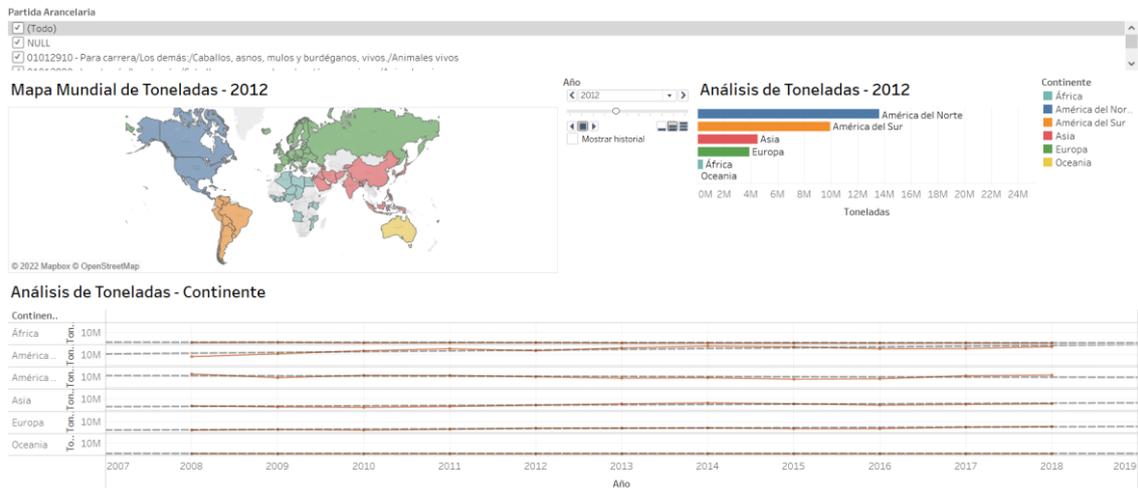


Nota: Esta gráfica muestra la visualización de líneas creada para analizar las toneladas importadas.

La Figura 13. muestra una gráfica de líneas de dos dimensiones, basado en ejes. Por un lado, el eje horizontal tiene los años de importación y en el eje vertical, está la cantidad en toneladas de las importaciones. El usuario, al deslizarse sobre un punto en la gráfica, observa los datos relevantes mencionados anteriormente para evaluar la información. La mencionada visualización tiene como parámetro el país del cual se importa. De esta manera, es posible filtrar el país que se desea evaluar, para analizar el comportamiento de las importaciones dentro del periodo establecido. Como se observa en la gráfica, existe la línea de tendencia que ayuda a las diferentes áreas de interés, evaluar el comportamiento de las importaciones a futuro.

Figura 14.

Tablero 2 – Modelo Visual Continentes



Nota: Esta gráfica muestra el segundo tablero creado.

Como se observa en la Figura 14, el segundo tablero es similar al primero mencionado. Se compone de tres gráficas principales, un mapa mundial, una gráfica de barras horizontales y una gráfica de líneas.

Figura 15.

Mapa Mundial - Continentes

Mapa Mundial de Toneladas - 2011



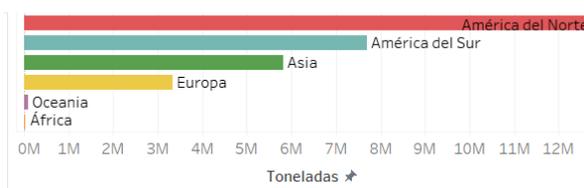
Nota: Esta gráfica muestra la visualización del mapa mundial de los continentes.

La Figura 15. muestra la visualización del mapa mundial implementado. El mismo se basa en un esquema geoespacial de dos dimensiones. En este caso, el mapa mundial se concentra en el análisis de las toneladas importadas a nivel de continente. Esto quiere decir que, al aplicar el parámetro página de Año, se observa el comportamiento de las importaciones de los continentes en base a toneladas durante los años 2008 y 2018. De igual manera, existe el parámetro de partida arancelaria, que brinda al usuario un enfoque de estudio y análisis personalizado para evaluar las importaciones.

Figura 16.

Gráfica de barras horizontales - Continentes

Análisis de Toneladas - 2015

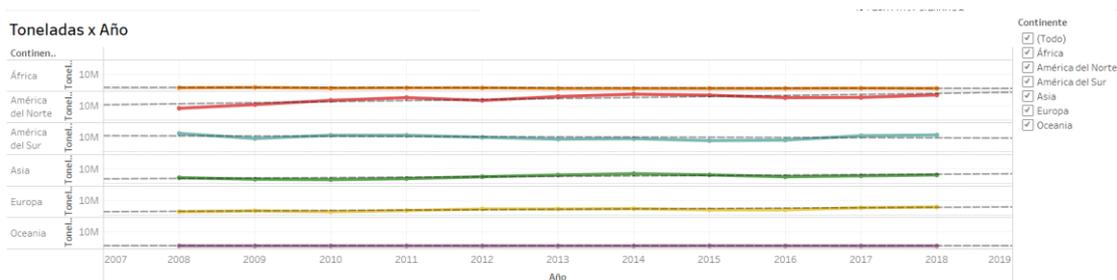


Nota: Esta gráfica muestra la visualización de barras horizontales de los continentes.

Como muestra la Figura 16, se implementó una gráfica de barras horizontales para representar las toneladas importadas a lo largo de tiempo para cada continente. El filtro de página mencionado anteriormente también afecta a esta hoja de trabajo, por lo que es posible visualizar el comportamiento de las importaciones de los distintos continentes a través del tiempo. Como se observa, se utilizó una paleta de colores para identificar de mejor manera a cada continente.

Figura 17.

Gráfica de Líneas - Continentes

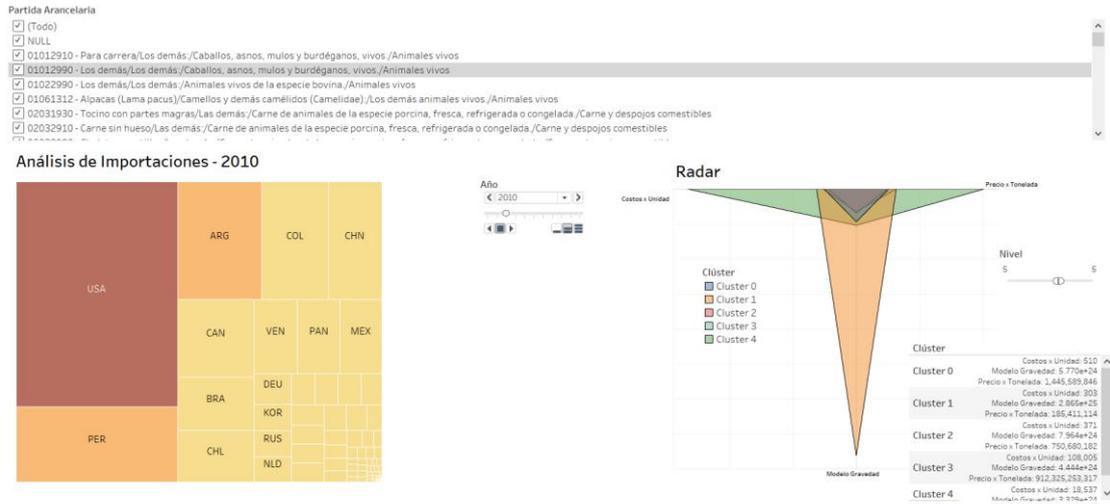


Nota: Esta visualización representa las importaciones de los distintos continentes.

La Figura 17. muestra una gráfica de líneas utilizada para demostrar al usuario final, los diferentes valores de toneladas con respecto a los continentes de los cuales importa el Ecuador. De igual manera, se incluyó la línea de tendencia dentro de la gráfica. La misma es de gran utilidad, ya que representa un patrón en el comportamiento de las importaciones de los distintos continentes a través de los años evaluados, en base a las toneladas de importación.

Figura 18.

Tablero 3 – Modelo Visual



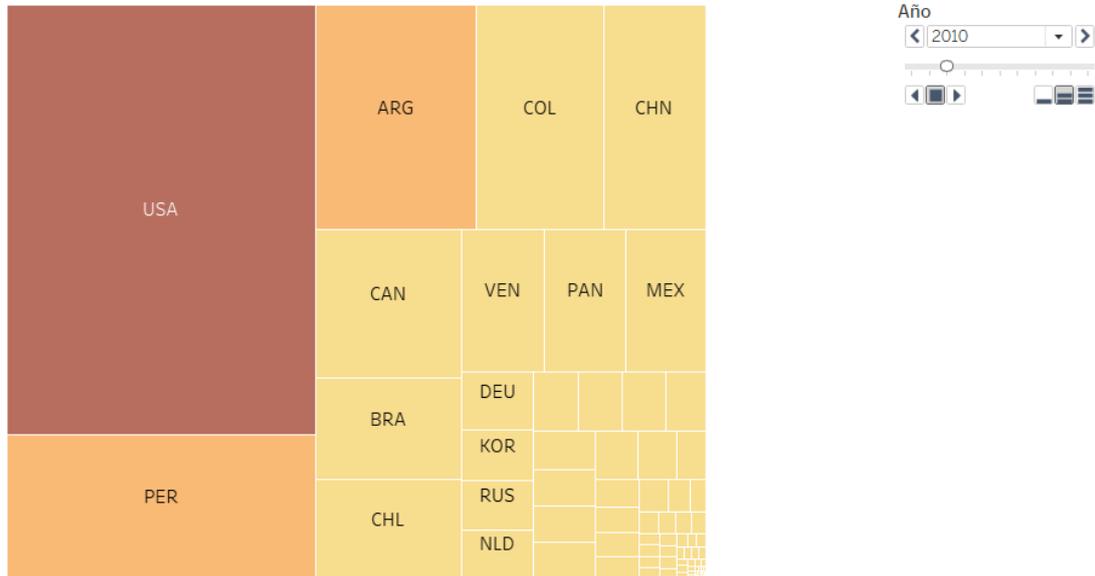
Nota: Esta figura representa el tercer tablero de visualizaciones.

La Figura 18. muestra el tercer tablero creado dentro del modelo visual. Como es posible observar, se componen de dos gráficas principales. Una gráfica de árbol y una gráfica de radar. Como se puede observar, las dos visualizaciones creadas en el tercer tablero, tienen como filtro, en la parte superior, la partida arancelaria que se desea evaluar.

Figura 19.

Gráfica de tipo árbol - Países

Análisis de Importaciones - 2010

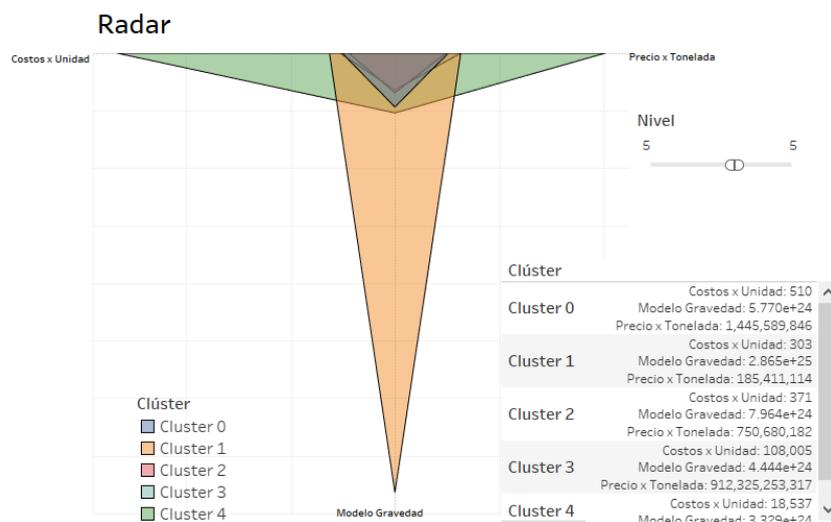


Nota: Esta visualización representa el análisis de toneladas de países.

Como muestra la Figura 19, la visualización de tipo árbol se emplea para demostrar una jerarquía. En este caso, se adaptó la visualización bidimensional para mostrar la participación de los países en la importación de un producto o servicio en específico. Dentro de la visualización, se utilizan rectángulos anidados para representar los países, y una paleta de color que la intensidad de la misma, hace referencia a una mayor cantidad de tonelaje de importación. En otras palabras, la gráfica muestra el tonelaje de importación de una partida arancelaria que se desea evaluar en el período seleccionado. Al igual que las otras visualizaciones mencionadas anteriormente, se emplea un filtro de tipo página para analizar el comportamiento de la importación de aquella partida arancelaria a lo largo del tiempo.

Figura 20.

Gráfica de radar - Clústers



Nota: Esta visualización muestra la clasificación por clústers.

La Figura 20. muestra la gráfica de radar implementada en el tercer tablero del modelo visual creado. La visualización de tipo radar se basa en coordenadas radiales para crear los diferentes vértices de la gráfica, en base a las variables seleccionadas. La gráfica de radar se creó en torno a tres variables analizadas: precio por tonelada, modelo de gravedad y costos por unidad. La mencionada gráfica evalúa una o varias partidas arancelarias a un nivel establecido, con el objetivo de clasificar las importaciones del mismo en diferentes clústers o agrupaciones, en base a las tres variables detalladas. Dicho esto, es posible visualizar cada clúster que se obtuvo como resultado de la aplicación de las técnicas de minería de datos y de esta manera, generar un pre análisis a la información. Además, existen filtros que se pueden utilizar como parámetros para evaluar la información a un nivel específico. También cuenta con un recuadro informativo en la parte inferior derecha, para visualizar los datos que hacen referencia a los clústers representados.

5. Evaluación

La evaluación del modelo de visualización de datos fue llevada a cabo por tres expertos en el área de estudio. La Tabla 3 muestra las distintas preguntas creadas para la aplicación del modelo TAM-BDA.

Tabla 3.

Tabla de Constructos

| Constructos | Ítem | Medida |
|--|-------------|--|
| Intención conductual de usar modelos de visualización de datos | BI1 | Estoy entusiasmado con el uso del modelo visual que abarca las importaciones del Ecuador, entre los años 2008 y 2018. |
| | BI2 | Es mi deseo ver la plena utilización y despliegue del modelo visual acerca de las importaciones del Ecuador, entre los años 2008 y 2018. |
| Actitud hacia los modelos de visualización de datos | ATT1 | Creo que la implementación de un modelo visual de importaciones del Ecuador entre los años 2008 y 2018, facilitará la toma de decisiones en tiempo real. |
| | ATT2 | Creo que el modelo de visualización de datos de las importaciones del Ecuador, a nivel mundial, entre los años 2008 y 2018; será mejor que los sistemas tradicionales de visualización de datos. |
| Utilidad percibida de los modelos de visualización de datos | PU1 | Creo que el uso del modelo de visualización de datos de las importaciones del Ecuador, a nivel mundial, entre los años 2008 y 2018; mejoraría la eficacia de las operaciones de mi trabajo. |
| | PU2 | Creo que el uso del modelo de visualización de datos puede aumentar la productividad de mi trabajo. |
| | PU3 | El uso del modelo de visualización de datos creado mejora el rendimiento del trabajo |
| Facilidad de uso percibida de los modelos de visualización de datos | PEU1 | El proceso de implementación/uso del modelo visual acerca de las importaciones del Ecuador es comprensible para mí |
| | PEU2 | Cree que es fácil integrar los modelos visuales de datos con los sistemas de información existentes. |
| Creencias en los beneficios de los modelos de visualización de datos | BENB1 | Mi equipo directivo cree en los beneficios de un proyecto de implementación de un modelo de visualización para el análisis de importaciones. |
| | BENB2 | Mis compañeros creen en los beneficios de un nuevo sistema de análisis de importaciones del Ecuador a nivel mundial. |

| | | |
|-------------------------------------|-------|---|
| | BENB3 | Creo en los beneficios de este nuevo modelo visual. |
| Calidad percibida del modelo visual | SQ1 | El modelo visual acerca de las importaciones del Ecuador, brinda información en tiempo y forma. |
| | SQ2 | El modelo visual presentado hace que la información sea muy accesible para los interesados. |
| | SQ3 | El modelo visual creado tiene la posibilidad de adaptarse de manera flexible a las nuevas demandas y condiciones del trabajo. |
| Calidad de la información percibida | IQ1 | Creo que la información disponible del modelo visual acerca de las importaciones del Ecuador a nivel mundial, entre los años 2008 y 2018; es confiable. |
| | IQ2 | Creo que la información disponible del modelo visual acerca de las importaciones del Ecuador a nivel mundial, entre los años 2008 y 2018 es precisa. |
| | IQ3 | El modelo de visualización de datos acerca de las importaciones del Ecuador puede proporcionar información a partir de un gran volumen de datos. |

Nota: Esta tabla muestra los constructos evaluados del modelo TAM

Los constructos utilizados son: intención conductual de usar modelos de visualizaciones de datos, actitud hacia los modelos de visualizaciones de datos, utilidad percibida de los modelos de visualización de datos, facilidad de uso percibida de los modelos de visualización de datos, creencias en los beneficios de los modelos de visualización de datos, calidad percibida del modelo visual y calidad de la información percibida.

Cada constructo contiene una cantidad de ítems que sirven de identificador para medir el mismo. El constructo de intención conductual (BI) se descompone en dos ítems que a su vez son reflejadas en dos medidas para observar y analizar el propósito del usuario en utilizar modelos de visualización de datos. Los mismos se miden con el ítem BI1 y BI2, y las medidas son “Estoy entusiasmado con el uso del modelo visual que abarca las importaciones del Ecuador, entre los años 2008 y 2018” y “Es mi deseo ver la plena utilización y despliegue del modelo visual acerca de las importaciones del Ecuador, entre los años 2008 y 2018.”, respectivamente. La actitud hacia los modelos de visualizaciones de datos (ATT) hace referencia a la actitud de los usuarios con la implementación del modelo visual presentado. Al igual que el constructo anterior, se crearon dos ítems, ATT1 y ATT2, con las medidas de “Creo que la implementación de un modelo visual de importaciones del Ecuador entre los años 2008 y 2018, facilitará la toma de decisiones en tiempo real.” y “Creo que el modelo de visualización de datos de las importaciones del Ecuador, a nivel mundial, entre los años 2008 y 2018; será mejor que los sistemas tradicionales de visualización de datos.”, respectivamente. Para la utilidad percibida de los modelos de visualización de datos (PU), se crearon tres medidas para calcular la perspectiva del usuario en cuanto a la utilidad que se puede tener del modelo visual presentado, dentro de su área de trabajo. La facilidad de uso percibida de los modelos de visualizaciones de datos (PEU) se refiere a la sencillez que existe al adaptar el modelo creado con otras tecnologías y la simplicidad de la implementación del mismo. Para medir esto, se utilizaron dos ítems (PEU1 y PEU2) dos medidas, respectivamente. El constructo de las creencias en los beneficios de los modelos de visualización de datos (BENB) se mide con tres ítems (BENB1, BENB2 y BENB3), que hacen referencia a la opinión de distintas personas acerca de los beneficios que tiene la implementación de un modelo de visualización de datos en el ámbito laboral. Por otro lado, el constructo de calidad percibida del modelo visual (SQ) mide, a través de tres ítems (SQ1, SQ2 Y SQ3), la adaptabilidad del modelo con respecto a las herramientas actuales. También mide la accesibilidad del modelo creado por diferentes partes interesadas y la forma en la que se presenta la información. Finalmente, la calidad de la información percibida (IQ) se mide con tres ítems (IQ1, IQ2, IQ3) hace referencia a la exactitud y precisión de la información presentada a través del modelo visual.

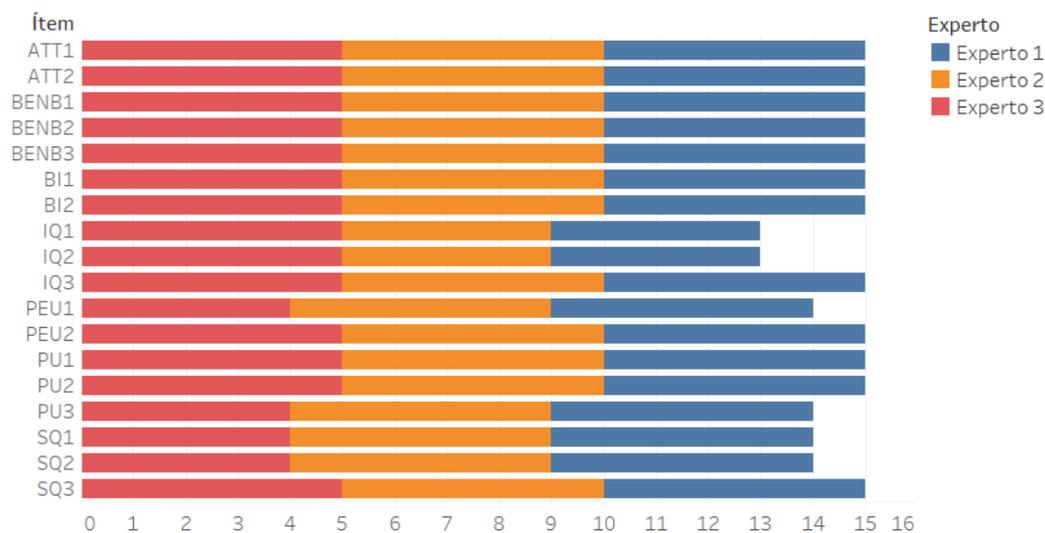
Las mencionadas medidas fueron presentadas a los expertos en un cuestionario digital. Cada medida fue valorada utilizando la escala de Likert de 5 puntos de acuerdo a lo establecido, siendo uno el valor de “Muy

en desacuerdo” y cinco, “Muy de acuerdo”. De esta manera, la evaluación se compone en dos partes principales. En la primera, los tres expertos en el área utilizaron el modelo visual para navegar a través de la información recopilada y generar conocimiento valioso. En la segunda etapa, los expertos llenaron la encuesta digital para evaluar el modelo después de su experiencia. En la dirección URL (<https://forms.gle/MLXbKzCwuHttG3nL8>) se encuentra el cuestionario implementado con las respectivas medidas.

Los resultados obtenidos de la implementación del modelo TAM, basado en una prueba Alpha de Cronbach, fue de 0.82. La prueba de Alpha de Cronbach es un cálculo estadístico que toma los resultados de una encuesta para analizar si los resultados de la misma, son fiables y si cuentan con inconsistencias, basado en la covarianza entre preguntas y respuestas obtenidas (Taber, 2018). En este caso, los valores de la prueba se encuentran dentro de un rango de 0 a 1. Si el valor obtenido es mayor a 0.8, significa que el modelo visual es fiable. De igual manera, si el valor obtenido es menor a 0.8, significa que existen inconsistencias dentro del modelo. Para realizar el cálculo estadístico, se utilizó la herramienta *RStudio*. Dicho esto, el resultado obtenido demuestra los constructos evaluados se encuentran en el rango de fiable. A continuación, en la Figura 21 se observan las respuestas obtenidas por los tres expertos en el área, en base a cada ítem mencionado anteriormente:

Figura 21.

Resultados – Modelo TAM



Nota: Esta visualización muestra los resultados de la encuesta realizada para el modelo TAM.

6. Discusión

Los resultados obtenidos en la investigación demuestran que la creación de un modelo visual que analice las importaciones del Ecuador entre el 2008 y 2018, facilita el uso y estudio de la información desde el punto de vista de las partes interesadas, como empresarios, gobiernos, entre otros. Como se mencionó anteriormente, las diferentes visualizaciones implementadas en el modelo visual son de gran utilidad para identificar patrones de comercio entre el Ecuador y los demás países de los cuales se importa. Además, se planteó un análisis enfocado en las importaciones del Ecuador, mas no de manera mundial, lo cual no existía previamente.

Los análisis de patrones de comportamiento basados en la visualización de datos es relevante para componer análisis multivariados, como lo demuestra el estudio de Jiang et al., (2018). Sin embargo, los análisis no consideran datos generados por procesos de minería de datos. El clústering es un claro ejemplo de proceso de minería de datos y permite al experto seleccionar grupos focalizados a través de la implementación de modelos preconstruidos con el aporte de estos mismos expertos. De igual manera, la selección de gráficas que faciliten la comprensión de las mismas por parte del usuario final es indispensable. Los tipos de visualizaciones son el encargado de transmitir los resultados al usuario que evalúa y utiliza el modelo creado, como lo demuestra el estudio realizado por Wei & Liran, (2017). Sin embargo, la selección del mismo debe optimizar la transmisión de información, utilizando visualizaciones intuitivas e informativas

que sean de navegación fácil y detalladas para el usuario. Para esto, se implementaron gráficas que simplifican la visualización de la información evaluada, a través del uso de distintos tipos de visualizaciones, tales como geo-espaciales y espacio-temporales. Las mencionadas visualizaciones aportan al usuario una perspectiva distinta a la información recopilada y genera conocimiento valioso que es de interés a los usuarios finales. Mediante el uso de paletas de colores distintivos para el usuario y una navegación armónica para el mismo, se consigue lo planteado en el modelo visual creado.

La ecuación del modelo de gravedad brinde un análisis basado en teoría de economía que aporta valor al modelo visual creado, al evaluar el flujo comercial entre dos países que comercian. La misma toma los valores PIB de los países que comercializan para ese año evaluado, y la distancia a la que se encuentran, respectivamente. Aunque la investigación realizada por Boar et al., (2017), analiza el comportamiento de del comercio exterior entre distintos países en base a las toneladas de importación, el presente modelo visual creado utiliza la ecuación del modelo de gravedad para ampliar el campo de estudio. Los mencionados autores también utilizaron gráficas de tipo nodo para visualizar la información analizada, pero en el presente estudio se crearon diferentes visualizaciones en base a una mayor cantidad de variables, como el modelo de gravedad, y no solamente las toneladas de importación, aplicado a distintos tipos de visualización. Además, en el presente modelo creado, a diferencia de los mencionados trabajos, existe la posibilidad de navegar y personalizar las visualizaciones de acuerdo a las necesidades del usuario.

Existen algunas limitaciones que se encontraron durante la realización de la investigación, como la falta de información de las fuentes de datos de donde se extrajo la misma. Esto es importante ya que se pierden algunos registros por la falta de información completa. De igual manera, los registros extraídos pueden contar con diferencias en la actualidad debido a las partidas arancelarias evaluadas. Las mismas fueron alteradas en el tiempo y pueden afectar las visualizaciones creadas.

En base al lineamiento establecido anteriormente, se refleja lo que se planteó con respecto a la teoría propuesta como la creación del modelo visual en torno al mockup y a la selección de los campos analizados, los cuales fueron solicitados por el experto en el área de estudio. Sin embargo, si bien se implementó un filtro para seleccionar la partida arancelaria que se desea evaluar, no se creó una navegación jerárquica de las partidas, ya que algunas partidas no poseen todos los niveles como otras.

7. Conclusión

En el presente estudio, se detalló la creación de un modelo visual desde la extracción de información de distintos repositorios de información como el Banco Mundial, hasta la creación de distintas visualizaciones de datos. El objetivo general del presente estudio fue la construcción de un modelo de visualizaciones de datos para reconocer los patrones de comportamiento de las importaciones del Ecuador a nivel mundial, dentro del período establecido. También se planteó la validación del mismo como un objetivo específico. Dicho esto, se siguieron un conjunto de pasos para la realización del modelo, el cual inicia con la recopilación de la información de distintas fuentes mencionadas, para posteriormente aplicar técnicas de preprocesamiento de datos y técnicas de minería de datos. Esto estructura y clasifica los datos colectados. Los datos procesados fueron alimentados a una herramienta, en este caso *Tableau*, para crear distintos tableros de información que sean de utilidad a grupos de interés. Dentro del modelo creado, se utilizó la ubicación física de los países, paletas de colores y distintos objetos gráficos para representar la mencionada información. La visualización de información, aporta a la toma de decisiones de distintos grupos de interés, y la creación de políticas, ya que es posible observar, el comportamiento de las importaciones del Ecuador.

El presente trabajo contribuye a la comprensión de las importaciones del Ecuador entre los años 2008 y 2018, ya que contiene información valiosa acerca de la misma, la cual no se ha explotado de manera entera previamente. Sin embargo, se podría ampliar el período de estudio hacia años anteriores para analizar otros efectos en cuanto a las importaciones del Ecuador. Algunos de estos son el cambio de moneda del Ecuador, la pandemia de COVID-19, entre otros casos. Asimismo, se podrían crear diferentes visualizaciones en base a los requerimientos solicitados. De esta manera, se abren nuevas ramas de investigación a partir del presente trabajo, como el análisis de las importaciones en base a distintas variables como la población de un país, valor PIB por persona, entre otras.

A través de la aplicación del modelo de aceptación tecnológica extendida (TAM-BDA), se concluye que el modelo demuestra la eficiencia y usabilidad de las visualizaciones creadas dentro del modelo visual propuesto. La interacción de usuarios a través de la posibilidad de personalización de gráficas ha sido implementada exitosamente. Estas interacciones han llevado a que distintos grupos de interés realicen el

análisis en base a los requerimientos que tienen. También se cumplieron con todos los objetivos planteados inicialmente. Al crear un modelo visual acerca de las importaciones del Ecuador entre el 2008 y 2018, utilizando información consolidada de distintas fuentes, se facilita el análisis de las importaciones y genera conocimiento valioso.

8. Referencias

- Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, 12(16), 4102–4107. <https://doi.org/10.3923/jeasci.2017.4102.4107>
- Banco Central del Ecuador. (2017). *Metodología. Información estadística mensual*.
- Boar, R., Iovanovici, A., & Ciocarlie, H. (2017). *Complex networks analysis of international import-export trade*. 31–34.
- Bornal, D. R., Silvestrini, M. M., Aparecida, L., Pio, S., Costa, A. C., Peche, P. M., Cristina, M., & Ramos, P. (2021). *Brazilian position in the international fresh fruit trade network*. 5, 1–13.
- Castro, S. (2021). *El Proceso de Visualización*. 1–30.
- Gupta, B., Kumar, R., & Kumar, A. (2020). Towards Information Discovery On Large Scale Data: state-of-the-art. *2018 International Conference on Soft-Computing and Network Security (ICSNS)*, 1–9.
- Huang, H. (2014). *Analysis of the influence factors of international trade flows based on the trade gravity model and the data of China's empirical*. 728–730.
- Islam, M., & Jin, S. (2019). An Overview of Data Visualization. *International Conference on Information Science and Communications Technologies: Applications, Trends and Opportunities, ICISCT 2019*. <https://doi.org/10.1109/ICISCT47635.2019.9012031>
- Jiang, M., Niu, L., Zhang, Y., Wang, Z., & Ren, X. (2018). A Big-Data-Analysis-Based Research Study on the Dependence on Foreign Trade and in the Trade Structure (2000-2015) of the Yunnan Province of China. *Proceedings - 3rd International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2018, 2018-Janua*, 235–239. <https://doi.org/10.1109/ICITBS.2018.00068>
- Mao, M., & Cheng, X. (2019). Evolution analysis of foreign trade network structure based on complex network SNA. *ACM International Conference Proceeding Series*, 136–140. <https://doi.org/10.1145/3377817.3377839>
- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061. <https://doi.org/10.1109/TKDE.2019.2962680>
- OMG. (2008). *Software & Systems Process Engineering Meta-Model Specification V2.0. April*, 236. <http://www.omg.org/spec/SPEM/2.0/PDF>
- Osman, A. S. (2019). *Data Mining Techniques: Review*. 2(1), 1–4. <https://www.educba.com/7-data>
- Shearer, C. (2000). *The CRISP-DM model: the new blueprint for data mining*. 5(4).
- Shen, H., & Duan, Z. (2020). Application research of clustering algorithm based on K-means in data mining. *Proceedings - 2020 International Conference on Computer Information and Big Data Applications, CIBDA 2020*, 66–69. <https://doi.org/10.1109/CIBDA50819.2020.00023>
- Shreyans Pathak, S. P. (2019). *Data Visualization Techniques, Model and Taxonomy*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-25797-2>
- Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Vazquez, J., & Tonon, L. (2021). *Modelo de gravedad de las exportaciones de cacao en grano del Ecuador. January*. <https://doi.org/10.33890/innova.v6.n1.2021.1591>
- Verma, S., Bhattacharyya, S. S., & Kumar, S. (2018). An extension of the technology acceptance model in the big data analytics system implementation environment. *Information Processing and Management*, 54(5), 791–806. <https://doi.org/10.1016/j.ipm.2018.01.004>
- Wei, R., & Liran, X. (2017). Research on the change of Vietnam's foreign-trade dependence based on big data analysis. *Proceedings - 2016 International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2016*, 253–256. <https://doi.org/10.1109/ICITBS.2016.50>
- Ye, S., Song, C., Cheng, C., Shen, S., & Gao, P. (2020). *Digital Trade Feature Map: A New Method for Visualization and Analysis of Spatial Patterns in Bilateral Trade*. 1–23.
- Zhou, Y., Lin, S., & Kong, S. (2020). *Analysis on the influence of green trade barrier's degree on Vegetable export in China based on gravity model*. 128–135. <https://doi.org/10.1109/MSIEID52046.2020.00031>

