



**COMILLAS**  
UNIVERSIDAD PONTIFICIA

ICAI

# MÁSTER UNIVERSITARIO EN INGENIERÍA DE TELECOMUNICACIONES

TRABAJO FIN DE MÁSTER

## **Diseño e implementación de un modelo para el ahorro de costes y emisiones asociados al comercio internacional**

Autor: Jose Antonio Leiva Orcoyen

Director: Iñigo García García

Madrid



Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título  
Diseño e implementación de un modelo para el ahorro de costes  
y emisiones asociadas al comercio internacional  
en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el  
curso académico 2022/23 es de mi autoría, original e inédito y  
no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido  
tomada de otros documentos está debidamente referenciada.



Fdo.: Jose Antonio Leiva Orcoyen

Fecha: 10 / 07 / 2023

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: Iñigo García García

Fecha: 10 / 07 / 2023



## **AUTORIZACIÓN PARA LA DIGITALIZACIÓN, DEPÓSITO Y DIVULGACIÓN EN RED DE PROYECTOS FIN DE GRADO, FIN DE MÁSTER, TESIS O MEMORIAS DE BACHILLERATO**

### **1º. Declaración de la autoría y acreditación de la misma.**

El autor D. Jose Antonio Leiva Orcoyen

DECLARA ser el titular de los derechos de propiedad intelectual de la obra: Diseño e implementación de un modelo para el ahorro de costes y emisiones asociados al comercio internacional, que ésta es una obra original, y que ostenta la condición de autor en el sentido que otorga la Ley de Propiedad Intelectual.

### **2º. Objeto y fines de la cesión.**

Con el fin de dar la máxima difusión a la obra citada a través del Repositorio institucional de la Universidad, el autor **CEDE** a la Universidad Pontificia Comillas, de forma gratuita y no exclusiva, por el máximo plazo legal y con ámbito universal, los derechos de digitalización, de archivo, de reproducción, de distribución y de comunicación pública, incluido el derecho de puesta a disposición electrónica, tal y como se describen en la Ley de Propiedad Intelectual. El derecho de transformación se cede a los únicos efectos de lo dispuesto en la letra a) del apartado siguiente.

### **3º. Condiciones de la cesión y acceso**

Sin perjuicio de la titularidad de la obra, que sigue correspondiendo a su autor, la cesión de derechos contemplada en esta licencia habilita para:

- a) Transformarla con el fin de adaptarla a cualquier tecnología que permita incorporarla a internet y hacerla accesible; incorporar metadatos para realizar el registro de la obra e incorporar “marcas de agua” o cualquier otro sistema de seguridad o de protección.
- b) Reproducir la en un soporte digital para su incorporación a una base de datos electrónica, incluyendo el derecho de reproducir y almacenar la obra en servidores, a los efectos de garantizar su seguridad, conservación y preservar el formato.
- c) Comunicarla, por defecto, a través de un archivo institucional abierto, accesible de modo libre y gratuito a través de internet.
- d) Cualquier otra forma de acceso (restringido, embargado, cerrado) deberá solicitarse expresamente y obedecer a causas justificadas.
- e) Asignar por defecto a estos trabajos una licencia Creative Commons.
- f) Asignar por defecto a estos trabajos un HANDLE (URL *persistente*).

### **4º. Derechos del autor.**

El autor, en tanto que titular de una obra tiene derecho a:

- a) Que la Universidad identifique claramente su nombre como autor de la misma
- b) Comunicar y dar publicidad a la obra en la versión que ceda y en otras posteriores a través de cualquier medio.
- c) Solicitar la retirada de la obra del repositorio por causa justificada.
- d) Recibir notificación fehaciente de cualquier reclamación que puedan formular terceras personas en relación con la obra y, en particular, de reclamaciones relativas a los derechos de propiedad intelectual sobre ella.

### **5º. Deberes del autor.**

El autor se compromete a:

- a) Garantizar que el compromiso que adquiere mediante el presente escrito no infringe ningún derecho de terceros, ya sean de propiedad industrial, intelectual o cualquier otro.
- b) Garantizar que el contenido de las obras no atenta contra los derechos al honor, a la intimidad y a la imagen de terceros.
- c) Asumir toda reclamación o responsabilidad, incluyendo las indemnizaciones por daños, que pudieran ejercitarse contra la Universidad por terceros que vieran infringidos sus derechos e intereses a causa de la cesión.

- d) Asumir la responsabilidad en el caso de que las instituciones fueran condenadas por infracción de derechos derivada de las obras objeto de la cesión.

**6º. Fines y funcionamiento del Repositorio Institucional.**

La obra se pondrá a disposición de los usuarios para que hagan de ella un uso justo y respetuoso con los derechos del autor, según lo permitido por la legislación aplicable, y con fines de estudio, investigación, o cualquier otro fin lícito. Con dicha finalidad, la Universidad asume los siguientes deberes y se reserva las siguientes facultades:

- La Universidad informará a los usuarios del archivo sobre los usos permitidos, y no garantiza ni asume responsabilidad alguna por otras formas en que los usuarios hagan un uso posterior de las obras no conforme con la legislación vigente. El uso posterior, más allá de la copia privada, requerirá que se cite la fuente y se reconozca la autoría, que no se obtenga beneficio comercial, y que no se realicen obras derivadas.
- La Universidad no revisará el contenido de las obras, que en todo caso permanecerá bajo la responsabilidad exclusiva del autor y no estará obligada a ejercitar acciones legales en nombre del autor en el supuesto de infracciones a derechos de propiedad intelectual derivados del depósito y archivo de las obras. El autor renuncia a cualquier reclamación frente a la Universidad por las formas no ajustadas a la legislación vigente en que los usuarios hagan uso de las obras.
- La Universidad adoptará las medidas necesarias para la preservación de la obra en un futuro.
- La Universidad se reserva la facultad de retirar la obra, previa notificación al autor, en supuestos suficientemente justificados, o en caso de reclamaciones de terceros.

Madrid, a 10 de Julio de 2023

**ACEPTA**



Fdo. Jose Antonio Leiva Orcoyen

Motivos para solicitar el acceso restringido, cerrado o embargado del trabajo en el Repositorio Institucional:



# MÁSTER UNIVERSITARIO EN INGENIERÍA DE TELECOMUNICACIONES

TRABAJO FIN DE MÁSTER

## **Diseño e implementación de un modelo para el ahorro de costes y emisiones asociados al comercio internacional**

Autor: Jose Antonio Leiva Orcoyen

Director: Iñigo García García

Madrid





# Agradecimientos

Tras muchos meses de investigación y desarrollo, el TFM llega a su fin y con ello mi etapa como estudiante de ICAI.

Me gustaría empezar este apartado agradeciendo a EY, a mi tutor Iñigo y al senior manager del departamento Zuriñe, la posibilidad de realizar este proyecto, así como el haber podido tener a mi alcance todas las herramientas necesarias para el desarrollo del mismo. Agradecer el seguimiento, el apoyo y las aportaciones tanto técnicas como fiscales que han llevado a la consecución exitosa del trabajo.

Gracias a mis compañeros y amigos Juan y Marina por las aportaciones tan enriquecedoras que han conseguido que eleve el proyecto a un nivel superior. Al mismo tiempo, agradecer a mis padres, Coral y Jose, por darme la oportunidad de estudiar en ICAI, sus apoyo constante e incondicional, así como sus aportaciones para hacer de este trabajo uno más profesional.

Por último, y no menos importante, quería dedicar este proyecto a Ignacio por sus críticas constructivas que han servido para mejorar y que el proyecto llegue a buen puerto, así como por el apoyo durante esos meses difíciles que parecían no ver el final.

A todos ellos,

GRACIAS.



# DISEÑO E IMPLEMENTACIÓN DE UN MODELO PARA EL AHORRO DE COSTES Y EMISIONES ASOCIADOS AL COMERCIO INTERNACIONAL

**Autor: Leiva Orcoyen, Jose Antonio.**

Director: García García, Iñigo.

Entidad Colaboradora: EY

## RESUMEN DEL PROYECTO

Desarrollo de un sistema de análisis y predicción de los costes involucrados en el comercio internacional, concretamente en el tránsito de las mercancías entre el lugar de origen y el lugar de destino, con el tratamiento aduanero. Para ello, se ha llevado a cabo un sistema de extracción, tratamiento, visualización y predicción de la información, aportándole al cliente un valor adicional esencial para su toma de decisiones estratégicas y posicionamiento de mercado, transformando sus cantidades ingentes de datos en información.

**Palabras clave:** Comercio Internacional, Aduanas, *Machine Learning*, Predicción.

### 1. Introducción

En una era de creciente interconexión y globalización, el comercio internacional se construye como una fuerza vibrante que da forma a la economía mundial. Es la sangre que corre por las venas de las naciones, tendiendo puentes entre continentes y culturas, impulsando el crecimiento económico y transformando las sociedades. Desde la antigua Ruta de la Seda hasta las modernas autopistas digitales, el intercambio de bienes, servicios e ideas a través de las fronteras ha desencadenado la innovación, alimentado la prosperidad y promovido tanto la cooperación como la competencia.

El comercio internacional permite a los países ampliar sus mercados y acceder a bienes y servicios que de otro modo no estarían disponibles en el mercado nacional. Como resultado de ello, el mercado es más competitivo y en última instancia, esto se traduce en precios más competitivos y en un producto más barato para el consumidor. (Heakal, 2023)

Así, el comercio exterior es una actividad que involucra la transferencia de bienes y servicios a través de las fronteras nacionales, que implica muchos costes que generalmente las empresas asumen de manera sistemática dando por hecho que no son objeto de eficiencia económica.

### 2. Definición del proyecto

El proyecto nace a través de una propuesta de EY a un cliente que realiza asiduamente transacciones internacionales y que busca analizar y sacar valor a la información.

Así, aparecen dos tipos de objetivos para el sistema desarrollado:

- Objetivos técnicos
  - Extracción automática de toda la información relativa a los DUA
  - Modelo de visualización y análisis de datos del comercio exterior
  - Modelo de *machine learning* con un elevado grado de precisión
  - Herramienta de predicción de costes en tiempo real
- Objetivos de negocio
  - La reducción de costes del comercio internacional eligiendo la estrategia adecuada
  - Mejora de la eficiencia de los procesos de la empresa
  - Predicción de los costes de una futura transacción internacional
  - Mejora de la competitividad de la empresa en el mercado global

El proyecto se basa, por tanto, en extraer la información de la AEAT, tratarla, modelarla para poder predecirla y finalmente mostrarla en un reporte de visualización. Para el desarrollo del sistema se ha hecho uso de distintas tecnologías:

- Fase de extracción: *web scrapping* con Python *Selenium*
- Fase de limpieza y transformación: Python *Sklearn*
- Fase de modelaje: *Machine Learning* con Python *Sklearn*
- Fase de visualización: *PowerBI*

Por último, la metodología seguida a lo largo de todo el proyecto fue la metodología *Agile* al ser la estándar en el desarrollo de proyectos del departamento, ya que “promueve un proceso disciplinado de gestión de proyectos que fomenta la inspección y la adaptación frecuentes; es una filosofía de liderazgo que fomenta el trabajo en equipo, la autoorganización y la responsabilidad, que permite la entrega rápida de software de alta calidad. Tiene un enfoque empresarial que alinea el desarrollo con las necesidades del cliente y los objetivos de la empresa” (Nuñez, 2016)

El proyecto ha tenido una duración real de 6 meses con puntos de control mensuales con cliente y reuniones semanales con el *scrum master*. Además, se han dado numerosas reuniones de apoyo técnico y de negocio tanto con el equipo de tecnología como con el equipo de fiscal de EY. Durante el desarrollo del proyecto se ha recibido una formación específica de *PowerBI*.

### 3. Descripción del modelo/sistema/herramienta

La herramienta consta de una primera fase de extracción de la información con su posterior limpieza y tratamiento de los datos. Una segunda fase de desarrollo de un modelo de predicción con inteligencia artificial. Y una última fase de reporte y visualización de la información extraída.

Se define un sistema de extracción automática de la información de los DUA (información que documenta los detalles de las mercancías en las aduanas) a través de la página web de la Agencia Tributaria. La información será estudiada desde un punto de vista de negocio para incluir únicamente las variables que conducen a la consecución del objetivo final. También se hará un análisis técnico para limpiar y transformar las variables de manera que se adecuen al modelo de predicción.

Posteriormente, se desarrolla un modelo de *machine learning* que sea capaz de predecir la variable que más valor aporta al cliente cuando una mercancía llega a una aduana, el valor en aduana.

Finalmente, se realiza un reporte de información desde donde el cliente será capaz de visualizar todos los datos, así como realizar un análisis extensivo de los mismos, al mismo tiempo que realiza predicciones en tiempo real de hipotéticas transacciones internacionales futuras.

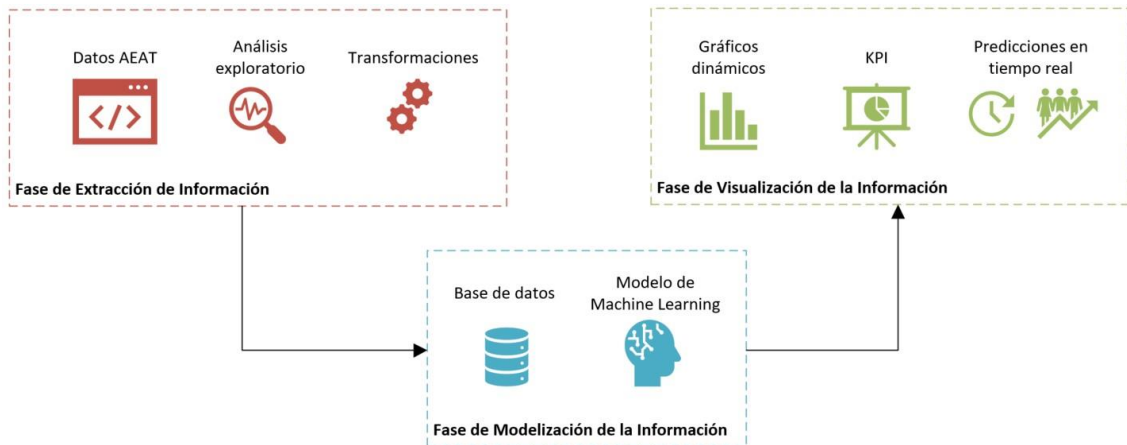


Ilustración 1 - Diseño de alto nivel del sistema desarrollado

#### 4. Resultados

En la primera fase se extrae toda la información referente a los DUA desde que existen registros para el cliente hasta mayo de 2023 directamente de la página web de la AEAT. Inicialmente se obtienen 47 variables, sin embargo, tras el proceso de limpieza y transformación de la información el modelo cuenta únicamente con 8 variables de las cuales 4 fueron numéricas (“Total Bultos”, “Total Factura (€)”, “Masa Bruta (Kg)” y “Unidades Suplementarias”) y las otras 4 categóricas (“Declarante”, “Localización de las mercancías”, “Código TARIC” y “Código Incoterms”)

Durante el proceso de tratamiento de la información se introdujeron en el modelo las variables categóricas usando *label encoding* para las variables que contenían categorías con un orden lógico y *one hot encoding* para las variables cuyas categorías podían tener un orden aleatorio. Se realizan transformaciones logarítmicas y de potencias al cuadrado con el objetivo de eliminar la asimetría y evitar la curtosis. Sin embargo, durante las iteraciones de los modelos se considera no incluirlas ya que empeoran el rendimiento del modelo. Si bien, se realizó una transformación logarítmica a la variable dependiente “Valor en aduana” que mejora considerablemente los resultados del modelo. Además, se normalizan todas las variables numéricas independientes.

La fase de desarrollo del modelo de *machine learning* se basa en numerosas iteraciones hasta la búsqueda de la combinación de variables y modelos más adecuada que dan como resultado la mejor predicción de la variable “Valor en aduana”. Se itera del modelo más simple al más complejo, dando comienzo con un modelo de regresión lineal incluyendo únicamente las variables numéricas más correlacionadas con la variable dependiente. Posteriormente se van introduciendo más variables al modelo, transformándolas e incluso incluyendo variables de nueva creación con sentido técnico y de negocio.

Las medidas para evaluar la precisión del modelo han sido el error cuadrático medio y un histograma de porcentaje de error entre los valores predichos y los valores reales (al ser un modelo de aprendizaje supervisado se cuenta con los datos reales de la variable de estudio) El ECM muchas veces castiga fuertemente al modelo cuando aparecen errores notorios, por lo que se combina con los histogramas para evaluar los modelos.

En la Ilustración 2 se puede visualizar los resultados de la primera iteración del modelo de regresión lineal con una gran cantidad de errores de predicción y un ECM mayor al 100%.

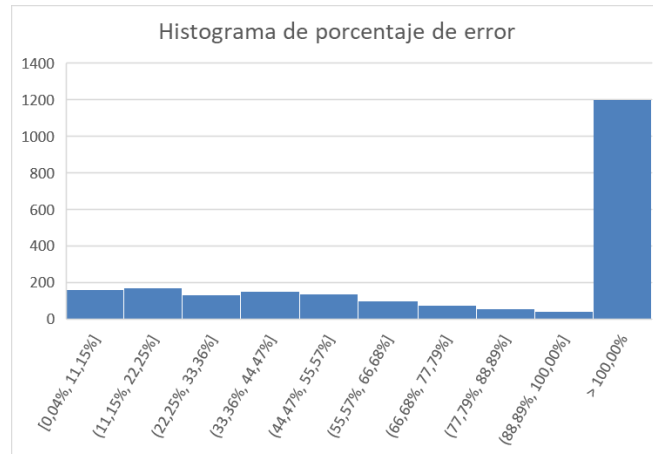


Ilustración 2 - Histograma de error 1ª iteración - Modelo de regresión lineal

Los modelos que se introducen en cada una de las iteraciones son: regresión lineal, regresión polinómica de grado 2 y 3, árbol de regresión, *bagging* con árboles de regresión, *boosting* y *random forest*.

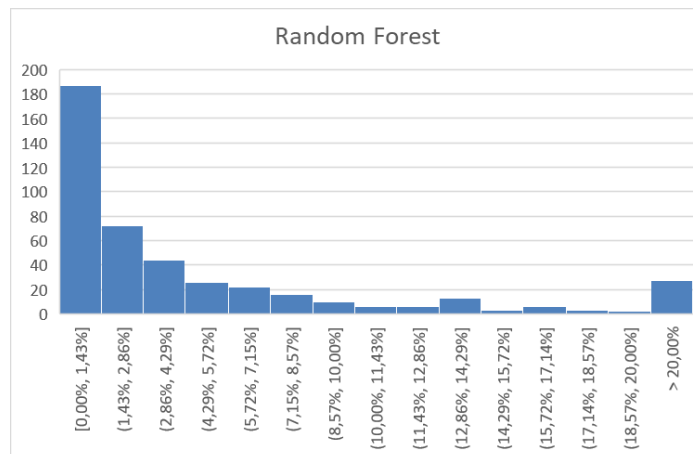


Ilustración 3 - Histograma de error de la última iteración - Modelo Random Forest

El modelo final elegido es un modelo *random forest*, ya que obtuvo un ECM del 6,59%. Además, el histograma de error, tal y como se aprecia en la Ilustración 3, toma una forma de “L” como era deseable. El modelo es capaz de predecir un 42% de la muestra de test con un error inferior al 1,5% y un 85% de la muestra es predicho con un error inferior al 10%. Solo un 6% de la muestra cuenta con un error superior al 20% pero nunca superior al 50%. Tras un análisis de estos registros, se identifica que aquellos DUA con

declarantes o códigos TARIC poco frecuentes (1 o 2 coincidencias en toda la base de datos) son más difíciles de predecir por el modelo, debido a la falta de información.

## **5. Conclusiones**

Con este sistema, la empresa es capaz de visualizar y analizar toda la información relativa a sus transacciones internacionales en un único cuadro de mando interactivo, el cual incluye los principales KPIs de negocio. Estos indicadores están vinculados y permite el filtrado individual o múltiple, lo cual permite a los usuarios finales simplificar el análisis de la información. Este análisis de la información en conjunto con la herramienta de predicción en tiempo real proporciona a la compañía los útiles necesarios para tomar decisiones de negocio que les posicionen en un lugar de ventaja respecto de sus competidores.

El sistema desarrollado es capaz de identificar gastos que son potencialmente reducibles, al mismo tiempo que es capaz de realizar predicciones en tiempo real de los costes asociados creando una mayor eficiencia a la hora de realizar una transacción internacional, ya que la empresa es capaz de anticiparse al mercado. Todo esto se traduce en un aumento de la competitividad de la empresa al colocarse en una posición de ventaja respecto de sus competidores.

Así, se puede concluir que, adoptando el poder del aprendizaje automático en el comercio internacional, se puede desbloquear un mundo de posibilidades ilimitadas, donde los conocimientos basados en datos allanan el camino para tomar decisiones más inteligentes, mejorar la eficiencia y lograr un crecimiento sin precedentes en el mercado global.

## **6. Referencias**

Heakal, R. (Marzo de 2023). International (Global) Trade: Definition, Benefits, Criticisms.

Obtenido de Investopedia: <https://www.investopedia.com/insights/what-is-international-trade/>

Núñez, E. A. (Mayo de 2016). Qué es la Metodología Agile. Obtenido de OpenWebinars:

<https://openwebinars.net/blog/que-es-la-metodologia-agile/>

# **COST AND EMISSIONS SAVINGS MODEL DESIGN ASSOCIATED WITH INTERNATIONAL TRADE**

**Author: Leiva Orcoyen, Jose Antonio.**

Supervisor: García García, Iñigo.

Collaborating Entity: EY

## **ABSTRACT**

Development of a system for analysis and prediction of the costs involved in international trade, specifically in the arrival of goods at destination customs. System of extraction, processing, visualization and prediction of information, which provides an essential value to the customer for decision making and strategic positioning.

**Keywords:** International trade, Customs, Machine Learning, Prediction.

## **1. Introduction**

In an era of increasing interconnectedness and globalization, international trade is building as a vibrant force shaping the world economy. It is the lifeblood that courses through the veins of nations, bridging continents and cultures, driving economic growth and transforming societies. From the ancient Silk Road to modern digital highways, the exchange of goods, services and ideas across borders has sparked innovation, fueled prosperity and promoted both cooperation and competition.

International trade allows countries to expand their markets and access goods and services that would otherwise be unavailable in the domestic market. As a result, the market is more competitive and ultimately this results in more competitive prices and a cheaper product for the consumer. (Heakal, 2023)

Thus, foreign trade is an activity that involves the transfer of goods and services across national borders, which involves many costs that are generally borne systematically by firms on the assumption that they are not subject to economic efficiency.

## **2. Project definition**

The project arose from a proposal from EY to a client that regularly carries out international transactions and seeks to analyze and extract value from the information.

Thus, there are two types of objectives for the developed system:

- Technical objectives
  - Automatic extraction of all information related to DUA.
  - Model for visualization and analysis of foreign trade data.
  - Machine learning model with a high degree of accuracy.
  - Real-time cost prediction tool.
- Business objectives
  - Reducing international trade costs by choosing the right strategy.
  - Improving the efficiency of the company's processes.
  - Predicting the costs of a future international transaction.
  - Improvement of the company's competitiveness in the global market.



The project is therefore based on extracting information from the AEAT, processing it, modeling it in order to predict it and finally displaying it in a visualization report. Different technologies have been used to develop the system:

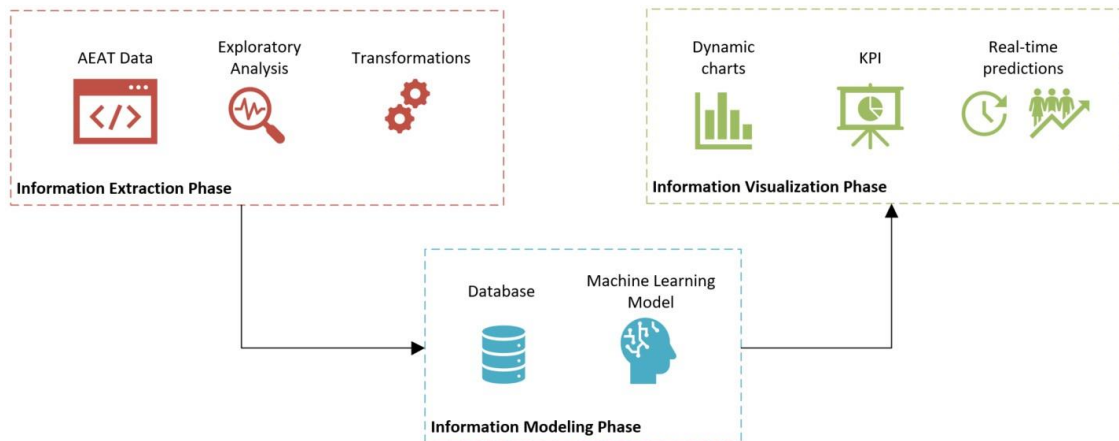
- Extraction phase: web scrapping with Python Selenium.
- Cleaning and transformation phase: Python Sklearn.
- Modeling phase: Machine Learning with Python Sklearn.
- Visualization phase: PowerBI.

Finally, the methodology followed throughout the project was the Agile methodology as it is the standard in the development of projects in the department, since it "promotes a disciplined project management process that encourages frequent inspection and adaptation; it is a leadership philosophy that encourages teamwork, self-organization and accountability, which enables the rapid delivery of high quality software. It has an entrepreneurial approach that aligns development with customer needs and business goals" (Nuñez, 2016)

The project has had an actual duration of 6 months with monthly checkpoints with client and weekly meetings with the scrum master. In addition, there have been numerous technical and business support meetings with both the technology team and EY's tax team. During the development of the project, specific PowerBI training has been received.

### 3. System description

The tool consists of a first phase of information extraction with subsequent data cleaning and processing. A second phase of development of a predictive model with artificial intelligence. And a last phase of reporting and visualization of the extracted information.



*Ilustración 4 - High Level Design of the developed system*

A system for the automatic extraction of DUA information (information documenting the details of the goods in customs) through the Tax Agency's web page is defined. The information will be studied from a business point of view to include only the variables that lead to the achievement of the final objective. A technical analysis will also be performed to clean and transform the variables to fit the prediction model.

Subsequently, a machine learning model is developed that is capable of predicting the variable that brings the most value to the customer when a merchandise arrives at a customs office, the customs value.

Finally, an information report is created from which the client will be able to visualize all the data, as well as perform an extensive analysis of them, while making real-time predictions of hypothetical future international transactions.

#### 4. Results analysis

In the first phase, all the information relating to DUA since there are records for the client until May 2023 is extracted directly from the AEAT website. Initially 47 variables were obtained, however, after the process of cleaning and transformation of the information the model has only 8 variables of which 4 were numerical ("Total Packages", "Total Invoice (€)", "Gross Mass (Kg)" and "Supplementary Units") and the other 4 categorical ("Declarer", "Location of the goods", "TARIC Code" and "Incoterms Code").

During the information processing process, the categorical variables were introduced into the model using label encoding for the variables containing categories with a logical order and one hot encoding for the variables whose categories could have a random order. Logarithmic and power squared transformations are performed in order to eliminate skewness and avoid kurtosis. However, during model iterations it is considered not to include them as they worsen model performance. However, a logarithmic transformation was made to the dependent variable "Customs value" which considerably improves the results of the model. In addition, all numerical independent variables are normalized.

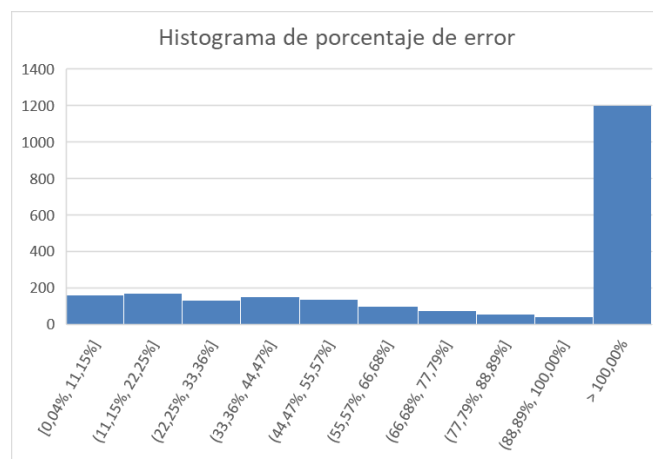


Ilustración 5 – Error histogram of the 1st iteration - Linear regression model

The development phase of the machine learning model is based on numerous iterations until the perfect combination of variables and models is found that results in the best prediction of the "Customs value" variable. It is iterated from the simplest to the most complex model, starting with a linear regression model including only the numerical variables most correlated with the dependent variable. Subsequently, more variables are introduced to the model, transforming them and even including newly created variables with technical and business sense.

The measures to evaluate the accuracy of the model have been the mean square error and a histogram of the percentage error between the predicted values and the real values (since it is a supervised learning model, the real data of the study variable are available).

Figure 5 shows the results of the first iteration of the linear regression model with a large number of prediction errors and an ECM greater than 100%.

The models introduced in each of the iterations are: linear regression, polynomial regression of degree 2 and 3, regression tree, bagging with regression trees, boosting and random forest.

The final model chosen is a random forest model, since it obtained an ECM of 6.59%. In addition, the error histogram, as shown in Figure 6, takes an "L" shape, as was desirable. The model is able to predict 42% of the test sample with an error of less than 1.5% and 85% of the sample is predicted with an error of less than 10%. Only 6% of the sample has an error greater than 20% but never greater than 50%. After an analysis of these records, it is identified that those DUA with infrequent declarants or TARIC codes (1 or 2 matches in the whole database) are more difficult to predict by the model due to the lack of information.

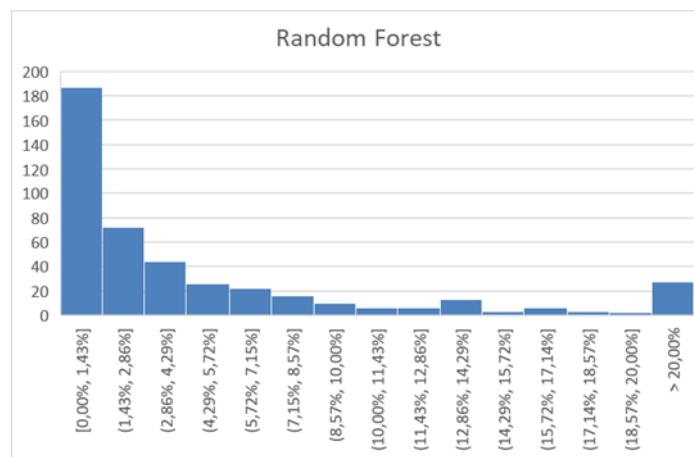


Ilustración 6 - Error histogram of the last iteration - Random Forest Model

## 5. Conclusions

With this system, the company is able to visualize and analyze all the information related to its international transactions using KPIs and filters that allow it to simplify data analysis. This information analysis in conjunction with the real-time forecasting tool provides the company with the necessary tools to make business decisions that put them at an advantage over their competitors.

The developed system is able to identify expenses that are potentially reducible, while at the same time it is able to make real-time predictions of the associated costs, creating greater efficiency when carrying out an international transaction, since the company is able to anticipate the market. All this translates into an increase in the company's competitiveness by placing it in an advantageous position with respect to its competitors.

Thus, it can be concluded that by embracing the power of machine learning in international trade, a world of unlimited possibilities can be unlocked, where data-driven insights pave the way for smarter decisions, improved efficiency and unprecedented growth in the global marketplace.

## **6. References**

Heakal, R. (Marzo de 2023). International (Global) Trade: Definition, Benefits, Criticisms.

Obtenido de Investopedia: <https://www.investopedia.com/insights/what-is-international-trade/>

Nuñez, E. A. (Mayo de 2016). Qué es la Metodología Agile. Obtenido de OpenWebinars:

<https://openwebinars.net/blog/que-es-la-metodologia-agile/>

## *Índice de la memoria*

<b>Capítulo 1. Introducción .....</b>	<b>6</b>
<b>Capítulo 2. Fiscalidad y tributación en el comercio internacional.....</b>	<b>12</b>
2.1 Modelos de comercio internacional.....	12
2.2 Fiscalidad en el comercio internacional.....	17
2.3 Derecho tributario internacional.....	18
2.4 Documento Único Administrativo.....	20
<b>Capítulo 3. Estado de la Cuestión .....</b>	<b>30</b>
3.1 Inteligencia Artificial.....	30
3.2 Blockchain .....	32
3.3 IoT .....	34
3.4 Digital Payments .....	37
3.5 Cloud Computing .....	39
3.6 Tech-Based Logistics.....	40
<b>Capítulo 4. Definición del Trabajo.....</b>	<b>43</b>
4.1 Justificación .....	43
4.2 Objetivos.....	45
4.3 Metodología.....	46
4.4 Planificación y Estimación Económica .....	50
4.5 Descripción de las tecnologías .....	54
<b>Capítulo 5. Sistema Desarrollado.....</b>	<b>64</b>
5.1 Extracción de la información .....	64
5.2 Limpieza y transformación de los datos .....	69
5.3 Modelos de Machine Learning.....	72
5.4 Visualización de la información.....	75
<b>Capítulo 6. Análisis de Resultados .....</b>	<b>83</b>
<b>Capítulo 7. Conclusiones y Trabajos Futuros .....</b>	<b>95</b>
<b>Capítulo 8. Bibliografía.....</b>	<b>98</b>

**ANEXO I 106**

8.1	Tipos de Incoterms 2020 (Santander Trade Markets, 2020).....	106
8.2	Diagrama de Gantt.....	123
8.3	Código Python Extracción DUA .....	125
8.4	Código Python Modelo de Regresión Lineal .....	142
8.5	Código Python Modelo de Regresión Polinómica.....	144
8.6	Código Python Modelo de Árbol de Decisión .....	146
8.7	Código Python Modelo de Bagging .....	148
8.8	Código Python Modelo de XGBoost.....	152
8.9	Código Python Modelo de Random Forest.....	154

**ANEXO II 156**

8.10	Alineación del proyecto con los ODS.....	156
------	--	-----

## *Índice de figuras*

Ilustración 1 - Diseño de alto nivel del sistema desarrollado.....	13
Ilustración 2 - Histograma de error 1ª iteración - Modelo de regresión lineal.....	14
Ilustración 3 - Histograma de error de la última iteración - Modelo Random Forest.....	14
Ilustración 4 - High Level Design of the developed system .....	17
Ilustración 5 – Error histogram of the 1st iteration - Linear regression model .....	18
Ilustración 6 - Error histogram of the last iteration - Random Forest Model.....	19
Ilustración 7 - Incoterms 2020 (DSV, 2020).....	26
Ilustración 8 - Página de búsqueda de DUA de la AEAT .....	65
Ilustración 9 - Etiquetas web del CSS de la página de búsqueda .....	66
Ilustración 10 - Consulta de información de un DUA .....	68
Ilustración 11 - Histograma de la variable "Total Divisas".....	69
Ilustración 12 - Histograma de la variable "Total Divisas" tras una transformación logarítmica.....	70
Ilustración 13 - Histograma de la variable "Valor Aduana" tras una transformación logarítmica.....	71
Ilustración 14 - Histograma de la variable "Total Divisas" tras una transformación cuadrada .....	72
Ilustración 15 - Página "Resumen" del reporte de PowerBI .....	77
Ilustración 16 - Página "Circuitos" del reporte de PowerBI .....	77
Ilustración 17 - Página "Representantes" del reporte de PowerBI .....	78
Ilustración 18 - Página "Orígenes" del reporte de PowerBI.....	79
Ilustración 19 - Página "Mercancías" del reporte de PowerBI.....	80
Ilustración 20 - Página "Aranceles" del reporte de PowerBI .....	80
Ilustración 21 - Página "Valor Aduanas" del reporte de PowerBI .....	81
Ilustración 22 - Página "Regímenes" del reporte de PowerBI .....	82
Ilustración 23 - Matriz de correlación de las variables numéricas .....	83
Ilustración 24 - Histograma de error de la 1ª iteración de regresión lineal.....	84
Ilustración 25 - Histograma de error de la 3ª iteración de regresión lineal.....	85

---

Ilustración 26 - Histograma de error de la 7ª iteración de regresión lineal .....	85
Ilustración 27 - Histograma de error de la 17ª iteración de regresión lineal .....	86
Ilustración 28 - Histograma de error de la 18ª iteración de regresión lineal .....	87
Ilustración 29 - Histograma de error de la 21ª iteración de regresión lineal .....	88
Ilustración 30 - Histograma de error de la iteración definitiva de regresión lineal con la muestra de test .....	89
Ilustración 31 - Histograma de error para una regresión polinómica de grado 2 .....	89
Ilustración 32 - Histograma de error para una regresión polinómica de grado 3 .....	90
Ilustración 33 - Histograma de error para un árbol de regresión .....	90
Ilustración 34 - Gráfica de codo para elegir el número óptimo de estimadores .....	91
Ilustración 35 - Histograma de error para el modelo de bagging con 4 estimadores .....	92
Ilustración 36 - Histograma de error para el modelo de XGBoost .....	92
Ilustración 37 - Histograma de error para el modelo de Random Forest .....	93
Ilustración 38 - Formulaciones EXW (Santander Trade Markets, 2020) .....	107
Ilustración 39 - Formulaciones FCA (Santander Trade Markets, 2020) .....	109
Ilustración 40 - Formulaciones FAS (Santander Trade Markets, 2020) .....	111
Ilustración 41 - Formulaciones FOB (Santander Trade Markets, 2020) .....	113
Ilustración 42 - Formulaciones CFR (Santander Trade Markets, 2020) .....	114
Ilustración 43 - Formulaciones CIF (Santander Trade Markets, 2020) .....	115
Ilustración 44 - Formulaciones CPT (Santander Trade Markets, 2020) .....	117
Ilustración 45 - Formulaciones CIP (Santander Trade Markets, 2020) .....	118
Ilustración 46 - Formulaciones DAP (Santander Trade Markets, 2020) .....	119
Ilustración 47 - Formulaciones DPU (Santander Trade Markets, 2020) .....	121
Ilustración 48 - Formulaciones DDP (Santander Trade Markets, 2020) .....	122
Ilustración 49 - Diagrama de Gantt Pt. 1 .....	123
Ilustración 50 - Diagrama de Gantt Pt. 2 .....	124



## *Índice de tablas*

Tabla 1 - Estimación de costes del proyecto .....53

## Capítulo 1. INTRODUCCIÓN

Si entras en un supermercado y encuentras plátanos de Costa Rica, café de Brasil y una botella de vino de Sudáfrica, estás experimentando los efectos del comercio internacional. (Heakal, 2023)

En una era de creciente interconexión y globalización, el comercio internacional se construye como una fuerza vibrante que da forma a la economía mundial. Es la sangre que corre por las venas de las naciones, tendiendo puentes entre continentes y culturas, impulsando el crecimiento económico y transformando las sociedades. Desde la antigua Ruta de la Seda hasta las modernas autopistas digitales, el intercambio de bienes, servicios e ideas a través de las fronteras ha desencadenado la innovación, alimentado la prosperidad y promovido tanto la cooperación como la competencia.

El comercio internacional permite a los países ampliar sus mercados y acceder a bienes y servicios que de otro modo no estarían disponibles en el mercado nacional. Como resultado de ello, el mercado es más competitivo y en última instancia, esto se traduce en precios más competitivos y en un producto más barato para el consumidor. (Heakal, 2023)

Es por ello, por lo que el comercio internacional desempeña un papel fundamental tanto a nivel macroeconómico como a nivel microeconómico. Es un motor clave del crecimiento económico ya que al ampliar los mercados más allá de las fronteras nacionales, los países pueden acceder a bases de clientes más amplias, aumentar las ventas e impulsar la producción. Además, este impulso de la producción se traduce en un fomento de la competencia, lo que estimula a las empresas a mejorar su eficiencia y las obliga a innovar, adoptar tecnologías avanzadas y optimizar sus procesos de producción.

Todo ello incentiva a mejorar la calidad de los productos y reducir costes beneficiando tanto a las empresas como los consumidores, lo que se traduce en una mejora del nivel de vida. Así, los consumidores se encuentran con una mayor variedad de bienes y servicios a precios

competitivos. Al importar productos de diferentes países, los consumidores tienen acceso a una amplia gama de opciones, incluyendo bienes que pueden no estar disponibles en el mercado nacional, además de tener acceso a ellos a un precio más bajo y con mayor calidad mejorando así su bienestar y aumentando su poder adquisitivo.

Gracias a esta expansión de los mercados los países pueden ampliar su base de exportación, diversificar sus economías y reducir la dependencia de un único mercado o industria. Esto es especialmente beneficioso para las naciones más pequeñas y menos desarrolladas que puedan tener mercados nacionales más limitados. Además, el acceso al mercado a través de acuerdos comerciales abre nuevas oportunidades de negocio, atrae a la inversión extranjera y apoya el desarrollo económico. De esta manera, aquellos países con políticas abiertas y favorables al comercio tenderán a atraer capital, tecnología, conocimientos de gestión y acceso a los mercados mundiales gracias a esta inversión extranjera que se desprende del comercio internacional. Ello desencadena que los países mejoren sus capacidades de producción, fomenten los avances tecnológicos y creen oportunidades de empleo.

De igual manera, el comercio internacional es un incentivo para empresas y países a la hora de diversificar y gestionar los riesgos. La dependencia de un único mercado nacional puede hacer que una empresa sea vulnerable a las fluctuaciones económicas, los cambios normativos o los riesgos específicos del sector. Mientras que, al expandirse a mercados extranjeros, las empresas pueden repartir sus riesgos entre diferentes países, industrias o segmentos de clientes. La diversificación a través del comercio internacional aumenta la resistencia económica y reduce la vulnerabilidad a las fluctuaciones del mercado nacional o a los acontecimientos geopolíticos.

Además, el comercio internacional y el desarrollo sostenible del planeta son conceptos interrelacionados que desempeñan un papel crucial en la configuración de la economía mundial. En el centro del desarrollo sostenible está la idea de equilibrar el crecimiento económico, la protección del medio ambiente y el bienestar social para garantizar la satisfacción de las necesidades de las generaciones presentes y futuras. El comercio internacional, por su parte, implica el intercambio de bienes, servicios e inversiones entre

países, promoviendo la interdependencia y la integración económicas. Cuando se alinean eficazmente, el desarrollo sostenible y el comercio internacional pueden reforzarse mutuamente, creando oportunidades para el crecimiento inclusivo, la protección del medio ambiente y el progreso social.

“La Agenda 2030 para el Desarrollo Sostenible define al comercio internacional como un motor para el crecimiento económico inclusivo y la reducción de la pobreza, que contribuye a la promoción del desarrollo sostenible.” (Tarazona, 2022) Al participar en el comercio, los países pueden acceder a mercados más amplios, diversificar sus economías y crear oportunidades de empleo. El aumento del comercio puede estimular las inversiones en sectores sostenibles, como las energías renovables, las tecnologías verdes y la agricultura respetuosa con el medio ambiente, fomentando la transición hacia una economía con bajas emisiones de carbono y eficiente en el uso de los recursos. Al integrar las consideraciones de sostenibilidad en las políticas comerciales, los gobiernos pueden fomentar modelos de producción y consumo sostenibles, promoviendo el uso eficiente de los recursos y la reducción del impacto medioambiental.

Asimismo, el comercio internacional puede facilitar la transferencia entre países de tecnologías y conocimientos respetuosos con el medio ambiente. A través del comercio, los países pueden acceder a soluciones innovadoras, conocimientos especializados y mejores prácticas que contribuyan al desarrollo sostenible. Los países en desarrollo, en particular, pueden beneficiarse de la transferencia de tecnología y de las iniciativas de creación de capacidad apoyadas por las asociaciones comerciales, lo que les permite saltar las vías tradicionales de desarrollo y adoptar prácticas sostenibles.

Sin embargo, cuando una empresa se embarca en la aventura del comercio internacional se encontrará con una serie de retos potenciales a los que deberá enfrentarse. En primer lugar, la aparición de riesgos políticos y económicos en función del país al que se esté tratando de acceder. La inestabilidad política, los cambios en las medidas tomadas por los gobiernos, las disputas comerciales o las tensiones geopolíticas pueden afectar de manera directa a las operaciones comerciales y crean incertidumbres que se traducen en pérdidas para la empresa.

La elección de los países con los que comercializar es muy importante ya que la estabilidad financiera de la empresa va a ser muy dependiente de las fluctuaciones monetarias y de tipo de cambio, así como de las recesiones económicas de los mercados extranjeros. Además, hay que tener en cuenta las características del mercado al que se pretende acceder, evaluando las condiciones económicas, la competencia existente y comparando en términos de precio y base de clientes las ofertas existentes.

Otros puntos importantes a la hora de seleccionar países con los que comercializar serán tanto la distancia geográfica como la comprensión cultural. Puede parecer que sean dos variables sin importancia, sin embargo, son puntos críticos a la hora de realizar negocios de manera satisfactoria. Por un lado, la distancia geográfica implica que cuanto más cerca esté el país de comercio, menores serán los costes y esfuerzos, además de que el reconocimiento de la marca es mayor cuanto más cerca se está del país de origen, y las preferencias de los clientes también son muy similares. (Gibb, s.f.)

Por otro lado, la economía global moderna es un popurrí de diversas culturas, ideas, costumbres y comportamientos sociales. A medida que el mercado mundial crece y da cabida a empresas y equipos interculturales y multinacionales, la sensibilidad cultural es una prioridad para garantizar el éxito internacional. Comprender el papel y la importancia de la cultura en los negocios internacionales es esencial para superar las barreras culturales que impiden el crecimiento y el desarrollo de las empresas. La adaptación cultural en los negocios internacionales abarca tanto la cultura organizativa como las culturas y tradiciones nacionales. Esto pasa por una serie de retos como comprender las prácticas comerciales locales, identificar las diferencias regionales, adaptar los modelos de negocio al mercado local, así como las políticas de recursos humanos o las técnicas de gestión. (Global HR Practices, s.f.)

Por último, aunque aparentemente el comercio internacional parezca realmente beneficioso para los países, no siempre es así. Esto provoca la aparición de impuestos, aranceles o cuotas que entorpecen la actividad comercial. Estos costes añadidos suelen crearse para proteger industrias incipientes o economías en vías de desarrollo, aunque también los utilizan las

economías más avanzadas con industrias desarrolladas. Las principales razones por las que se imponen aranceles son la protección del empleo doméstico, la protección de los consumidores locales, el desarrollo de industrias nacientes, la seguridad nacional, o como respuesta o represalia ante alguna acción comercial o política perjudicial para el país de destino. (Radcliffe, 2023)

Estas barreras al comercio internacional se pueden traducir en forma de aranceles que pueden ser fijos, se paga una cantidad por unidad, o variables, el precio del arancel es un porcentaje del valor total de la transacción. Sin embargo, las barreras también se pueden traducir como no arancelarias, al no haber ningún pago monetario per se. Se trata de licencias a ciertas empresas o sectores que reciben una excepción ante determinadas restricciones, o cuotas de importación (generalmente están asociadas con la emisión de licencias) donde se permite el comercio internacional, pero con límites en la cantidad del producto o servicio comercializado. (Oftex International Sales, 2022)

Por una parte, a las barreras no arancelarias, hay que añadirle la extensa documentación necesaria para poder realizar una transacción internacional. Partiendo del DUA (Documento Único Administrativo), la declaración de importación o exportación en soporte papel que será requerido en la aduana. Este documento con carácter de declaración tributaria aporta información detallada (valor, cantidad, clase, origen, destino, etc.) acerca del producto de la transacción. Además, será necesario también un certificado EORI, que no es más que la asignación por parte de la autoridad aduanera de un número identificativo único en todo el territorio europeo para empresas y personas registradas como exportadoras e importadores. Adicionalmente, será necesaria más documentación que dependerá del tipo de producto o mercancía a transporte como facturas comerciales, facturas proforma, facturas aduaneras, facturas consulares, certificados de origen, listas de contenido o el cuaderno ATA y CTD, entre otros. (Infoautónomos, 2022)

Por otra parte, a las barreras arancelarias habrá que añadir el impuesto indirecto del territorio de destino. En el caso de España se pagará el IVA para la Península y Baleares, el IGIC en Canarias y el IPSI en Ceuta y Melilla. Este impuesto correspondiente al producto importado

se calculará conforme a su valor. En el caso del IVA se aplicará un 21% sobre el valor del producto, aunque dependiendo de la mercancía se puede designar el 10% del IVA reducido o el 4% del IVA superreducido. (Infoautónomos, 2022)

Así, se puede visualizar como el panorama para iniciar una andadura en el comercio internacional tiene amplios beneficios y ventajas que conllevan a su vez riesgos y retos que pueden parecer desalentadores. Sin embargo, con las estrategias adecuadas y un enfoque proactivo, estos obstáculos pueden superarse, dando lugar a fructíferas oportunidades para las empresas. Para navegar por el complejo panorama del comercio internacional, las empresas deben apostar por la innovación, la adaptabilidad y la colaboración. Aprovechando los avances tecnológicos, manteniéndose informadas sobre los cambios normativos, fomentando relaciones sólidas con los socios e invirtiendo en el desarrollo del talento, las empresas pueden posicionarse para prosperar en el mercado global.

En este ámbito, comprender y modelar con precisión los costes es crucial para las empresas que buscan optimizar sus operaciones y maximizar la rentabilidad. A medida que las empresas navegan por las complejidades de las cadenas de suministro mundiales, las fluctuaciones de los tipos de cambio, las normativas comerciales y la logística del transporte, se hace cada vez más evidente la necesidad de contar con modelos de costes sólidos. Al profundizar en los factores de coste como aranceles, gastos de transporte, derechos de aduana, seguros y tasas administrativas, las empresas pueden obtener información valiosa sobre el verdadero coste de participar en el comercio internacional.

Este proyecto se adentra en el mundo de la modelización de costes en el comercio internacional, explorando diversas fuentes de datos y herramientas analíticas que permiten a las empresas tomar decisiones informadas, racionalizar sus operaciones y obtener una ventaja competitiva en el mercado global.

## **Capítulo 2. FISCALIDAD Y TRIBUTACIÓN EN EL COMERCIO INTERNACIONAL**

El comercio internacional se produce cuando los bienes y servicios cruzan las fronteras nacionales a cambio de dinero o de los bienes y servicios de otra nación. A lo largo de los años ha crecido espectacularmente como reflejo de la mayor demanda de bienes y servicios que no existen o no pueden producirse localmente, reflejando la creciente internacionalización de la producción.

El comercio, por tanto, vincula a los países generando una importante interdependencia económica, política y social. Para la mayoría de los Estados, el comercio es una forma fácil de generar ingresos y puestos de trabajo. Además, para muchas naciones en desarrollo, suele ser un componente fundamental de los planes de desarrollo. Así pues, en una economía política internacional altamente integrada, los Estados se ven obligados a regular el comercio para maximizar sus beneficios y limitar sus costes para sus economías. Al hacerlo, los Estados se reservan el derecho de adoptar diversas políticas para alcanzar estos objetivos, lo que genera no pocas tensiones entre las naciones. Y, sin embargo, las políticas comerciales de un Estado pueden imponer fácilmente costosos problemas de ajuste socioeconómico a otro Estado. Sin un conjunto de normas y procedimientos internacionales, las políticas comerciales de orientación nacional podrían socavar fácilmente toda la estructura de producción y comercio.

### **2.1 *MODELOS DE COMERCIO INTERNACIONAL***

Entre los siglos XVI y XVIII, no existían normas comerciales internacionales. Los primeros Estados europeos trataron de generar excedentes comerciales como fuente de riqueza para los productores locales, la realeza y, más tarde, el Estado burocrático. Para ayudar a las industrias locales a despegar, se desalentaban las importaciones de bienes intermedios si



significaban que la gente compraría importaciones en lugar de comprar bienes producidos localmente.

Hoy en día, la mayoría de los académicos y responsables políticos siguen favoreciendo un sistema de comercio internacional económicamente liberal dentro de un orden que se supone que se está liberalizando o abriendo de par en par. Y, sin embargo, en casos como la crisis financiera mundial de 2008, la mayoría de las naciones tienden a comportarse de forma mercantilista y a adoptar medidas proteccionistas cuando sus intereses nacionales se ven amenazados. (Robinson, Gómez, & Ruiz, 2015)

El trueque de bienes o servicios entre diferentes pueblos es una práctica milenaria, probablemente tan antigua como la historia de la humanidad. El comercio internacional, sin embargo, se refiere específicamente a un intercambio entre miembros de diferentes naciones, y los relatos y explicaciones de dicho comercio comienzan sólo con el surgimiento del Estado-nación moderno al final de la Edad Media europea. A medida que los pensadores políticos y los filósofos comenzaron a examinar la naturaleza y la función de la nación, el comercio con otros países se convirtió en un tema particular de su investigación. No es de extrañar, por tanto, que uno de los primeros intentos de describir la función del comercio internacional se produjera en el seno del pensamiento nacionalista conocido como mercantilismo.

El análisis mercantilista, que alcanzó su máxima influencia en el pensamiento europeo en los siglos XVI y XVII, se centraba directamente en el bienestar de la nación. Insistía en que la adquisición de riqueza, sobre todo en forma de oro, era de suma importancia para la política nacional. Los mercantilistas tomaban las virtudes del oro casi como un artículo de fe, en consecuencia, nunca trataron de explicar adecuadamente por qué la búsqueda de oro merecía tan alta prioridad en sus planes económicos.

El mercantilismo se basaba en la convicción de que los intereses nacionales están inevitablemente en conflicto, de que una nación sólo puede aumentar su comercio a expensas de otras naciones. Así, los gobiernos se vieron abocados a imponer controles de precios y salarios, fomentar las industrias nacionales, promover las exportaciones de productos

acabados y las importaciones de materias primas, limitando al mismo tiempo las exportaciones de materias primas y las importaciones de productos acabados. El Estado se esforzaba por proporcionar a sus ciudadanos el monopolio de los recursos y las salidas comerciales de sus colonias.

La política comercial dictada por la filosofía mercantilista era, en consecuencia, sencilla: fomentar las exportaciones, desalentar las importaciones y llevarse en oro el producto del excedente de exportación resultante. Las ideas de los mercantilistas a menudo eran intelectualmente superficiales y, de hecho, su política comercial puede haber sido poco más que una racionalización de los intereses de una clase mercantil en ascenso que quería mercados más amplios -de ahí el énfasis en la expansión de las exportaciones- junto con la protección contra la competencia en forma de bienes importados.

A mediados del siglo XVIII comenzó a gestarse una fuerte reacción contra las actitudes mercantilistas. En Francia, los economistas conocidos como fisiócratas exigieron la libertad de producción y de comercio. En Inglaterra, el economista Adam Smith demostró en su libro “La riqueza de las naciones” (1776) las ventajas de eliminar las restricciones comerciales. Economistas y hombres de negocios expresaron su oposición a los derechos de aduana excesivamente elevados y a menudo prohibitivos e instaron a la negociación de acuerdos comerciales con potencias extranjeras. Este cambio de actitud condujo a la firma de una serie de acuerdos que encarnaban las nuevas ideas liberales sobre el comercio, entre ellos el Tratado Anglo-Francés de 1786, que puso fin a lo que había sido una guerra económica entre ambos países.

Después de Adam Smith, los principios básicos del mercantilismo ya no se consideraban defendibles. Sin embargo, esto no significaba que las naciones abandonaran todas las políticas mercantilistas. Las políticas económicas restrictivas se justificaban ahora alegando que, hasta cierto punto, el gobierno debía mantener las mercancías extranjeras fuera del mercado nacional para proteger la producción nacional de la competencia exterior. Con este fin, se introdujeron cada vez más gravámenes aduaneros, que sustituyeron a las prohibiciones totales de las importaciones, cada vez menos frecuentes.

Un triunfo de las ideas liberales fue el acuerdo comercial anglo-francés de 1860, que establecía que los derechos protectores franceses debían reducirse a un máximo del 25 por ciento en un plazo de cinco años, con la entrada libre en Gran Bretaña de todos los productos franceses excepto los vinos. A este acuerdo siguieron otros pactos comerciales europeos.

La escuela británica de economía clásica comenzó en gran medida como una reacción contra las incoherencias del pensamiento mercantilista. Adam Smith fue el fundador de esta escuela en el siglo XVIII; su famosa obra, “La riqueza de las naciones” (1776), es en parte un tratado anti mercantilista. En el libro, Smith subraya la importancia de la especialización como fuente de aumento de la producción, y trata el comercio internacional como un caso particular de especialización: en un mundo en el que los recursos productivos son escasos y las necesidades humanas no pueden satisfacerse por completo, cada nación debe especializarse en la producción de bienes que está particularmente bien equipada para producir; debe exportar parte de esta producción, recibiendo a cambio otros bienes que no puede producir tan fácilmente. Smith no se extendió mucho en estas ideas, pero otro economista clásico, David Ricardo, las desarrolló en el principio de la ventaja comparativa, un principio que todavía se encuentra, en gran medida como Ricardo lo explicó, en los libros de texto contemporáneos sobre comercio internacional.

Independientemente de lo que diga la teoría de la ventaja comparativa sobre las virtudes del comercio sin restricciones, todas las naciones interfieren en cierta medida en las transacciones internacionales.

- Se pueden imponer aranceles a las importaciones, en algunos casos tan costosos que impiden completamente la entrada del producto en cuestión.
- Las cuotas pueden limitar el volumen permitido de importaciones. Pueden ofrecerse subvenciones estatales para fomentar las exportaciones.
- Las exportaciones de capital monetario pueden restringirse o prohibirse.
- La inversión extranjera en instalaciones y equipos nacionales puede verse limitada de forma similar.

Estas interferencias pueden ser simplemente el resultado de la defensa de intereses especiales, porque determinados grupos sufren como consecuencia de la competencia de las importaciones, o un gobierno puede imponer restricciones porque se siente obligado a tener en cuenta factores que la ventaja comparativa deja de lado. Es interesante observar que, por lo que respecta a los bienes y servicios, la pauta general de interferencia sigue el viejo dictado mercantilista de desalentar las importaciones y fomentar las exportaciones. (Bertrand, Wonnacott, Allais, Balassa, & Robinson, 2023)

Así, mientras muchos liberales económicos aplauden al GATT (Acuerdo General sobre Comercio y Aranceles) y a la OMC por hacer realidad muchos de los sueños de los partidarios de una mayor apertura del sistema internacional de producción y comercio, aún siguen coexisten varias tendencias contrarias a este objetivo.

Mientras que muchos grupos de interés apoyan el libre comercio y las políticas de la OMC, otros, especialmente en las naciones en desarrollo, son cada vez más críticos. Las dificultades en las negociaciones multilaterales reflejan las tensiones entre el Norte y el Sur, pero también el argumento de las naciones en desarrollo de que la normativa comercial refleja predominantemente los intereses de las naciones industrializadas del Norte.

Los países en desarrollo tienen ahora una influencia cada vez mayor en las negociaciones multilaterales, basada en su importancia para los Estados desarrollados como mercados y fuentes de mano de obra para las empresas transnacionales. Paradójicamente, la globalización se ha visto socavada por fuerzas y políticas económicas que han generado una mayor demanda de protección en las naciones desarrolladas y en desarrollo. En muchos casos, el Estado es relativamente demasiado débil para impedir que los intereses privados desempeñen el papel de guardianes entre los intereses nacionales e internacionales. Pero, al mismo tiempo, muchos Estados son todavía lo bastante fuertes frente a las calamidades internacionales como para rechazar muchas de las fuerzas que debilitarían su poder.

La estructura comercial liberal parece estar dando paso a un sistema comercial gestionado que mezcla prácticas liberales y mercantilistas. (Robinson, Gómez, & Ruiz, 2015)

## **2.2 FISCALIDAD EN EL COMERCIO INTERNACIONAL**

Bajo este marco se encuentra la regulación del comercio internacional que da paso a las distintas formas de comerciar entre los distintos países mezclando prácticas liberales y mercantilistas al mismo tiempo.

La fiscalidad en el comercio internacional desempeña un papel importante en la configuración del panorama económico mundial y repercute en las transacciones transfronterizas. Los gobiernos aplican impuestos al comercio internacional principalmente para recaudar ingresos, proteger las industrias nacionales y regular la actividad económica. El comercio internacional puede estar sujeto a distintos tipos de impuestos, como los derechos de aduana, el impuesto sobre el valor añadido (IVA), los impuestos especiales y los impuestos de exportación e importación.

Los derechos de aduana son una de las formas más comunes de imposición en el comercio internacional. Se aplican a las mercancías importadas cuando cruzan las fronteras nacionales y se basan en factores como el valor, el peso o la cantidad del producto. Los derechos de aduana sirven como fuente de ingresos para los gobiernos y también pueden utilizarse para proteger a las industrias nacionales haciendo que los bienes importados sean relativamente más caros, fomentando así la producción y el consumo nacionales.

Los impuestos sobre el valor añadido (IVA) son otro componente importante de la fiscalidad del comercio internacional. El IVA es un impuesto sobre el consumo que grava el valor añadido en cada etapa de la cadena de suministro. En el comercio internacional, el IVA suele aplicarse a la importación de bienes y servicios y puede reembolsarse o eximirse en el caso de las exportaciones. El IVA constituye una importante fuente de ingresos para los gobiernos y está diseñado para garantizar que el consumo se grave con independencia de que los bienes se produzcan en el país o se importen. (Cámaras de Comercio. Servicios de Estudios, 2003)

Los impuestos especiales gravan bienes específicos como el tabaco, el alcohol o los artículos de lujo. Estos impuestos se utilizan a menudo para desalentar ciertos patrones de consumo o para promover la salud pública. En el comercio internacional, los impuestos especiales

pueden aplicarse a los bienes importados para crear igualdad de condiciones y proteger a las industrias nacionales de la competencia desleal.

En la fiscalidad del comercio internacional influyen diversos factores, como los acuerdos comerciales bilaterales y multilaterales, los bloques comerciales regionales y las organizaciones internacionales. Los acuerdos comerciales suelen incluir disposiciones destinadas a reducir o eliminar determinadas barreras comerciales, incluidos los aranceles y otros impuestos, para promover un comercio más libre y la integración económica. Organizaciones internacionales como la Organización Mundial del Comercio (OMC) y bloques comerciales regionales como la Unión Europea (UE) también desempeñan un papel en la configuración de la fiscalidad del comercio internacional a través de negociaciones, normativas y mecanismos de resolución de disputas.

Así, esto puede tener efectos tanto positivos como negativos. Por un lado, los impuestos generan ingresos para los gobiernos, protegen las industrias nacionales y apoyan las políticas económicas. Por otro lado, una fiscalidad excesiva o proteccionista puede obstaculizar los flujos comerciales, aumentar los costes para las empresas y los consumidores y crear barreras comerciales. Lograr el equilibrio adecuado entre la generación de ingresos, la protección de la industria nacional y la facilitación del comercio internacional es crucial para el crecimiento económico y el desarrollo sostenibles.

### ***2.3 DERECHO TRIBUTARIO INTERNACIONAL***

El Derecho tributario internacional hace referencia al marco jurídico que regula la fiscalidad de las transacciones y actividades transfronterizas entre distintos países. Abarca un complejo conjunto de normas, tratados, acuerdos y reglamentos destinados a evitar la doble imposición, promover la equidad y prevenir la evasión fiscal en el contexto internacional. El derecho fiscal internacional aborda la asignación de derechos tributarios entre países, la determinación de la renta imponible y el intercambio de información entre autoridades fiscales. (Ramírez, 2021)

Uno de los principios fundamentales del derecho fiscal internacional es evitar la doble imposición, que se produce cuando los mismos ingresos o activos están sujetos a impuestos en más de una jurisdicción. Para mitigar esta situación, los países celebran tratados fiscales bilaterales o multilaterales, conocidos como convenios para evitar la doble imposición o convenios fiscales, que establecen normas para asignar derechos fiscales y resolver conflictos. Estos tratados suelen incluir disposiciones para evitar la doble imposición mediante mecanismos de eliminación o reducción de las obligaciones fiscales a través de métodos como los créditos o las exenciones fiscales. (Morales, 2020)

Otro aspecto clave de la legislación fiscal internacional es la prevención de la evasión fiscal y las prácticas agresivas de planificación fiscal. Los gobiernos colaboran a través de iniciativas internacionales como la Organización para la Cooperación y el Desarrollo Económico (OCDE) y el G20 para desarrollar y aplicar medidas que combatan las prácticas fiscales perjudiciales y garanticen la transparencia. Estas iniciativas incluyen el proyecto BEPS (*Base Erosion and Profit Shifting*), cuyo objetivo es abordar las estrategias de elusión fiscal empleadas por las empresas multinacionales, y el CRS (*Common Reporting Standard*), que facilita el intercambio automático de información sobre cuentas financieras entre las jurisdicciones participantes. (Agbonlahor, 2015)

La fijación de precios de transferencia es otro ámbito fundamental de la legislación fiscal internacional. Se refiere a la fijación de precios de las transacciones entre entidades relacionadas dentro de las empresas multinacionales. El principio de plena competencia es un principio rector a la hora de determinar los precios de transferencia adecuados, garantizando que las transacciones se valoren como lo harían entre entidades no relacionadas. Los países tienen normas específicas y requisitos de documentación relacionados con los precios de transferencia para evitar el desplazamiento de beneficios y garantizar la asignación justa de los ingresos imposables.

La legislación fiscal internacional también incluye normas relativas a la residencia fiscal y a la imposición de las rentas extranjeras. Cada país tiene sus propios criterios para determinar el estatus de residencia fiscal, lo que afecta al alcance de la tributación. Muchos países

aplican un sistema fiscal territorial, que grava las rentas obtenidas dentro de su jurisdicción, mientras que otros aplican un sistema fiscal mundial, en el que los residentes están sujetos a tributación por sus rentas globales. Los convenios fiscales suelen establecer normas de desempate para resolver los conflictos de residencia y evitar la doble imposición.

La aplicación de la legislación fiscal internacional se basa en la cooperación y el intercambio de información entre las autoridades fiscales. Los países han celebrado acuerdos y adoptado mecanismos para facilitar el intercambio de información relacionada con los impuestos, como el intercambio de información sobre cuentas financieras en el marco del CRS. Las organizaciones internacionales, como la OCDE y el Fondo Monetario Internacional (FMI), desempeñan un papel vital en el desarrollo y la promoción de normas para el intercambio de información y ayudan a los países a aplicarlas.

## **2.4 DOCUMENTO ÚNICO ADMINISTRATIVO**

El DUA, o Documento Único Administrativo, es un documento aduanero que se utiliza en el comercio internacional para realizar operaciones de importación y exportación de mercancías. Este documento es esencial para el proceso de despacho aduanero y permite a los importadores y exportadores cumplir con las obligaciones legales y fiscales en el proceso de comercio internacional.

Se trata de un documento que contiene información detallada sobre la mercancía que se importa o exporta, como el origen y el destino de la mercancía, su valor y peso, el modo de transporte y los impuestos que se deben pagar. También incluye información sobre la empresa importadora o exportadora, así como sobre los intermediarios que participan en la operación.

Además, el DUA es obligatorio para todas las operaciones de importación y exportación que se realizan en el territorio de la Unión Europea, aunque es importante señalar que puede presentarse en formato físico o electrónico, dependiendo del país y de las normas específicas de cada operación.



El Documento Único Administrativo sirve para diferentes propósitos en el comercio internacional.

En primer lugar, es un documento aduanero obligatorio que debe ser presentado a las autoridades aduaneras antes de que se pueda liberar la mercancía importada o exportada.

Además de su función aduanera, el DUA también sirve como documento de declaración fiscal para el pago de impuestos y aranceles que permite a los importadores y exportadores cumplir con las obligaciones fiscales establecidas por las autoridades aduaneras y tributarias del país de origen y de destino.

Otro propósito es el control de las estadísticas comerciales. El DUA proporciona información detallada sobre las operaciones de importación y exportación que se realizan en un país y es utilizado por las autoridades para la elaboración de estadísticas comerciales que permiten a los gobiernos tomar decisiones importantes en materia económica.

Asimismo, también sirve como documento de transporte y garantiza que la mercancía llegue a su destino final sin problemas. Contiene información sobre el modo de transporte, el itinerario, el embalaje y la identificación de la mercancía, lo que permite a los intermediarios del transporte garantizar la correcta entrega de la mercancía. (Across Logistics, 2023)

A pesar de ser un proceso burocrático de control aduanero, el DUA puede facilitar a las empresas y comerciantes información valiosa acerca de por donde optimizar las transacciones internacionales. De un DUA se puede extraer información acerca de los impuestos y aranceles a pagar, al mismo tiempo que podemos conocer el código TARIC con que se ha registrado la mercancía, el Incoterm que va a regir la transacción, el circuito seguido por la mercancía en la aduana de entrada o salida, así como el país de origen o destino. A continuación se detalla información acerca de estos términos.

### **2.4.1 EL CÓDIGO TARIC**

El código TARIC (*Tarif Intégrée de la Communauté*) es un sistema utilizado en la Unión Europea (UE) para la clasificación de mercancías en el comercio internacional. Es un sistema

armonizado de clasificación de productos que combina el Arancel Aduanero Común (AAC) de la UE y las medidas adicionales de la UE, como los derechos antidumping y los contingentes arancelarios.

Se trata de un código numérico de diez dígitos que proporciona una descripción detallada de las mercancías, incluida información sobre derechos de aduana, medidas comerciales y otras normativas aplicables. Los seis primeros dígitos del código TARIC corresponden al código del Sistema Armonizado (SA), que es un sistema de clasificación mundial utilizado por la mayoría de los países para normalizar la clasificación de las mercancías. (Veigler Business School, 2022)

Se utiliza a efectos aduaneros en la UE, facilitando la recaudación de derechos de aduana, las estadísticas de importación y exportación y la aplicación de diversas medidas de política comercial. Ayuda a determinar los tipos de derechos de aduana aplicables, los regímenes comerciales preferenciales y cualquier medida comercial adicional que pueda aplicarse a mercancías concretas.

Las empresas que participan en el comercio internacional dentro de la UE deben utilizar el código TARIC para clasificar correctamente sus productos a efectos aduaneros. El código es importante para determinar el importe exacto de los derechos de aduana que deben pagarse, cumplir la normativa sobre importación y exportación y cumplir los requisitos de documentación aduanera.

Para encontrar el código de un producto concreto, las empresas pueden consultar la base de datos TARIC, de acceso público en el sitio web de la Comisión Europea. La base de datos permite a los usuarios buscar códigos a partir de descripciones de productos, palabras clave o códigos del SA, y ofrece información sobre derechos de aduana, medidas comerciales y requisitos adicionales relacionados con mercancías específicas.

Como se ha comentado anteriormente el código TARIC es específico de la Unión Europea y sus Estados miembros. Otros países pueden tener sus propios sistemas de clasificación

aduanera, como el *Harmonized Tariff Schedule* (HTS) utilizado en Estados Unidos o el *Customs Tariff* utilizado en Canadá.

Elegir con precisión el código TARIC correcto para sus mercancías es crucial para garantizar el cumplimiento de la normativa aduanera y determinar los derechos de aduana y las medidas comerciales adecuadas.

En primer lugar, es esencial conocer a fondo el producto que va a clasificar. Habrá que considerar su composición, características, función y uso previsto. Es importante revisar las especificaciones técnicas, la documentación sobre el producto o las normas industriales que le ayuden a comprender su clasificación.

Se deberá consultar la base de datos TARIC de la Comisión Europea que permitirá buscar el código adecuado en función de varios criterios, como la descripción del producto, palabras clave o códigos del SA. (European Commission, s.f.)

Una vez obtenidas las posibles coincidencias, se analizarán las descripciones y atributos de cada código. También es aconsejable revisar cualquier orientación oficial o nota explicativa proporcionada por las autoridades aduaneras o la Comisión Europea en relación con la clasificación de productos específicos. Estos recursos ofrecen a menudo aclaraciones y perspectivas adicionales sobre la correcta clasificación de las mercancías.

Para mayor precisión se recomienda hacer referencias cruzadas con productos o mercancías similares en la base de datos TARIC buscando productos que se parezcan en términos de función, composición o uso previsto.

#### **2.4.2 INCOTERMS**

Siguiendo con los términos de un DUA, se encuentran los Incoterms, abreviatura de *International Commercial Terms*. Se trata de un conjunto de términos comerciales estándar reconocidos internacionalmente que definen los derechos y obligaciones de compradores y vendedores en los contratos comerciales internacionales. Aportan claridad y uniformidad en la interpretación de elementos clave de la transacción, como el reparto de costes, la

asignación de riesgos y el momento en que la responsabilidad se transfiere del vendedor al comprador.

Refiriéndose en sus contratos a un término de la ICC (Cámara de Comercio Internacional), el comprador y el vendedor reducen la incertidumbre propia a toda transacción internacional, como prácticas comerciales e interpretaciones diferentes de un país al otro. Vuelven más precisas sus responsabilidades y obligaciones respectivas para la entrega de mercancías y las obligaciones de documentos del vendedor. De esta forma los Incoterms, aunque son facultativos, son cláusulas estandarizadas y reconocidas que permiten evitar litigios, distribuyendo claramente entre el comprador y el vendedor: las obligaciones, los riesgos, los gastos. Además, los Incoterms distinguen el tema de la transferencia de riesgos y la transferencia de propiedad, ésta última sometida a la ley que rige el contrato. Concretamente, en un contrato de venta internacional los Incoterms clarificarán los siguientes puntos:

- ▶ Situar el punto crítico de la transferencia de los riesgos del vendedor al comprador en el proceso de envío de mercancías (riesgos de pérdida, deterioro, hurto de las mercancías), permitiendo así que quien asume estos riesgos pueda tomar medidas preventivas, sobre todo en cuanto a seguros.
- ▶ Indicar quién (el vendedor o el comprador) debe cerrar el contrato de transporte.
- ▶ Repartir entre ambas partes los gastos logísticos y administrativos en las diferentes etapas del proceso.
- ▶ Precisar quién está a cargo del embalaje, el marcado de la mercancía, las operaciones de mantenimiento, de carga y descarga de las mercancías, o la carga o transferencia del contenido de los contenedores, al igual que las operaciones de inspección.
- ▶ Fijar las obligaciones respectivas para la realización de las formalidades de exportación y/o importación, el pago de los derechos e impuestos de importación, al igual que el suministro de los documentos.

La elección del Incoterm hace plenamente parte de la negociación comercial. Debe efectuarse según las capacidades de organización de la empresa, el medio de transporte utilizado, el nivel de servicio que se desea aportar al cliente o recibir del proveedor, o según

las costumbres del mercado, las prácticas de la competencia, etc. El Incoterm elegido debe adaptarse tanto a las mercancías que serán despachadas como al medio de transporte. (Santander Trade Markets, 2020)

Una entrega directa a la salida significa que la mercancía viaja por cuenta y riesgo del comprador, es decir:

- ▶ desde su disponibilidad en los locales del vendedor (EXW)
- ▶ desde el momento en que fue entregada al transportista para ser despachada (FCA, FAS, FOB, CFR, CIF, CPT y CIP)

Los Incoterms de entrega directa a la salida hacen que el comprador (en una proporción mayor o menor) asuma los costes y riesgos inherentes al transporte de mercancías.


Una venta en la llegada significa que la mercancía viaja por cuenta y riesgo del vendedor hasta el punto/puerto convenido. Esto corresponde a 3 Incoterms:

- ▶ hasta el final del transporte marítimo y del desembarco (DAP);
- ▶ hasta el punto de destino (DPU, DDP). (Santander Trade Markets, 2020)

En la ilustración 7 se muestra una infografía a modo resumen de los distintos Incoterms 2020, se incluirá un mayor detalle de cada uno en el Anexo.

## Reglas Incoterms®

EDICIÓN 2020



DESCRIPCIÓN	Vendedor	Transportista	Franco al costado del buque	A bordo	A la llegada	Franco al costado del buque	Destino	Comprador	FLETE/ TRANSMISIÓN DEL RIESGO	MÁS DETALLES
<b>EXW</b> Ex Works	VENDEDOR							COMPRADOR	<b>Flete</b> Instalaciones del Vendedor <b>Riesgo</b> Instalaciones del Vendedor	El vendedor solo es responsable de entregar las mercancías en sus instalaciones propias o externas designadas por el mismo. El comprador asume todo el riesgo desde allí hasta el destino.
<b>FCA</b> Free Carrier	VENDEDOR							COMPRADOR	<b>Flete</b> Transportista <b>Riesgo</b> Transportista	Necesario concretar punto de entrega de la mercancía. Si el mismo es en las instalaciones del vendedor, la mercancía se entrega con su carga en los medios de transporte dispuestos por el comprador. Si el lugar designado es otro, la mercancía se entrega cuando está preparada para la descarga sobre los medios de transporte del vendedor en el punto acordado. El riesgo se transmite con la entrega de la mercancía.
<b>CPT</b> Carriage Paid to	VENDEDOR							COMPRADOR	<b>Flete</b> Destino <b>Riesgo</b> Primer transportista	El vendedor paga el transporte hasta el destino acordado, pero entrega las mercancías al comprador (con transmisión del riesgo) poniéndola en poder del porteador contratado por dicho vendedor.
<b>CIP</b> Carriage and Insurance Paid to	VENDEDOR							COMPRADOR	<b>Flete</b> Destino <b>Riesgo</b> Primer transportista	El vendedor paga el transporte hasta el destino acordado, pero entrega las mercancías al comprador (con transmisión del riesgo) poniéndola en poder del porteador contratado por dicho vendedor. Este Incoterm exige al vendedor que contrate una cobertura de seguro en las condiciones ICCA o similar. <small>*En los Incoterms 2010® la cobertura de seguro limitada era sufiatente</small>
<b>DAP</b> Delivered at Place	VENDEDOR							COMPRADOR	<b>Flete</b> Lugar de destino <b>Riesgo</b> Mercancía en medios de transporte en destino	El vendedor entrega la mercancía, con transmisión del riesgo, cuando ésta se pone a disposición del comprador en los medios de transporte preparada para la descarga en el lugar de destino acordado.
<b>DPU</b> <sup>New</sup> Delivered at Place Unloaded	VENDEDOR							COMPRADOR	<b>Flete</b> Destino <b>Riesgo</b> Destino	El vendedor entrega la mercancía, con transmisión del riesgo, cuando ésta, una vez descargada de los medios de transporte, se pone a disposición del comprador en el lugar de destino acordado.
<b>DDP</b> Delivered Duty Paid	VENDEDOR							COMPRADOR	<b>Flete</b> Destino <b>Riesgo</b> Destino	El vendedor entrega la mercancía, con transmisión del riesgo, cuando la pone a disposición del comprador despachada para la importación en los medios de transporte de llegada preparada para la descarga en el lugar de destino acordado.
<b>FAS</b> Free Alongside Ship	VENDEDOR							COMPRADOR	<b>Flete</b> Al costado del buque en el puerto de salida <b>Riesgo</b> Al costado del buque en el puerto de salida	El vendedor entrega la mercancía al comprador cuando la mercancía se coloca al costado del buque (o barcaza) designados por el comprador en el puerto de embarque. A partir de este punto, el riesgo lo asume el comprador.
<b>FOB</b> Free on board	VENDEDOR							COMPRADOR	<b>Flete</b> A bordo del buque <b>Riesgo</b> A bordo del buque	El vendedor entrega la mercancía a bordo del buque designado por el comprador en el puerto de embarque. El riesgo se transmite cuando las mercancías están a bordo del buque.
<b>CFR</b> Cost and Freight	VENDEDOR							COMPRADOR	<b>Flete</b> Puerto de destino <b>Riesgo</b> A bordo del buque	El vendedor asume el coste y el flete, derechos no pagados, hasta el puerto de destino convenido. La entrega se produce y el riesgo se transmite cuando las mercancías están a bordo del buque.
<b>CIF</b> Cost, Insurance and Freight	VENDEDOR							COMPRADOR	<b>Flete</b> Puerto de destino <b>Riesgo</b> A bordo del buque	El vendedor asume el coste del seguro y flete, derechos no pagados, hasta el puerto de destino convenido. La entrega se produce y el riesgo se transmite cuando las mercancías están a bordo del buque. Este Incoterm exige al vendedor que contrate una cobertura de seguro limitada en las condiciones ICCA o similar.

El lugar debe especificarse siempre con la mayor precisión posible, no solo un nombre de ciudad.

**COSTE** **RIESGO** **SEGURO** **TRANSPORTE POR MAR Y VÍAS NAVEGABLES INTERIORES**

*Ilustración 7 - Incoterms 2020 (DSV, 2020)*

### **2.4.3 TIPOS DE CIRCUITOS**

El despacho aduanero es el proceso que engloba todos los trámites que es necesario realizar y la documentación que debe presentarse ante las autoridades aduaneras, para introducir o exportar mercancías de un territorio en el marco de una operación comercio internacional.

Por lo general se trata de poner a disposición de la aduana una serie de documentos relativos tanto al importador y exportador, como a la mercancía y su origen y destino. La documentación solicitada puede variar, por ejemplo, en función de los países de origen o destino o del tipo de mercancía a transportar.

En el momento en que se solicita el despacho aduanero de una mercancía, la autoridad aduanera le asigna un circuito de entrada, que implicará la necesidad o no de una revisión posterior. La asignación de uno u otro canal se realiza combinando modelos estadísticos y otros parámetros que también incluyen controles aleatorios. Por este motivo esta asignación dependerá de varios factores como el origen y destino de la mercancía, el historial del exportador o importador, el hecho de que se trate de un operador con certificado OEA o en el peor de los casos por puro azar. (Stock logistic, 2018)

#### ***2.4.3.1 El circuito verde***

El más beneficioso de los circuitos aduaneros. Permite obtener el levante de las mercancías inmediatamente y que estas puedan retirarse y poner rumbo al destino, tanto en importaciones como en exportaciones. (Arrondo, 2019)

#### ***2.4.3.2 El circuito naranja***

El circuito naranja implica la presentación de la documentación comercial a través de la web de Agencia Tributaria, accediendo al número MRN (o número de DUA) asignado en el despacho. Una vez el inspector de aduanas a quien han asignado el despacho ha analizado la documentación aportada, devuelve una de estas posibles respuestas:



- ▶ Levante de las mercancías, pues ha comprobado que la documentación es correcta y el despacho se ha realizado conforme a dicha documentación.
- ▶ Requerimientos. Solicitudes por parte de la administración de aportar más documentación que disipe las dudas sobre las mercancías o la declaración (certificados para aduaneros, declaraciones del importador/exportador de índole técnica, imágenes del producto, entre otros).
- ▶ Inspección física de las mercancías. Esto convertiría el circuito naranja en uno rojo. La administración considera insuficientes los documentos o aclaraciones presentados y desea ver las mercancías. (Arrondo, 2019)

#### ***2.4.3.3 El circuito rojo***

Este es el más problemático de los circuitos aduaneros. Implica la obligatoria presentación de la documentación comercial (al igual que en el circuito naranja) pero conlleva, salvo en situaciones muy concretas, la inspección física de las mercancías.

Al presentar la documentación a través de la web, el inspector asignado comprueba la documentación y emite un rechazo al agente de aduanas con una de las siguientes opciones:

- ▶ Solicita el posicionamiento en terminal, en un punto de inspección fronterizo (PIF), para inspeccionar las mercancías. En este punto, el inspector realiza la apertura del contenedor, desprecintando el mismo, y con una simple vista puede comprobar el etiquetado o marcaje de las cajas, puede realizar un “pasillo” para extraer una muestra al azar u optar por el traslado del contenedor a otro punto más seguro o donde la inspección pueda ser posible.
- ▶ Solicita el traslado a un almacén autorizado para realizar la inspección, esto lleva implícito el vaciado del contenedor, además de todos los costes de descarga, desplazamiento, costes del propio almacén, etc.
- ▶ Solicita el control del contenedor mediante escáner. Esta es la mejor de las situaciones dentro de un circuito rojo. El levante de las mercancías está condicionado a la conformidad del escáner. Si éste es conforme, la mercancía puede salir



inmediatamente a su destino. En caso de no conformidad, se procederá a la inspección física más pormenorizada de las mercancías. (Arrondo, 2019)

#### ***2.4.3.4 El circuito amarillo***

Este circuito no pertenece necesariamente a los circuitos aduaneros, pues aparece cuando las mercancías todavía no han llegado a puerto o no tienen su declaración sumaria datada y, por lo tanto, no están bajo vigilancia aduanera. Se trata de un circuito para aduanero, que busca que las partidas pre declaradas cumplan y cuenten con los certificados requeridos. (Arrondo, 2019)

## **Capítulo 3. ESTADO DE LA CUESTIÓN**

A medida que las cadenas de suministro se alargan y se interrelacionan cada vez más, su gestión se ha convertido en una tarea considerablemente difícil. Además, factores externos como la volatilidad del mercado, las fluctuaciones inducidas por pandemias y una mayor atención al impacto medioambiental de las cadenas de suministro han desencadenado la regionalización y creado una demanda de optimización. Curiosamente, esto ha estimulado la llegada de la tecnología comercial para ayudar a las PYME y a los *stakeholders* a centrarse en la resistencia de la cadena de suministro para hacer que el comercio mundial sea más eficiente, inclusivo y equitativo.

Estas tecnologías no solo promueven plataformas de descubrimiento de compradores y vendedores, sino que también permiten a los proveedores de servicios, como los financiadores del comercio, los proveedores de seguros y las empresas de logística, comunicarse entre sí y garantizar la interoperabilidad dentro del ecosistema. En pocas palabras, la tecnología tiene el poder de hacer que el comercio transfronterizo sea más eficiente, inclusivo y equitativo para las PYME de todo el mundo. (Mukewar, 2022)

### **3.1 INTELIGENCIA ARTIFICIAL**

La Inteligencia Artificial (IA) ha logrado avances significativos en varios sectores, incluida la gestión de la cadena de suministro. Las tecnologías de IA se están utilizando para optimizar y agilizar las operaciones de la cadena de suministro, mejorar la eficiencia y potenciar los procesos de toma de decisiones. A continuación, se exponen algunos ejemplos de aplicaciones de la IA tanto en el comercio internacional como en la cadena de suministro de una empresa en su conjunto.

Un área en la que la IA está demostrando su valor es la previsión de la demanda. Los algoritmos pueden analizar grandes cantidades de datos históricos, tendencias del mercado y factores externos para proporcionar previsiones precisas de la demanda. Esto ayuda a

optimizar los niveles de inventario, reducir las roturas de stock y mejorar la satisfacción general del cliente.

Al predecir los patrones de demanda, los plazos de entrega y los puntos óptimos de reposición, la IA puede optimizar los niveles de inventario lo que permite a las empresas reducir el exceso de inventario, minimizar los costes de mantenimiento y mejorar la rotación del inventario.

En relación a esto, en la gestión de almacenes se automatizan diversas tareas como el seguimiento de inventarios, la preparación de pedidos y la reposición. Los robots y vehículos autónomos equipados con capacidades de IA pueden navegar por los almacenes, localizar artículos y realizar tareas de forma eficiente, reduciendo errores y costes laborales.

La planificación y optimización de la cadena de suministro son otras áreas clave en las que destaca la IA. Pueden generar programas de producción optimizados, planes de ruta y estrategias de distribución teniendo en cuenta factores como los costes de transporte, las capacidades de producción y las necesidades de los clientes mejorando la eficiencia operativa.

Asimismo, las herramientas de análisis basadas en IA proporcionan visibilidad en tiempo real de la cadena de suministro al agregar y analizar datos de diversas fuentes. Esto permite a las empresas realizar un seguimiento de los envíos, supervisar el rendimiento de los proveedores e identificar cuellos de botella o interrupciones de forma proactiva.

El mantenimiento predictivo es otra área en la que la IA aporta importantes beneficios. Mediante el análisis de los datos de los sensores de los equipos y la maquinaria, los sistemas pueden predecir y prevenir fallos ayudando a las empresas a minimizar el tiempo de inactividad, optimizar los programas de mantenimiento y reducir los costes generales de mantenimiento.

Gracias a la IA y en relación con el IoT se consigue que vehículos autónomos y drones transporten mercancías dentro de la cadena de suministro. Estas tecnologías permiten una

planificación eficiente de las rutas, la detección de obstáculos y la toma de decisiones adaptativa, mejorando la velocidad de entrega y reduciendo los costes.

La gestión de riesgos es otro apartado en el que la IA es capaz de convertirla en más robusta analizando datos de diversas fuentes, como redes sociales, noticias e informes meteorológicos. Esto ayuda a identificar posibles riesgos e interrupciones, lo que permite a las empresas tomar medidas proactivas para mitigarlos y garantizar la continuidad del negocio.

Además, se están utilizando *chatbots* y asistentes virtuales potenciados por IA para gestionar las consultas de los clientes, hacer un seguimiento de los pedidos y proporcionar información en tiempo real. Estas herramientas mejoran la experiencia del cliente, reducen los tiempos de respuesta y liberan recursos humanos para tareas más complejas.

### **3.2 BLOCKCHAIN**

La tecnología *Blockchain* tiene el potencial de transformar el comercio internacional mejorando la transparencia, la seguridad y la eficiencia en toda la cadena de suministro. *Blockchain* permite la creación de un libro de contabilidad inmutable y descentralizado donde se pueden registrar las transacciones y los datos relacionados con el movimiento de mercancías. Esto garantiza la transparencia y la rendición de cuentas en toda la cadena de suministro, permitiendo a las partes interesadas rastrear y verificar el origen, la ubicación y el estado de las mercancías en cada etapa.

Además, esta tecnología puede ayudar en los procesos aduaneros y de cumplimiento de la normativa proporcionando un registro seguro e inmutable del origen de las mercancías, certificaciones e información relacionada con el cumplimiento. Esto puede agilizar el despacho de aduanas, reducir los retrasos y mejorar el cumplimiento normativo a través de las fronteras.

Otra ventaja significativa es el uso de contratos inteligentes. *Blockchain* facilita el uso de acuerdos autoejecutables con reglas y condiciones predefinidas. Estos contratos automatizan

y hacen cumplir los términos y condiciones entre las partes implicadas en el comercio internacional, como compradores, vendedores y compañías navieras. Los contratos inteligentes pueden agilizar y acelerar procesos como el despacho de aduanas, la liquidación de pagos y la validación de documentación.

Aparte de los contratos inteligentes, la tecnología *blockchain* permite la digitalización de muchos otros documentos utilizados durante toda la transacción comercial. Tradicionalmente, el comercio internacional implicaba numerosos documentos en papel, como conocimientos de embarque, facturas y certificados de origen. *Blockchain* permite la digitalización y el almacenamiento seguro de estos documentos, reduciendo el papeleo, eliminando el riesgo de fraude y permitiendo una verificación e intercambio de documentos más rápidos.

En relación al punto anterior, otro aspecto importante es la mayor seguridad que ofrece *blockchain*. Cuenta con sólidas características de seguridad, como el cifrado de datos, el almacenamiento descentralizado y los mecanismos de consenso. Estas características dificultan que agentes malintencionados manipulen o alteren los registros de transacciones, garantizando la integridad y autenticidad de la información relacionada con el comercio.

Relacionado con la seguridad está la prevención de falsificaciones. *Blockchain* puede utilizarse para rastrear y autenticar productos a lo largo de la cadena de suministro, reduciendo el riesgo de que mercancías falsificadas entren en el comercio internacional. Al registrar la información del producto, incluidos los números de serie, los detalles de fabricación y el historial de propiedad, *blockchain* permite la trazabilidad y la verificación de la autenticidad.

Esta tecnología también puede tener un impacto significativo en la financiación del comercio. Puede agilizar y mejorar el acceso de las pequeñas y medianas empresas (PYME) a la financiación del comercio proporcionando un registro transparente y auditable de las transacciones. Al compartir datos comerciales en la cadena de bloques, las PYME pueden crear un historial crediticio verificable y mitigar los riesgos para los prestamistas, mejorando así su acceso a la financiación.

Además, la naturaleza transparente e inmutable de la cadena de bloques puede simplificar la resolución de litigios en el comercio internacional. Todos los datos y la documentación pertinentes de las transacciones se almacenan de forma segura en la cadena de bloques, lo que permite una resolución eficaz y auditable de los litigios al proporcionar una fuente irrefutable de la verdad.

Asimismo, también facilita la colaboración y el intercambio de datos entre las múltiples partes interesadas en las redes de comercio internacional. Permitiendo un acceso seguro y autorizado a los datos pertinentes, así las partes pueden intercambiar información y coordinarse con mayor eficacia.

Por último, *blockchain* puede apoyar las iniciativas de sostenibilidad en el comercio internacional. Permite la trazabilidad y la verificación de las prácticas sostenibles a lo largo de la cadena de suministro. Permite a consumidores y empresas rastrear y validar el impacto medioambiental y social de los productos, fomentando la transparencia y el abastecimiento responsable.

Aunque la cadena de bloques es muy prometedora para el comercio internacional, su adopción generalizada se enfrenta a retos como la interoperabilidad entre diferentes plataformas de cadena de bloques, los marcos jurídicos y normativos y la necesidad de colaboración en todo el sector. Sin embargo, el potencial de la tecnología para mejorar la confianza, la eficiencia y la seguridad hace que sea un área de creciente interés y exploración dentro del sector comercial.

### **3.3 IoT**

El Internet de los objetos (IoT) está desempeñando un papel transformador en las cadenas de suministro al conectar objetos y dispositivos físicos a Internet, lo que les permite recopilar e intercambiar datos. La tecnología IoT proporciona visibilidad en tiempo real, aumenta la automatización y mejora la capacidad de toma de decisiones en toda la cadena de suministro. A continuación, se exponen distintas aplicaciones para el comercio, si bien, muchas de ellas

están estrechamente relacionadas con las ya mencionadas en el apartado de la inteligencia artificial.

Una aplicación significativa de IoT en las cadenas de suministro es la gestión de inventarios. Los sensores IoT pueden desplegarse en almacenes, instalaciones de almacenamiento e incluso en productos o paquetes individuales para realizar un seguimiento automático de los niveles de inventario. Estos sensores pueden controlar en tiempo real factores como los niveles de existencias, las fechas de caducidad y las condiciones ambientales (por ejemplo, temperatura, humedad). Esto permite una gestión proactiva del inventario, reduce las roturas de existencias, minimiza el exceso de inventario y mejora la eficiencia general de la cadena de suministro.

Los dispositivos IoT, como los rastreadores GPS y las etiquetas RFID, también desempeñan un papel crucial en el seguimiento de activos. Al fijar estos dispositivos a contenedores de envío, palés o vehículos, es posible realizar un seguimiento en tiempo real de sus ubicaciones y movimientos. Esto permite un seguimiento y control precisos de las mercancías a lo largo de la cadena de suministro, proporcionando visibilidad y facilitando la toma de decisiones a tiempo.

La monitorización del estado es otra área en la que brilla IoT. Los sensores IoT pueden supervisar el estado y el rendimiento de los equipos, la maquinaria y los vehículos de la cadena de suministro. Al medir factores como la temperatura, la presión, la vibración o el consumo de energía, estos sensores proporcionan datos para detectar anomalías, predecir las necesidades de mantenimiento y optimizar los programas de mantenimiento. Esto ayuda a reducir el tiempo de inactividad y mejorar la eficiencia operativa.

La tecnología IoT permite una visibilidad integral de la cadena de suministro mediante la recopilación de datos en diversos puntos de contacto, como la fabricación, el almacenamiento, el transporte y la venta al por menor. Al facilitar la supervisión en tiempo real del inventario, los envíos y las condiciones ambientales, IoT mejora la visibilidad y ayuda a identificar cuellos de botella, optimizar procesos y mejorar el rendimiento general de la cadena de suministro.

Los datos generados por los dispositivos IoT pueden combinarse con análisis avanzados y técnicas de aprendizaje automático para el análisis predictivo. Mediante el análisis de datos históricos y en tiempo real, las empresas pueden predecir patrones de demanda, identificar posibles interrupciones, optimizar el enrutamiento y la programación, y tomar decisiones basadas en datos para mejorar la eficiencia y la satisfacción del cliente.

IoT también permite operaciones de almacenamiento inteligentes. Al automatizar y optimizar varios procesos, los dispositivos IoT contribuyen a una gestión eficiente del almacén. Por ejemplo, pueden automatizar la reposición de inventarios, hacer un seguimiento de los activos dentro de los almacenes, optimizar la utilización del espacio y habilitar sistemas autónomos de manipulación de materiales. Esto mejora la precisión de los pedidos, acelera el cumplimiento, reduce los costes de mano de obra y mejora la productividad del almacén.

En industrias como la farmacéutica y la alimentaria, es fundamental mantener unas condiciones de temperatura adecuadas durante el transporte y el almacenamiento. Los sensores IoT desempeñan un papel crucial en la gestión de la cadena de frío al controlar los niveles de temperatura y humedad en tiempo real. Esto garantiza el cumplimiento de la normativa y los estándares de calidad, previniendo el deterioro, reduciendo los residuos y asegurando la integridad del producto a lo largo de la cadena de frío.

Por último, los dispositivos IoT permiten a las empresas ofrecer experiencias de cliente personalizadas y proactivas. A través de los dispositivos IoT en los hogares, los productos pueden reordenarse automáticamente cuando se agotan las existencias, lo que conduce a una experiencia de cliente cómoda y sin problemas. Los datos de IoT también pueden aprovecharse para obtener información sobre las preferencias y el comportamiento de los clientes, lo que facilita la realización de campañas de marketing específicas y mejora la satisfacción del cliente.



### **3.4 DIGITAL PAYMENTS**

Los pagos digitales han transformado significativamente el panorama del comercio internacional al ofrecer formas rápidas, seguras y cómodas de realizar transacciones financieras transfronterizas. Han surgido varios métodos y plataformas para facilitar los pagos digitales que se exponen a continuación.

Un método común es la transferencia electrónica de fondos (TEF), que consiste en transferir fondos electrónicamente de una cuenta bancaria a otra utilizando sistemas de pago como transferencias bancarias, cámaras de compensación automatizadas (ACH) o plataformas de pago electrónico. Las TEF permiten a las empresas realizar pagos directos y seguros a proveedores, prestadores de servicios o socios comerciales de todo el mundo.

Las pasarelas de pago en línea desempeñan un papel crucial a la hora de facilitar los pagos digitales de bienes y servicios en el comercio internacional. Estas pasarelas procesan de forma segura los pagos realizados a través de tarjetas de crédito, débito o monederos electrónicos, lo que permite a las empresas aceptar pagos de clientes situados en cualquier parte del mundo. Entre las pasarelas de pago en línea más populares están *PayPal*, *Stripe* y *Square*.

Las soluciones de pago por móvil también han ganado terreno en el comercio internacional, sobre todo en regiones con una alta penetración del móvil. Las aplicaciones de pago por móvil y los monederos digitales permiten a empresas y consumidores realizar transacciones utilizando sus teléfonos inteligentes. Estas aplicaciones suelen ofrecer funciones como el escaneo de códigos QR, los pagos sin contacto y la integración con sistemas bancarios, lo que facilita el envío y la recepción de pagos transfronterizos.

Las criptomonedas, como Bitcoin y Ethereum, ofrecen una forma alternativa de pago digital en el comercio internacional. Utilizando la tecnología *blockchain*, las criptomonedas ofrecen ventajas como tiempos de liquidación rápidos y comisiones de transacción reducidas. Algunas empresas y particulares optan por aceptar criptomonedas como método de pago, en particular para las transacciones transfronterizas.

Las plataformas de financiación del comercio ofrecen soluciones de pago seguras y simplificadas para el comercio internacional. Estas plataformas incorporan características como cartas de crédito, financiación de la cadena de suministro y financiación de facturas, facilitando las transacciones internacionales al tiempo que mitigan los riesgos y mejoran la gestión del flujo de caja.

Los sistemas de pago entre particulares (P2P) permiten transferencias directas de fondos entre particulares o empresas, evitando a los intermediarios tradicionales. Estos sistemas utilizan la tecnología y los monederos digitales para facilitar pagos seguros e instantáneos a través de las fronteras, lo que los hace útiles para el comercio internacional a pequeña escala o las transacciones en las que participan autónomos, consultores o proveedores de servicios.

Las plataformas de pago digitales suelen integrar servicios de cambio de divisas, lo que permite a las empresas convertir divisas durante las transacciones internacionales. Estas plataformas ofrecen tipos de cambio competitivos, precios transparentes y plazos de liquidación más rápidos en comparación con los métodos tradicionales de cambio de divisas, lo que reduce los costes y las complejidades asociadas a las conversiones de divisas en el comercio internacional.

El cumplimiento de los requisitos normativos y unas medidas de seguridad sólidas son prioritarios para las plataformas de pago digitales en el comercio internacional. Aplican medidas como procedimientos de conocimiento del cliente (KYC), controles contra el blanqueo de capitales (AML) y tecnologías de cifrado para garantizar transacciones seguras y conformes.

En resumen, los pagos digitales en el comercio internacional ofrecen ventajas como plazos de liquidación más rápidos, menores costes de transacción, mayor seguridad y transparencia.

### **3.5 CLOUD COMPUTING**

La computación en nube se ha convertido en una tecnología fundamental en la gestión de la cadena de suministro, ya que aporta numerosas ventajas y permite mejorar la colaboración, la escalabilidad y la eficiencia.

Una de las principales ventajas de la computación en nube en las cadenas de suministro es su capacidad para almacenar y gestionar datos. En lugar de depender de servidores locales o dispositivos de almacenamiento físico, las empresas pueden almacenar y gestionar sus datos en la nube, lo que ofrece un acceso sencillo, copias de seguridad de los datos y garantiza su integridad, al tiempo que reduce los costes de infraestructura.

Las plataformas basadas en la nube también proporcionan un repositorio centralizado para almacenar y compartir datos de la cadena de suministro entre múltiples partes interesadas. Esto facilita la visibilidad en tiempo real de los niveles de inventario, el estado de los pedidos, los datos de producción y la información logística. Las plataformas de visibilidad de la cadena de suministro basadas en la nube permiten una colaboración sin fisuras, mejoran la toma de decisiones y aumentan la capacidad de respuesta a las demandas de los clientes.

La computación en nube también desempeña un papel vital en la gestión de almacenes e inventarios. Los sistemas agilizan y optimizan las operaciones de almacén proporcionando visibilidad en tiempo real de los niveles de inventario, facilitando una eficiente preparación y reposición de pedidos y automatizando el seguimiento del inventario. La escalabilidad de las soluciones basadas en la nube garantiza que las empresas puedan adaptarse y gestionar la fluctuación de la demanda y la evolución de las necesidades de la cadena de suministro.

Además del almacenamiento de datos y la colaboración, la computación en nube apoya la planificación de la demanda y el suministro. Las soluciones basadas en la nube aprovechan la potencia de cálculo y la capacidad de almacenamiento para procesar y analizar grandes volúmenes de datos, lo que permite realizar previsiones precisas, detectar la demanda y optimizar los procesos de planificación de la cadena de suministro. Las empresas pueden

simular diferentes escenarios, evaluar el impacto de los cambios y tomar decisiones basadas en datos para mejorar la eficiencia operativa y la satisfacción del cliente.

La computación en nube también admite potentes capacidades de análisis e inteligencia empresarial para los datos de la cadena de suministro. Al aprovechar las plataformas analíticas basadas en la nube, las empresas pueden obtener información a partir de grandes cantidades de datos recogidos de diversas fuentes. Esta información ayuda a identificar tendencias, detectar anomalías, optimizar procesos y respaldar la toma de decisiones estratégicas.

Los sistemas ERP basados en la nube son otra aplicación valiosa en la gestión de la cadena de suministro. Proporcionan un conjunto completo de aplicaciones para gestionar diversos aspectos de las operaciones de la cadena de suministro, como el aprovisionamiento, la gestión de inventarios, la planificación de la producción y el cumplimiento de pedidos. Las soluciones ERP basadas en la nube ofrecen flexibilidad, escalabilidad y fácil integración con otras plataformas y tecnologías basadas en la nube.

La gestión del transporte es otro ámbito en el que la computación en nube aporta importantes ventajas. Los sistemas de gestión del transporte (TMS) ayudan a optimizar las operaciones de transporte proporcionando herramientas para la planificación de rutas, la selección de transportistas, la optimización de la carga y el seguimiento de la carga. Permiten la visibilidad en tiempo real, la colaboración con transportistas y socios logísticos y la toma de decisiones basada en datos, lo que reduce los costes de transporte y mejora el rendimiento de las entregas.

### ***3.6 TECH-BASED LOGISTICS***

La tecnología desempeña un papel crucial en las cadenas de suministro modernas, revolucionando las operaciones logísticas e impulsando la eficiencia. Se utilizan diversas soluciones logísticas basadas en la tecnología para agilizar los procesos y mejorar la gestión de la cadena de suministro.

Los sistemas de gestión del transporte (TMS), como ya se ha mencionado anteriormente, son herramientas clave para optimizar las operaciones de transporte. Estas plataformas utilizan datos y análisis en tiempo real para mejorar la planificación de rutas, la consolidación de cargas, la selección de transportistas y el seguimiento de la carga. Al aumentar la visibilidad y la precisión, las plataformas TMS ayudan a reducir los costes de transporte y mejorar la eficiencia de las entregas.

Los sistemas de gestión de almacenes (SGA) automatizan y optimizan las operaciones de almacén. Proporcionan visibilidad en tiempo real de los niveles de inventario, agilizan los procesos de *picking* y embalaje y mejoran la eficiencia general del almacén. Las soluciones de SGA permiten a las empresas gestionar eficazmente el inventario, satisfacer los pedidos y asignar los recursos de mano de obra.

Las tecnologías GPS y telemáticas permiten el seguimiento y la supervisión de los envíos en tiempo real. A través de dispositivos con GPS y plataformas telemáticas, las empresas pueden rastrear la ubicación, el estado y la condición de las mercancías en tránsito. Esto mejora la visibilidad de la cadena de suministro y permite una gestión proactiva de las operaciones logísticas.

La robótica y las tecnologías de automatización están transformando las operaciones logísticas. Los vehículos de guiado automático (AGV) y los sistemas robóticos se encargan de tareas repetitivas en los almacenes, mejorando la eficiencia, reduciendo los costes de mano de obra y minimizando los errores. Los drones y los vehículos autónomos también se están utilizando para la entrega y el transporte de última milla, proporcionando opciones más rápidas y flexibles.

Las tecnologías de Realidad Aumentada (RA) y Realidad Virtual (RV) encuentran aplicaciones en formación, mantenimiento y asistencia remota en las cadenas de suministro. Mejoran la productividad de los trabajadores, aumentan la eficacia de la formación y permiten realizar inspecciones y simulaciones virtuales, reduciendo el tiempo de inactividad y mejorando la eficiencia operativa.

Estas soluciones logísticas basadas en la tecnología han transformado las cadenas de suministro, aumentando la visibilidad, optimizando las operaciones, reduciendo los costes y mejorando la satisfacción del cliente. Las empresas que adoptan e integran eficazmente estas tecnologías obtienen una ventaja competitiva en el dinámico y vertiginoso panorama actual de la cadena de suministro.

## **Capítulo 4. DEFINICIÓN DEL TRABAJO**

### **4.1 JUSTIFICACIÓN**

La idea de analizar y predecir los costes del comercio internacional tratando de optimizar los gastos involucrados en el mismo nace a partir de un proyecto que propone EY para una empresa que realiza asiduamente transacciones internacionales. El proyecto se basa en una automatización del proceso fiscal que conlleva la compraventa de mercancías en el ámbito internacional a través de las DUA, aranceles e impuestos aduaneros.

Así, con este nuevo proyecto no solo se realizaría la automatización del proceso fiscal llevado a cabo cada vez que la empresa realiza una importación o exportación, sino que va más allá y trataría de analizar el sistema de costes de estas transacciones para la empresa correspondiente. Aprovechando que EY tiene acceso a todos los datos disponibles, el modelo tendrá todas las herramientas necesarias para poder identificar todas aquellas áreas de mejora en cuanto a optimización de costes durante el comercio internacional.

El interés por el comercio internacional se combina con la importancia de la optimización de costes como beneficio para la empresa, reduciendo costes y mejorando la eficiencia de sus operaciones comerciales.

La relevancia del uso de la inteligencia artificial es el pilar clave de este proyecto ya que trata de explorar de una forma efectiva las distintas estrategias comerciales. Prediciendo el sistema de costes y gastos, y explotando al máximo los datos disponibles será posible encontrar la forma más eficiente de realizar una importación o exportación.

Por otro lado, se tiene en cuenta también el potencial impacto en la industria. La implementación del modelo ayudará a las empresas a mejorar su competitividad y rentabilidad, generando efectos positivos en la economía en general.

Por último y no por ello menos importante, la implementación de este proyecto promueve de manera indirecta un desarrollo sostenible durante toda la cadena de valor.

- Se enfoca hacia una menor emisión de gases de efecto invernadero en el transporte de mercancías. La optimización de costes ayuda a encontrar rutas más eficientes y reducir el número de viajes necesarios, lo que puede tener un impacto positivo en el medio ambiente.
- Lleva a una mayor utilización de energías renovables en el transporte de mercancías. La búsqueda de rutas y modos de transporte más eficientes y económicos puede llevar a la utilización de medios de transporte más sostenibles, como el transporte ferroviario o el transporte marítimo, que pueden utilizar energías renovables para su funcionamiento.
- Ayuda a encontrar soluciones más eficientes para la gestión de la cadena de suministro, lo que reduce la cantidad de productos que se desperdician o se descartan en el proceso de transporte. Así como reducir el consumo de recursos y a minimizar el impacto ambiental del transporte de mercancías.

En resumen, el desarrollo del proyecto proporciona una ventaja competitiva a la empresa al predecir los costes comerciales de transacciones internacionales, lo que provoca que la compañía se pueda anticipar al mercado tomando decisiones estratégicas en base a los resultados expuestos por el modelo. Además, se produce un beneficio generalizado en la economía en su conjunto a largo plazo, dado el caso de que el modelo se implantase de manera genérica en el sector conllevaría el ahorro de costes que se traduce en inversiones en otros puntos de la cadena de suministro o incluso en otros sectores. Y, por último, se produce un beneficio en cuanto al desarrollo sostenible reduciendo la emisión de gases en las mercancías, una mayor utilización de las energías renovables, y una mayor reducción de los deshechos minimizando el impacto ambiental.



## **4.2 OBJETIVOS**

Los objetivos técnicos de este proyecto son:

- Extracción automática de toda la información relativa a los DUA de las diferentes fuentes de información
- Modelo de visualización y análisis de datos del comercio exterior
- Modelo de *machine learning* con un elevado grado de precisión
- Herramienta de predicción de costes en tiempo real

Los objetivos de negocio que se buscan con la implementación de este proyecto son:

- La reducción de costes del comercio internacional eligiendo la estrategia adecuada.
- Mejora de la eficiencia de los procesos de la empresa.
- Predicción de los costes de una futura transacción internacional.
- Mejora de la competitividad de la empresa en el mercado global.

Si bien, es cierto, que el objetivo principal del proyecto busca la optimización de costes del proceso de compraventa que una empresa emprende durante una transacción de mercancías internacional. Sin embargo, de este objetivo base se desprenden todos los de arriba que finalmente acaban en el objetivo final que es que la empresa se encuentre en una posición más ventajosa que sus competidores, así no solo habremos conseguido reducir costes sino hacer más eficiente toda la cadena de suministro obteniendo mayores ingresos. Al final todos los objetivos están interrelacionados unos con otros con el denominador común de reducir y predecir costes para formular la estrategia más adecuada al mercado internacional.

Además, como objetivo añadido se incluye la búsqueda de una cadena de suministro y transacción internacional de mercancías lo más sostenible posible. Este objetivo es doble ya que no solo busca un desarrollo sostenible amable con el planeta sino también una fuente de ingresos para la compañía que podría beneficiarse de ayudas y contribuciones internacionales por buenas prácticas sostenibles, así como un impacto positivo en el

consumidor y la sociedad poniendo a la empresa, una vez más, en una posición ventajosa en su sector.

### **4.3 METODOLOGÍA**

Al ser un proyecto propio de EY la metodología de trabajo seguida será la misma que la del resto de proyectos que se realizan en el departamento. En este caso se habla de una metodología *agile* para la consecución del proyecto.

La metodología *agile* es un enfoque de la gestión de proyectos que hace hincapié en la flexibilidad, la colaboración y la entrega de valor a los clientes a través de mejoras iterativas. Permite la retroalimentación frecuente y la mejora continua, dando prioridad al valor para el cliente e implicando una estrecha colaboración con los clientes a lo largo de todo el proyecto.

Para definir *Agile*, el método se basa en el concepto de olas continuas o *sprints* de planificación y ejecución de proyectos, lo que permite adaptar y desarrollar continuamente el plan, alcance y diseño a lo largo del proyecto.

De acuerdo con la definición de metodología *Agile*, los proyectos *Agile* requieren un enfoque iterativo, que admite entregas incrementales, frecuentes y coherentes de productos funcionales al cliente. Este enfoque innovador garantiza que el equipo de proyecto pueda entregar productos concretos de manera coherente sin retrasarse por los cambios y los requisitos cambiantes.

*Agile* tiene un alto nivel de participación del cliente e incluye revisiones frecuentes del progreso tanto con el equipo del proyecto como con el cliente. (Wrike, 2023)

El desarrollo *agile* de software se refiere a un grupo de metodologías de desarrollo de software basadas en el desarrollo iterativo, donde los requisitos y las soluciones evolucionan a través de la colaboración entre equipos multifuncionales autoorganizados. Los métodos o los procesos ágiles generalmente promueven un proceso disciplinado de gestión de proyectos que fomenta la inspección y la adaptación frecuentes, una filosofía de liderazgo que fomenta

el trabajo en equipo, la autoorganización y la responsabilidad, un conjunto de mejores prácticas de ingeniería destinadas a permitir la entrega rápida de software de alta calidad y un enfoque empresarial que alinea el desarrollo con las necesidades del cliente y los objetivos de la empresa. (Nuñez, 2016) El desarrollo agile se refiere a cualquier proceso de desarrollo que esté alineado con los conceptos del Manifiesto Agile.

1. La máxima prioridad es satisfacer al cliente a través de la entrega anticipada y continua de software valioso.
2. Aprecio de los requisitos cambiantes, incluso al final del desarrollo. Los procesos Agile aprovechan el cambio para la ventaja competitiva del cliente.
3. Entrega de software funcional con frecuencia, desde un par de semanas hasta un par de meses, con preferencia por el plazo más corto.
4. Los empresarios y desarrolladores deben trabajar juntos todos los días durante todo el proyecto.
5. Desarrollo de proyectos en torno a personas motivadas. Provisión del entorno y del apoyo necesario, así como un depósito de confianza en el equipo para que haga el trabajo.
6. El método más eficiente y efectivo para transmitir información a un equipo de desarrollo y dentro de él es la conversación cara a cara.
7. El software funcional es la medida principal del progreso.
8. Los procesos Agile promueven el desarrollo sostenible. Los patrocinadores, desarrolladores y usuarios deben poder mantener un ritmo constante indefinidamente.
9. La atención continua a la excelencia técnica y al buen diseño mejora la agilidad.
10. La simplicidad, el arte de maximizar la cantidad de trabajo no realizado, es esencial.
11. Las mejores arquitecturas, requisitos y diseños surgen de equipos autoorganizados.
12. A intervalos regulares, el equipo reflexiona sobre cómo ser más efectivo, luego sintoniza y ajusta su comportamiento en consecuencia. (Wrike, 2023)

En una metodología ágil, suele haber varios roles clave implicados en el proyecto. Los nombres y responsabilidades específicos pueden variar en función del marco ágil que se utilice (por ejemplo, Scrum, Kanban), pero es habitual encontrar los siguientes roles:

- ▶ **Product owner:** El propietario del producto representa los intereses de las partes interesadas y es responsable de gestionar el backlog del producto. Define y prioriza las características, los requisitos y la visión general del proyecto. El propietario del producto colabora con el equipo de desarrollo, proporciona aclaraciones y acepta o rechaza el trabajo terminado.

En el caso del proyecto, el *product owner* será el papel del senior mánager del departamento, al ser la persona en contacto constante con el cliente y la que nos aterriza los distintos requerimientos del proyecto a elementos desarrollables.

- ▶ **Scrum Master:** En Scrum, el scrum master actúa como facilitador y entrenador del equipo de desarrollo. Se asegura de que se sigan los principios y prácticas ágiles, elimina cualquier obstáculo que impida el progreso del equipo y ayuda a fomentar un entorno productivo y autoorganizado. El scrum master también ayuda a llevar a cabo reuniones eficaces, como la planificación del sprint, los stand-ups diarios y las retrospectivas.

En este caso, el *scrum master* del proyecto será uno de los *seniors* del departamento que, en este caso concreto coincide con el tutor del proyecto en el lado de EY. Será el encargado de guiar todo el proyecto y el que controle los avances del mismo.

- ▶ **Equipo de desarrollo:** El equipo de desarrollo está formado por las personas responsables de entregar los incrementos del proyecto. Suelen incluir desarrolladores, probadores, diseñadores y cualquier otra función necesaria. El equipo trabaja en colaboración en las tareas identificadas en el backlog del sprint y se esfuerza por entregar un software funcional y de alta calidad al final de cada sprint. Se autoorganizan y son interfuncionales, asumiendo la propiedad colectiva del trabajo y tomando decisiones sobre cómo llevar a cabo las tareas.

Para el caso de este proyecto, el equipo de desarrollo estará comprendido únicamente por mi persona, al ser un requerimiento tanto mío como de ICAI.

- ▶ **Stakeholders:** Las partes interesadas son individuos o grupos que tienen un interés en el proyecto y sus resultados. Pueden ser usuarios finales, clientes, directivos o cualquier otra parte afectada por el proyecto. Las partes interesadas aportan información, comentarios y requisitos al propietario del producto, participan en las revisiones de los *sprints* e influyen en la dirección del proyecto.

En este caso, los *stakeholders* serán por un lado el cliente final al que se le entrega la herramienta, y por extensión EY al ser beneficiario de los honorarios del proyecto, así como la reputación adquirida durante la realización del mismo. Por otro lado, podríamos encasillar en este rol a ICAI, al ser quién evaluará el desarrollo del proyecto.

Otros roles que pueden estar presentes en un proyecto ágil son:

- ▶ **Entrenador ágil:** Un entrenador ágil es un profesional con experiencia que proporciona orientación y entrenamiento al equipo y a la organización para adoptar e implementar prácticas ágiles de manera efectiva. Ayudan a fomentar una mentalidad ágil, proporcionan capacitación, facilitan talleres y abordan cualquier desafío o bloqueo.
- ▶ **Diseñador de UX:** Un diseñador de experiencia de usuario (UX) se centra en crear interfaces e interacciones intuitivas y fáciles de usar para el producto. Colaboran con el equipo de desarrollo y el propietario del producto para comprender las necesidades del usuario, realizar investigaciones de usuarios, crear *wireframes* y diseñar la experiencia general del usuario.
- ▶ **Especialista en control de calidad:** El especialista en control de calidad es responsable de garantizar la calidad de las entregas a lo largo del proyecto. Colabora con el equipo de desarrollo para definir estrategias de prueba, ejecutar casos de prueba, informar y hacer un seguimiento de los defectos y validar que el producto cumple los requisitos especificados.

## 4.4 PLANIFICACIÓN Y ESTIMACIÓN ECONÓMICA

El proyecto se va a dividir en seis partes bien diferenciadas a lo largo de sus seis meses de duración. Se dará comienzo el 9 de enero a la vuelta de vacaciones y finalizará unos pocos días antes de la defensa del trabajo de fin de máster, se estima alrededor del 9 de julio. Aunque estas sean las fechas oficiales, marcadas en combinación con los requerimientos del cliente para EY junto con los requerimientos de ICAI de cara al desarrollo del proyecto, en realidad el proyecto lleva latente desde enero de 2022 de manera intermitente.

A lo largo de todo el proyecto (tal y como se puede apreciar en el diagrama de Gantt incluido en el Anexo) se han establecido un conjunto de reuniones y formaciones que harán que el proyecto llegue a buen puerto. Para empezar, se han organizado reuniones mensuales con el *Product Owner* para que tenga un control sobre el avance de los desarrollos del proyecto, como un punto de situación. Si bien es cierto, que durante el inicio del proyecto estas reuniones serán más recurrentes debido a la necesidad de estar en mayor contacto tanto con él como con el cliente a fin de conocer de manera detallada los requerimientos del proyecto.

Por otro lado, se establecen reuniones con el equipo de EY en puntos del proyecto de mayor complejidad que requieren de un apoyo, ya sea por el lado técnico para el desarrollo de la herramienta, o del lado fiscal para un mayor entendimiento de los objetivos y terminologías usadas en el ámbito aduanero y de comercio internacional.

Además, se establecerán reuniones semanales o *weeklies* con el tutor del proyecto en EY. Estas reuniones estaban pensadas en un primer momento para ser diarias en consonancia con la metodología *agile*, pero la logística del proyecto en combinación con el resto de los proyectos del departamento no lo permitía. Así, semanalmente se producirá un punto de control con el tutor del proyecto para enseñar avances, bloqueos y recibir un *feedback* cercano y directo, pero sobre todo rápido, que permite un redireccionamiento a tiempo.

Existirán también reuniones con ICAI en varias ocasiones. El objetivo de estas reuniones será esclarecer los objetivos que se pretenden con el proyecto y adecuarlo a las exigencias que la escuela pide para un trabajo de fin de máster. Al mismo tiempo, las reuniones se

usarán como seguimiento parcial del proyecto, así como para realizar la presentación final del mismo en la defensa del trabajo.

Por último, se ha establecido una semana del proyecto para asistir a una formación de *PowerBI*, proporcionada por EY. Se trata de una semana en la que la formación ocupará la mayoría del tiempo y servirá para aprender las técnicas necesarias para realizar un reporte dinámico en *PowerBI* y realizar una excelente visualización de la información extraída del proyecto.

A continuación, se exponen las tareas a realizar para cada una de las fases en el tiempo, aunque para un vistazo rápido se va a incluir también un diagrama de Gantt en el Anexo de esta memoria.

En primer lugar, la fase de *kick-off*. Se trata de una fase de toma de contacto con cliente y estudio del propio proyecto a desarrollar. Una primera parte de toma de requisitos en la que el *product owner* mantendrá comunicaciones con el cliente para determinar exactamente cuáles son sus necesidades y requerimientos de cara a la realización del proyecto. Una vez se conocen y se aterrizan estas necesidades, se da paso a un estudio tanto de viabilidad como de aprendizaje tanto desde un punto de vista técnico de desarrollo de la herramienta, como un punto de vista fiscal. Al fin y al cabo, la especialidad del equipo de desarrollo es el desarrollo software y por tanto necesita de una gran inmersión en los conceptos e ideas en los que se va a ubicar el proyecto dentro del ámbito de la fiscalidad y la tributación en el comercio internacional. Cuando ya se tiene sobre la mesa todas las herramientas y conceptos, así como las necesidades y requerimientos del propio cliente, se realiza una propuesta de proyecto tanto a nivel técnico como fiscal/legal que se le traslada al cliente.

Una vez cliente, *product owner* e ICAI han validado la propuesta de proyecto da comienzo como tal el desarrollo del mismo. Comienza la segunda fase del trabajo, la extracción de datos. Esta fase del desarrollo es una de las más importantes, sino la más importante, ya que es el punto en el que se adquiere la materia prima con la que se va a desarrollar todo el resto del proyecto. El resto de las fases del trabajo van a depender de que se haga correctamente. La extracción de datos se hará extrayendo la información de los DUA presentados por el

cliente en la AEAT desde el inicio de 2020 hasta inicios de 2023. Se extraerán los datos tanto de las importaciones como de las exportaciones y el proceso de extracción se dividirá en dos partes. En primer lugar, se extraerán los datos planos a priori más simples y alcanzables, y en segundo lugar se extraerá la información contenida en tablas complejas.

Pasado el periodo de extracción de la información, se da paso el tratamiento y análisis de los datos. Esta fase comenzará con un estudio fiscal en el que se organizará una reunión con el equipo de EY para analizar la información que se ha extraído y sacar conclusiones sobre la dirección a seguir en el desarrollo del proyecto. Tras esto se hará un análisis descriptivo y exploratorio de los datos que servirá para quitar la información que no conviene mantener y limpiar los datos en general. Se realizarán también un tratamiento de los datos practicando las transformaciones pertinentes y necesarias para que los datos queden perfectamente expuestos de cara a su introducción en un futuro modelo de *machine learning*. Por último, se realizará un análisis de los KPI derivado del estudio fiscal inicial para poder encarar correctamente en el futuro la visualización de los datos.

La cuarta fase del proyecto constará del desarrollo de un modelo de inteligencia artificial. Se hará un estudio técnico al inicio de esta fase para evaluar las distintas opciones a realizar de acuerdo al tipo de modelo que se usará. Una vez decidido el tipo de modelo se da paso a la definición e implementación del mismo. Por último, se hará un análisis de los resultados devueltos por el modelo y se validarán los datos. Es probable que la definición del modelo y el análisis de los resultados se haga de manera conjunto debido a la necesidad de pivotar de un modelo a otro debido a que los resultados no concuerden con los esperados.

En quinto lugar, estará la fase de *reporting*. Esta fase será la más importante para el cliente ya que es la que le aporta realmente el valor que se desprende del proyecto. La correcta realización de esta fase le proporciona al cliente las herramientas y la información que necesita para tomar las decisiones pertinentes que le permitan marcar la estrategia futura de su compañía de cara a la realización de transacciones internacionales. Esta fase dará comienzo con un estudio fiscal en el que se organizará una reunión con el equipo de EY para



concretar más en detalle los KPI y puntos sobre los que se quiere hacer hincapié para mostrar en el *dashboard* a implementar.

Se realizará una formación de *PowerBI*, impartida por EY, para una excelente realización del reporte de información del proyecto. Así, tras la formación se dará paso al desarrollo del informe donde se expondrán los KPI principales del proyecto.

Se incluirá en esta fase de *reporting*, aunque no se trate de temas de visualización de datos propiamente dichos, una parte de análisis de resultados del proyecto en su conjunto, así como un análisis panorámico a futuro de cara a mejoras del proyecto o posibles nuevas aplicaciones y herramientas surgidas a partir de la realización de este.

<b>Función</b>	<b>Horas de Trabajo</b>	<b>Coste (€)</b>
Estudio Fiscal	144	1.296
Estudio Técnico	112	1.008
Extracción de información	280	2.520
Exploración y tratamiento de la información	152	1.368
Desarrollo de modelos de inteligencia artificial	128	1.152
Formación de <i>PowerBI</i>	40	500
Desarrollo de reporte de visualización de datos	296	2.664
Documentación	120	1.080
<b>TOTAL</b>	<b>1272 horas</b>	<b>11.588 €</b>

*Tabla 1 - Estimación de costes del proyecto*

Finalmente, la fase de documentación del proyecto. Esta fase es transversal a todo el proyecto, y se irá completando a lo largo de los seis meses de duración. Esta parte contiene toda la documentación requerida por ICAI para el correcto desarrollo del proyecto. Así,

incluye el Anexo A, una presentación de seguimiento, el Anexo B, una memoria descriptiva y la presentación final del proyecto.

## **4.5 DESCRIPCIÓN DE LAS TECNOLOGÍAS**

A lo largo de todo el proyecto se irán utilizando distintas tecnologías que harán posible la consecución de los objetivos propuestos. A continuación, se exponen aquellas tecnologías utilizadas en las fases de extracción, modelado y visualización de datos, al ser las partes más tecnológicas de todo el proyecto que requieren de una explicación más en detalle.

### **4.5.1 WEB SCRAPPING Y SELENIUM**

En primer lugar, para la fase de extracción de datos la tecnología utilizada ha sido la de *web scrapping* con Python. En este proyecto se hará uso de *Selenium* con Python para poder llevar a cabo la extracción de la información. Los datos a extraer se encuentran alojados en la página web de la Agencia Tributaria de forma visual, sin haber opción de acceder a alguna base de datos o API que nos pueda devolver la información requerida de una forma más sencilla. Es por esto, por lo que se recurrirá al *scrapeo* web de la página de la AEAT tanto para el apartado de importaciones como para el apartado de exportaciones.

El *web scraping*, también conocido como *web harvesting* o extracción de datos web, es el proceso de extracción automática de datos de sitios web. Consiste en utilizar un programa o script para acceder y recuperar información específica de páginas web, normalmente de forma automatizada y repetitiva.

El *web scraping* permite recopilar datos de múltiples fuentes de Internet sin tener que visitar manualmente cada sitio web y copiar y pegar la información. Al automatizar el proceso de extracción de datos, se puede ahorrar tiempo y esfuerzo en la recopilación y organización de datos para diversos fines, como la investigación, el análisis o la creación de aplicaciones. (Hunter, 2021)

Por su parte, *Selenium* es un marco de trabajo de código abierto y un conjunto de herramientas que permite realizar pruebas automatizadas de navegadores y *scrapeo* web. (Selenium, 2020) Proporciona una forma independiente de la plataforma para automatizar las interacciones con los navegadores web y realizar diversas tareas, como hacer clic en botones, rellenar formularios, navegar por páginas y extraer datos. (Abdullaiev & Ibrahimova, 2020)

Desarrollado originalmente como una herramienta de pruebas para aplicaciones web, *Selenium* ha ganado popularidad como una potente herramienta de *scrapeo* web debido a su capacidad para manejar sitios web dinámicos que dependen en gran medida de JavaScript para la interactividad. Es compatible con varios lenguajes de programación, como Python, Java, C#, Ruby y JavaScript, por lo que es ampliamente accesible para desarrolladores de distintas plataformas.

El componente central de *Selenium* es el *WebDriver*, que actúa como puente de comunicación entre el código y el navegador web. El *WebDriver* permite controlar el navegador mediante programación, emular las interacciones del usuario y recuperar datos de las páginas web.

Algunas de las principales características y ventajas de *Selenium* son:

- ▶ Compatibilidad entre navegadores: *Selenium* soporta múltiples navegadores, lo que le permite escribir scripts de prueba o realizar *web scraping* a través de diferentes navegadores sin cambios significativos en el código.
- ▶ Manejo de contenido dinámico: *Selenium* destaca en la interacción con páginas web dinámicas que dependen de JavaScript para la renderización del contenido. Se puede esperar a que los elementos se carguen, manejar peticiones AJAX y actualizar dinámicamente el contenido de la página.
- ▶ Localización e interacción de elementos: *Selenium* proporciona una amplia gama de métodos para localizar e interactuar con elementos en páginas web, incluyendo la búsqueda de elementos por ID, nombre de clase, selectores CSS, XPath, y más. Se

pueden realizar acciones como hacer clic en botones, introducir texto, enviar formularios o extraer datos de elementos.

- ▶ Soporte para múltiples lenguajes de programación
- ▶ Comunidad y ecosistema: *Selenium* tiene una comunidad grande y activa de desarrolladores que contribuyen a su desarrollo, proporcionan soporte y crean recursos útiles. También se integra bien con otras bibliotecas de pruebas y *scraping*, por lo que es una herramienta versátil para tareas relacionadas con la web.

## 4.5.2 MACHINE LEARNING

En segundo lugar, para la fase de modelización de los datos se hará uso de las distintas librerías de Python que permiten la implementación de modelos de *machine learning*. Se hará uso de modelos de regresión capaces de predecir el valor a pagar en una aduana en función de unos parámetros predefinidos.

### 4.5.2.1 Modelos de regresión y aprendizaje supervisado

En el modelado estadístico, el análisis de regresión es un conjunto de procesos estadísticos para estimar las relaciones entre una variable dependiente (a menudo denominada variable de "resultado" o "respuesta", o "etiqueta" en el lenguaje del aprendizaje automático) y una o más variables independientes (a menudo denominadas "predictores", "covariables", "variables explicativas" o "características"). La forma más común de análisis de regresión es la regresión lineal, en la que se encuentra la línea (o una combinación lineal más compleja) que mejor se ajusta a los datos según un criterio matemático específico.

El análisis de regresión se utiliza principalmente para dos fines conceptualmente distintos. En primer lugar, el análisis de regresión se utiliza ampliamente para la predicción y la previsión, donde su uso se solapa sustancialmente con el campo del aprendizaje automático. En segundo lugar, en algunas situaciones el análisis de regresión puede utilizarse para inferir relaciones causales entre las variables independientes y dependientes. (Freedman, 2009) (Cook & Weisberg, 1982)

Se optará por modelos de regresión con aprendizaje supervisado ya que los datos obtenidos nos permiten contrastar la información predicha por el modelo con los datos reales.

El aprendizaje supervisado es un paradigma de aprendizaje automático para problemas en los que los datos disponibles consisten en ejemplos etiquetados, lo que significa que cada punto de datos contiene características (covariables) y una etiqueta asociada. El objetivo de los algoritmos de aprendizaje supervisado es aprender una función que asigne vectores de características (entradas) a etiquetas (salida), basándose en pares de ejemplos de entrada-salida (Russell & Norvig, 2010). Infiere una función a partir de datos de entrenamiento etiquetados que consisten en un conjunto de ejemplos de entrenamiento. (Mohri, Rostamizadeh, & Talwalkar, 2012) En el aprendizaje supervisado, cada ejemplo es un par formado por un objeto de entrada (normalmente un vector) y un valor de salida deseado (también llamado señal de supervisión). Un algoritmo de aprendizaje supervisado analiza los datos de entrenamiento y produce una función inferida, que puede utilizarse para mapear nuevos ejemplos. Un escenario óptimo permitirá al algoritmo determinar correctamente las etiquetas de clase para instancias no vistas. Esto requiere que el algoritmo de aprendizaje generalice a partir de los datos de entrenamiento a situaciones no vistas de una manera "razonable". Esta calidad estadística de un algoritmo se mide a través del denominado error de generalización. (Wikipedia, 2023)

De entre los distintos tipos de modelos de regresión de aprendizaje supervisado se exponen a continuación algunos ejemplos que podrán optar a ser el modelo elegido para la implementación de este proyecto.

#### ***4.5.2.2 Regresión Lineal***

La **regresión lineal** es un enfoque lineal para modelizar la relación entre una respuesta escalar y una o más variables explicativas (también conocidas como variables dependientes e independientes). El caso de una variable explicativa se denomina regresión lineal simple; para más de una, el proceso se denomina regresión lineal múltiple (Freedman, 2009)

En la regresión lineal, las relaciones se modelan utilizando funciones predictoras lineales cuyos parámetros desconocidos del modelo se estiman a partir de los datos. Estos modelos se denominan modelos lineales. (Seal, 1967) Lo más habitual es que se suponga que la media condicional de la respuesta dados los valores de las variables explicativas (o predictoras) es una función afín de dichos valores; con menos frecuencia, se utiliza la mediana condicional o algún otro cuantil. Como todas las formas de análisis de regresión, la regresión lineal se centra en la distribución de probabilidad condicional de la respuesta dados los valores de los predictores, en lugar de en la distribución de probabilidad conjunta de todas estas variables, que es el dominio del análisis multivariante. (Wikipedia, 2023)

#### ***4.5.2.3 Regresión Polinómica***

La **regresión polinómica** es una forma de análisis de regresión en la que la relación entre la variable independiente  $x$  y la variable dependiente  $y$  se modela como un polinomio de  $n$ -ésimo grado en  $x$ . La regresión polinómica ajusta una relación no lineal entre el valor de  $x$  y la media condicional correspondiente de  $y$ . Aunque la regresión polinómica ajusta un modelo no lineal a los datos, como problema de estimación estadística es lineal, en el sentido de que la función de regresión es lineal en los parámetros desconocidos que se estiman a partir de los datos. Por este motivo, la regresión polinómica se considera un caso especial de la regresión lineal múltiple.

Las variables explicativas (independientes) resultantes de la expansión polinómica de las variables "base" se conocen como términos de grado superior. Estas variables también se utilizan en entornos de clasificación. (Chang, Hsieh, Chang, Ringgaard, & Lin, 2010) (Wikipedia, 2022)

#### ***4.5.2.4 Árboles de decisión***

El aprendizaje de **árboles de decisión** es un enfoque de aprendizaje supervisado utilizado en estadística, minería de datos y aprendizaje automático. En este formalismo, se utiliza un árbol de decisión de clasificación o regresión como modelo predictivo para extraer conclusiones sobre un conjunto de observaciones.

Los modelos de árbol en los que la variable objetivo puede tomar un conjunto discreto de valores se denominan árboles de clasificación; en estas estructuras de árbol, las hojas representan etiquetas de clase y las ramas representan conjunciones de características que conducen a esas etiquetas de clase. Los árboles de decisión en los que la variable objetivo puede tomar valores continuos (normalmente números reales) se denominan árboles de regresión. En términos más generales, el concepto de árbol de regresión puede extenderse a cualquier tipo de objeto dotado de disimilitudes entre pares, como las secuencias categóricas. (Studer, Ritschard, Gabadinho, & Müller, 2011)

Los árboles de decisión se encuentran entre los algoritmos de aprendizaje automático más populares dada su inteligibilidad y simplicidad. (Wu, y otros, 2008)

En el análisis de decisiones, un árbol de decisión se puede utilizar para representar visual y explícitamente las decisiones y la toma de decisiones. En minería de datos, un árbol de decisión describe datos (pero el árbol de clasificación resultante puede ser una entrada para la toma de decisiones). (Wikipedia, 2023)

#### ***4.5.2.5 Bagging***

En el aprendizaje automático, el *bagging* (abreviatura de *bootstrap aggregating*) es una técnica utilizada para mejorar la precisión y solidez de los modelos predictivos. El *bagging* consiste en crear múltiples subconjuntos del conjunto de datos de entrenamiento original mediante un proceso denominado muestreo *bootstrap*, en el que las muestras se extraen con reemplazo. Cada subconjunto se utiliza para entrenar un modelo distinto, a menudo del mismo tipo, y la predicción final se realiza sumando las predicciones de todos los modelos individuales. (Fushiki, 2010)

Su funcionamiento consiste en tres fases. En primer lugar, se extraen muestras aleatorias del conjunto de datos de entrenamiento original con reemplazo donde cada subconjunto tiene el mismo tamaño que el conjunto de datos original, pero algunas muestras pueden aparecer varias veces, mientras que otras pueden omitirse.

En segundo lugar, se entrena un modelo independiente en cada subconjunto utilizando el algoritmo de entrenamiento elegido. Por ejemplo, si utiliza árboles de decisión, cada subconjunto se utilizaría para entrenar un modelo de árbol de decisión.

Una vez entrenados todos los modelos, se realizan predicciones sobre datos no vistos utilizando cada modelo individual. La predicción final se obtiene agregando las predicciones de todos los modelos, aunque este método de agregación varía en función del tipo de problema; para las tareas de clasificación, el enfoque más común es la votación por mayoría, mientras que, para las tareas de regresión, las predicciones se suelen promediar.

El principal objetivo del *bagging* es reducir la varianza de los modelos y mejorar su capacidad de generalización. Al generar diferentes subconjuntos de datos, el ensamblaje introduce diversidad entre los modelos, lo que ayuda a reducir el sobreajuste y hace que el conjunto sea más robusto. (Wikipedia, 2023)

#### **4.5.2.6 Boosting**

En el aprendizaje automático, el *boosting* es una técnica utilizada para combinar modelos débiles o básicos en un modelo predictivo fuerte. A diferencia del *bagging*, en el que los modelos se construyen de forma independiente, el *boosting* construye los modelos secuencialmente de forma que cada modelo posterior intenta corregir los errores cometidos por los modelos anteriores. La idea es crear un conjunto sólido centrándose en los casos mal clasificados y dándoles más peso durante el proceso de entrenamiento.

En primer lugar, a cada instancia del conjunto de datos de entrenamiento se le asigna un peso inicial. Inicialmente, todos los pesos son iguales. Después un modelo débil, a menudo denominado aprendiz débil, se entrena en el conjunto de datos de entrenamiento utilizando los pesos actuales de las instancias. El aprendiz débil puede ser cualquier algoritmo capaz de realizar predicciones, como árboles de decisión (p. ej., *AdaBoost*), modelos lineales (p. ej., *Gradient Boosting*) o redes neuronales (p. ej., *XGBoost*, *LightGBM*). En este caso se hará uso de *XGBoost*.



A continuación, se evalúa el rendimiento del modelo débil y se identifican las instancias con errores más elevados. Así, estas instancias reciben pesos más altos, lo que significa que se vuelven más influyentes durante el entrenamiento del siguiente modelo débil. De este modo, los modelos posteriores se centran en las instancias previamente difíciles.

La predicción del modelo débil se combina con las predicciones realizadas por los modelos entrenados previamente mediante una suma ponderada u otra regla de combinación. Las ponderaciones suelen basarse en el rendimiento del modelo débil.

Este proceso se repite de forma iterativa, y cada iteración tiene como objetivo mejorar el rendimiento global del modelo centrándose en las instancias mal clasificadas. De esta manera, la predicción final se obtiene agregando las predicciones de todos los modelos débiles, normalmente mediante un voto mayoritario ponderado o un promedio ponderado.

Los algoritmos de *boosting*, como *AdaBoost (Adaptive Boosting)*, *Gradient Boosting*, *XGBoost* y *LightGBM*, se han aplicado con éxito en diversos ámbitos y han alcanzado un rendimiento puntero en muchas tareas de aprendizaje automático. Son especialmente eficaces en el manejo de patrones complejos en los datos y son conocidos por su capacidad para reducir el sesgo y mejorar la precisión general de los modelos.

Los modelos de *boosting* tienden a ser más intensivos computacionalmente en comparación con los de *bagging* debido a la naturaleza secuencial del entrenamiento, pero pueden producir predicciones potentes y muy precisas cuando se aplican correctamente.

#### ***4.5.2.7 Random Forest***

Los bosques aleatorios o bosques de decisión aleatorios son un método de aprendizaje conjunto para clasificación, regresión y otras tareas que funciona construyendo una multitud de árboles de decisión en el momento del entrenamiento. Para tareas de clasificación, la salida del bosque aleatorio es la clase seleccionada por la mayoría de los árboles. Para tareas de regresión, se devuelve la predicción media o promedio de los árboles individuales. (Ho T. K., 1995) (Ho T. K., The Random Subspace Method for Constructing Decision Forests, 1998) Los bosques de decisión aleatorios corrigen el hábito de los árboles de decisión de

sobre ajustarse a su conjunto de entrenamiento. (Hastie, Tibshirani, & Friedman, 2008) En general, los bosques aleatorios superan a los árboles de decisión, pero su precisión es menor que la de los árboles potenciados por gradiente. Sin embargo, las características de los datos pueden afectar a su rendimiento. (Madeh & E., 2020) (Piryonesi & El-Diraby, 2021) (Wikipedia, 2023)

De forma similar al *bagging*, *Random Forest* comienza creando múltiples subconjuntos del conjunto de datos de entrenamiento original a través del muestreo *bootstrap*. Cada subconjunto, también conocido como "*bag*" o "muestra *bootstrap*", se crea seleccionando aleatoriamente muestras del conjunto de datos original con reemplazo.

Además de muestrear los datos, *Random Forest* introduce un nivel adicional de aleatoriedad seleccionando al azar un subconjunto de características (columnas) en cada división de un árbol de decisión. Esto ayuda a aumentar la diversidad entre los árboles individuales y evita el sobreajuste a características específicas.

Así, para cada muestra *bootstrap*, se entrena un árbol de decisión utilizando las características seleccionadas. Los árboles de decisión suelen crecer hasta su máxima profundidad sin poda, lo que puede dar lugar a árboles individuales que se ajusten en exceso a los datos de entrenamiento.

Una vez entrenados todos los árboles de decisión, las predicciones se realizan agregando las predicciones de todos los árboles individuales. Para las tareas de clasificación, el método de agregación más común es la votación por mayoría, en la que la clase con el mayor número de votos de los árboles se selecciona como predicción final. En las tareas de regresión, las predicciones suelen promediarse entre todos los árboles.

Las principales ventajas de los modelos *Random Forest* son su capacidad para manejar datos de alta dimensión, capturar interacciones complejas entre características y proporcionar estimaciones de la importancia de las características. También son resistentes al sobreajuste y funcionan bien en presencia de datos ruidosos o ausentes.

### **4.5.3 POWERBI**

Por último, para la fase de visualización de los datos se ha utilizado la herramienta de Microsoft *PowerBI*.

Microsoft *PowerBI* es un producto de software de visualización de datos interactivos desarrollado por Microsoft con un enfoque principal en la inteligencia empresarial (Microsoft, s.f.) que forma parte de Microsoft Power Platform. *PowerBI* es una colección de servicios de software, aplicaciones y conectores que trabajan juntos para convertir fuentes de datos no relacionadas en perspectivas coherentes, visualmente inmersivas e interactivas. Los datos pueden introducirse leyendo directamente desde una base de datos, una página web o archivos estructurados como hojas de cálculo, CSV, XML y JSON.

*PowerBI* proporciona servicios de BI (inteligencia empresarial) basados en la nube, conocidos como "*PowerBI Services*", junto con una interfaz basada en escritorio, denominada "*PowerBI Desktop*". Ofrece capacidades de almacén de datos, incluyendo preparación de datos, descubrimiento de datos y cuadros de mando interactivos. (Gartner, Inc., s.f.) (Wikipedia, 2023)

## Capítulo 5. SISTEMA DESARROLLADO

El proyecto consta de 3 partes bien diferenciadas que en conjunto conforman la solución propuesta al cliente y al cual le aporta el valor necesario para una toma de decisiones guiada y supervisada por el propio proyecto que le llevará a posicionarse estratégicamente en el sector del comercio internacional.

El sistema está dividido en varias fases. Una fase de extracción de los datos y la información en primer lugar, seguida de una limpieza y transformación de los datos para adecuarlos al modelo. En segundo lugar, una fase de modelización de los datos donde se elegirán distintos modelos de *machine learning* que servirán para predecir el valor a pagar en la aduana por una mercancía determinada. Por último, una fase de visualización de los datos desde donde el cliente podrá visualizar de manera rápida todas sus transacciones según distintos criterios y podrá realizar una predicción en tiempo real facilitando información básica de la mercancía a importar.

### 5.1 EXTRACCIÓN DE LA INFORMACIÓN

Inicialmente se optó por extraer el mayor número de datos posibles para así después poder tener la capacidad de decidir cuales usar y cuáles no. Así se comenzó extrayendo la información histórica del cliente relativa a todas las transacciones de importaciones y de exportaciones desde el 1 de enero del 2020 hasta mayo del 2023.

Para realizar la extracción de estos datos se utilizaron las librerías de *scrappeo web* de *Selenium* para *Python*. La extracción comenzaba con un archivo Excel desde el cual el proceso de Python leía las fechas de búsqueda en las cuales hubiese habido un DUA de importación o exportación. Esto se hizo así ya que la página web de la Agencia Tributaria solo permitía hacer búsquedas con un rango máximo de 3 meses. Así, el archivo Excel contenía todas las fechas de inicio y fin del rango de búsqueda empezando el 1 de enero de 2020 y terminando el 31 de mayo de 2023. Cada fila del Excel contenía dos columnas, la

primera coincidía con la fecha de inicio del rango y la segunda con la fecha de fin. Como el rango máximo de búsqueda era de tres meses, una de las filas del Excel podía tener, por ejemplo, la fecha del 01/01/2020 en la primera columna y la fecha del 31/03/2020 en la segunda. En la Ilustración 8 se puede ver el aspecto de la principal de consulta de importaciones de la AEAT desde la que el proceso comenzaba la búsqueda.

**CONSULTA DE IMPORTACIONES**

Excel | Excel Tributos

Búsqueda como

Importador  Declarante

Fecha Admisión desde: 04/01/2023

Fecha Admisión hasta: 05/31/2023

DUA:   3

Despachados: Si  No  Todos

Buscar Limpiar

Nº del D.U.A.	Tipo	F.Admisión	Importador	Declarante	Círculo	Actuario	F.Levante	C.S.V. Levante	Sol desp
23 ES 000855 3	DUA Declaración Completa	03-04-2023	N00	A0	AMARILLO		03-04-2023	UUJVP9PYRS	
23 ES 000855 3	DUA Declaración Completa	07-04-2023	N00	A0	AMARILLO		11-04-2023	4VUHHXJ6AK	
23 ES 000855 3	DUA Declaración Completa	22-04-2023	N00	A0	VERDE	DESAUT	22-04-2023	EE6SQHF2AC	

Ilustración 8 - Página de búsqueda de DUA de la AEAT

El proceso se encargaba de hacer un flujo iterativo para realizar todas las consultas necesarias para cumplir con el rango de búsqueda hasta la fecha final de extracción de los datos. En cada iteración el proceso cogía las fechas del archivo Excel, las colocaba en las etiquetas correspondientes de la página web y pulsaba en el botón “Buscar”. Este procesamiento de búsqueda de etiquetas, introducción de fechas y *click* en el botón se realiza a través de *Selenium*. Sin embargo, que *Selenium* pudiese identificar estos elementos y poder realizar acciones sobre ellos fue necesario un trabajo previo de identificación de los componentes web de la página que interesaba reconocer para poder realizar acciones sobre ellos desde *Selenium*. En la Ilustración 9 se puede ver un fragmento a modo de ejemplo de la estructura CSS de la página web de búsqueda de importaciones/exportaciones. Aunque aparentemente parece un proceso sencillo, realmente no lo era ya que la estructura de la página puede variar en función de la fecha que se incluya en la búsqueda o el tipo de DUA pudiendo aparecer elementos nuevos como requerimientos u otras partes estructurales que no se tuvieron en

cuenta en primera instancia. Es por esto, por lo que la fase de extracción fue un trabajo largo y laborioso basado en la técnica de prueba-error.

```
<legend class="oculto"> Condiciones de búsqueda de la consulta </legend>
<ul class="AEAT_form">
  <li class="dos_cols">
  <li class="cua_cols">
  <li class="">
  <li class="dos_cols" title="Fecha en formato dd-mm-aaaa">
  <li class="dos_cols" title="Fecha en formato dd-mm-aaaa">
    <label for="wFecOpeHdia"> Fecha Admisión hasta </label>
    <input class="text input_mask mask_date_us datepicker" type="date" id="wFecOpeHdia" name="wFecOpeHdia"
      autocomplete="off" value="31-05-2023" size="10">
  </li>
  <li>
    <label for="wAnio">
    <input type="text" name="wAnio" id="wAnio" value maxlength="2" size="2">
  </li>
  <li>
    <input type="text" name="wPais" id="wPais" value maxlength="2" size="2" aria-label="Sin literal">
  </li>
  <li>
    <input type="text" name="CABECERA_recinto" id="CABECERA_recinto" value maxlength="6" size="6" aria-label="Sin
      literal">
  </li>
  <li>
    <input type="text" name="wPrefijo" id="wPrefijo" value="3" maxlength="1" size="1" aria-label="Sin literal">
  </li>
  <li>
    <input type="text" name="wNum" id="wNum" value maxlength="6" size="6" aria-label="Sin literal">
```

Ilustración 9 - Etiquetas web del CSS de la página de búsqueda

Selenium es capaz de encontrar los componentes de una web de distintas maneras:

- ▶ Por el ID de la etiqueta: fue forma más simple de encontrar componentes al tratarse de un de un identificador único sobre toda la página.
- ▶ Por el nombre de la clase de la etiqueta: era muy útil sobre todo para todas aquellas etiquetas repetidas o que constituían una lista de *bullet points*.
- ▶ Por el CSS Selector: se usaba, sobre todo, para buscar aquellas etiquetas que contenían links y que podían confundirse con otros componentes de la página.
- ▶ Por el XPATH de la etiqueta: era la forma más parecida a buscar las etiquetas por ID ya que se iba siguiendo el camino de herencia de la etiqueta mirando a sus etiquetas padres. Fue muy útil con etiquetas que no tenían ID y se podían confundir con otras partes de la página web, sin embargo, era la forma más laboriosa.

- ▶ Por el tipo de etiqueta: se usaba de forma similar a la clase de la etiqueta para buscar etiquetas repetitivas, pero sobre todo para evitar aquellos componentes que se encontraban ocultos.

Volviendo al proceso de búsqueda de DUA, el proceso iteraba sobre todos los rangos de fecha establecidos. En cada iteración el proceso incluía las fechas y pulsaba sobre el botón buscar. En este punto, se mostraban en la parte inferior de la página, como se puede apreciar en la Ilustración 6, los DUA referentes a este intervalo de tiempo. Sin embargo, en aquellos periodos en los que el número de DUA era muy elevado, no se mostraban en la página todas los DUA correspondientes al periodo. Existía un botón “ver más” que pasaba a la siguiente página y mostraba más DUA, en cada página se mostraban 10 DUA por lo que en aquellos periodos en los que hubiese más de 10 DUA el proceso se complicaba. En cambio, existía una alternativa pudiendo descargar las referencias de todos los DUA del período elegido en un archivo Excel pulsando sobre un botón para dicho efecto. Así, el proceso en cada iteración incluía las fechas de búsqueda, pulsaba en el botón “Buscar” y posteriormente en el botón “Excel” para descargar todos los DUA. La complejidad de este apartado pasaba por que el proceso debía esperar a que terminase la descarga del archivo Excel para poder seguir con la siguiente iteración. Este paso dio muchos problemas ya que el tiempo de descarga dependía de múltiples factores y nunca era uniforme. Finalmente se consiguió solventar con una solución que buscaba unos archivos temporales que aparecen cuando comienza una descarga y desaparecen cuando finaliza.

Una vez el proceso termina la iteración sobre todas las fechas propuestas selecciona todos los archivos Excel descargados y los junta en un solo para facilitar su manejo. Aquí comienzo un nuevo proceso iterativo en el que se extrae del Excel uno a uno los DUA y se consultan en la web de la AEAT. Esta vez, en vez de ser con un motor de búsqueda, se accedía a la página del DUA específico añadiendo a una URL fija la extensión con la referencia del DUA determinado.

En la Ilustración 10 se puede ver un fragmento a modo de ejemplo de la página que muestra el detalle de información de un DUA en concreto. Como norma general, cada DUA estaba

estructurada en diferentes partes: “Sumario”, “DUA” y “Partidas”. Aunque, como se ha mencionado antes esta estructura puede variar dependiendo del DUA, de la aparición de requerimientos o del número de partidas que contenga la mercancía pertinente.

CONSULTA DEL D.U.A. DE IMPORTACIÓN

Acciones sobre el DUA [Imprimir](#)

Datos del DUA [Datos Contables](#) [Datos de Gestión](#) [Relación de Importaciones](#)

Consultas relacionadas [YUA](#) [Sistema de Liquidación IVA Diferido](#) [Mis Expedientes](#)

**Sumario**

[Sumario](#) [DatosGestión](#) [DatosContables](#) [TributosCalculados](#) [GarantiasActivas](#) [Requerimientos](#) [Correos Enviados](#) [SLID](#) [Ver Todo](#)

Requerimientos

D.U.A.: [REDACTED]

Nº de Partidas: 19

Fecha de Admisión: 03-04-2023

Importador: [REDACTED] Sujeto IVA Diferido

Declarante: [REDACTED] Operador Económico Autorizado

Partida	Código mercancía	País origen	Preferencia	Régimen	Tipos posibles de deudas	Valor Estadístico	Masa Neta	Bultos
00001	[REDACTED]	CN	1 00	4000	Deuda real y deuda real sin determinar	21.081,30 €	2.164,100 Kg.	19
00002	[REDACTED]	CN	1 00	4000	Deuda real y deuda real sin determinar	1.157,55 €	126,000 Kg.	2
00003	[REDACTED]	CN	1 00	4000	Deuda real y deuda real sin determinar	6.776,54 €	823,000 Kg.	125
00004	[REDACTED]	CN	1 00	4000	Deuda real y deuda real sin determinar	944,50 €	143,000 Kg.	13
00005	[REDACTED]	CN	1 00	4000	Deuda real y deuda real sin determinar	1.065,68 €	102,500 Kg.	7

Ilustración 10 - Consulta de información de un DUA

Así, una vez en la página de detalle de información se identificaron todos y cada uno de los componentes que contenían información relevante para el análisis del DUA, de manera que después *Selenium* fuese capaz de extraer la información contenida en cada uno de los componentes. Otra de las dificultades de este apartado fue la poca homogeneidad de los componentes ya que, mientras los datos planos ofrecían una complejidad media, las tablas y gráficos fueron un desafío mayor a la hora de extraer la información.

Toda la información extraída, los datos planos y las tablas, se guardaban en un *dataframe* para posteriormente volcarlo a un Excel maestro que contendría toda la información de todos y cada uno de los DUA inspeccionados.

Aparte de la información contenida en el DUA el proceso se encargaba de descargarse los requerimientos de aquellos DUA que los tuviesen, en formato PDF, para almacenarlos en una carpeta. Además, se volcaba en el Excel maestro si el DUA correspondiente había tenido requerimientos o no.



## 5.2 LIMPIEZA Y TRANSFORMACIÓN DE LOS DATOS

Una vez tenemos todos los datos relativos a las DUA de importación y exportación en Excel se da paso al preprocesamiento de la información, para adecuarla a la entrada de un modelo de *machine learning*.

En un primer acercamiento, la limpieza y transformación de variables se hizo con el lenguaje de programación R. Inicialmente, se hizo una limpieza de variables desechando aquellas que no aportaban ningún valor técnico ni ningún valor de negocio. Así, pasamos de 47 variables que teníamos al inicio a 27. Posteriormente se renombraron los nombres de las variables para que el procesamiento fuese más simple en R teniendo nombres más cortos y simples.

En primer lugar, se hicieron un número de transformaciones que detallaré a continuación, pero que finalmente no sirvieron para nada ya que como se detallará en el apartado 6. Análisis de los resultados solo empeoran la capacidad de predicción de nuestro modelo.

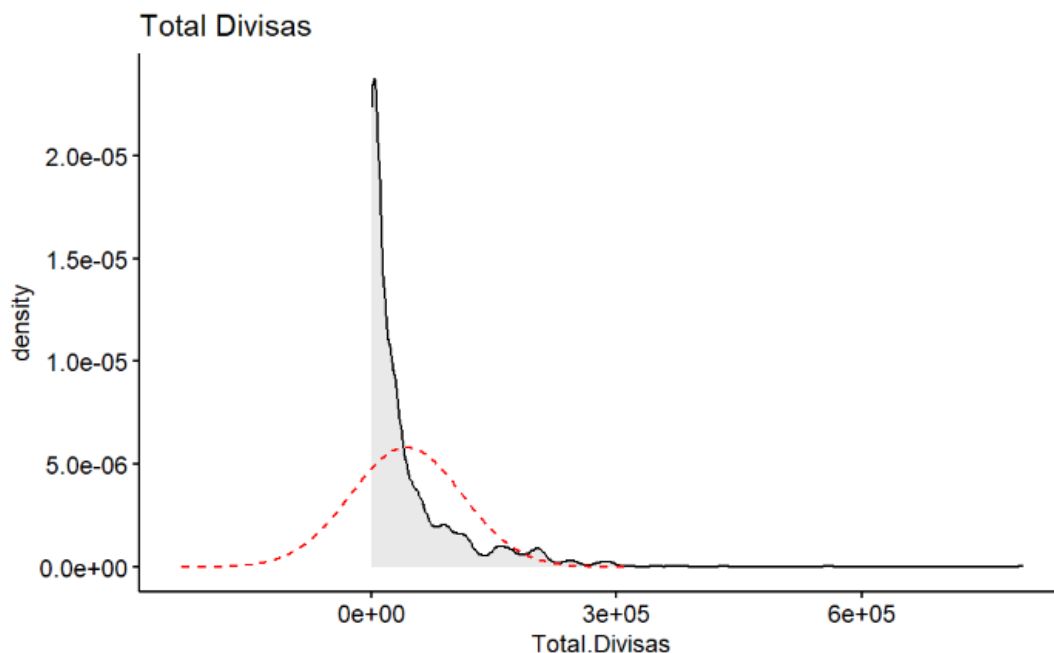
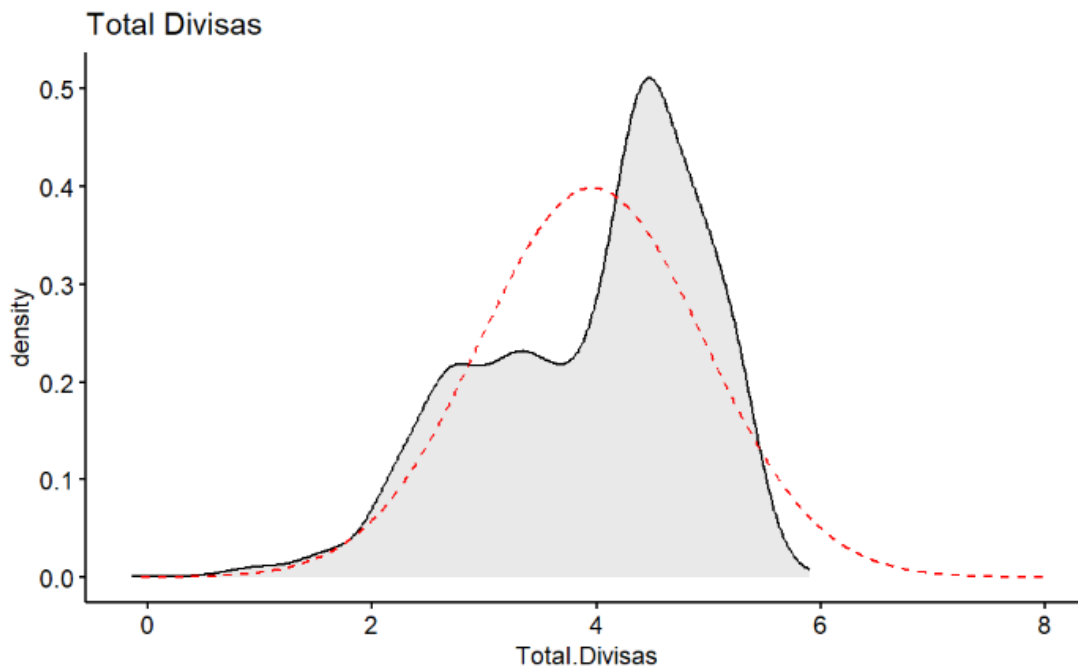


Ilustración 11 - Histograma de la variable "Total Divisas"

Así, lo que se buscaba en un inicio era que el conjunto de datos cuantitativos fuese lo más similar a una distribución normal, evitando valores atípicos y una asimetría excesiva. Con esto se eliminaron en un primer momento todos los valores nulos y aquellos valores atípicos que alejaban a la distribución de la normalidad. Se hicieron test de asimetría con resultados muy deficientes que se alejaban de la forma de una distribución normal.

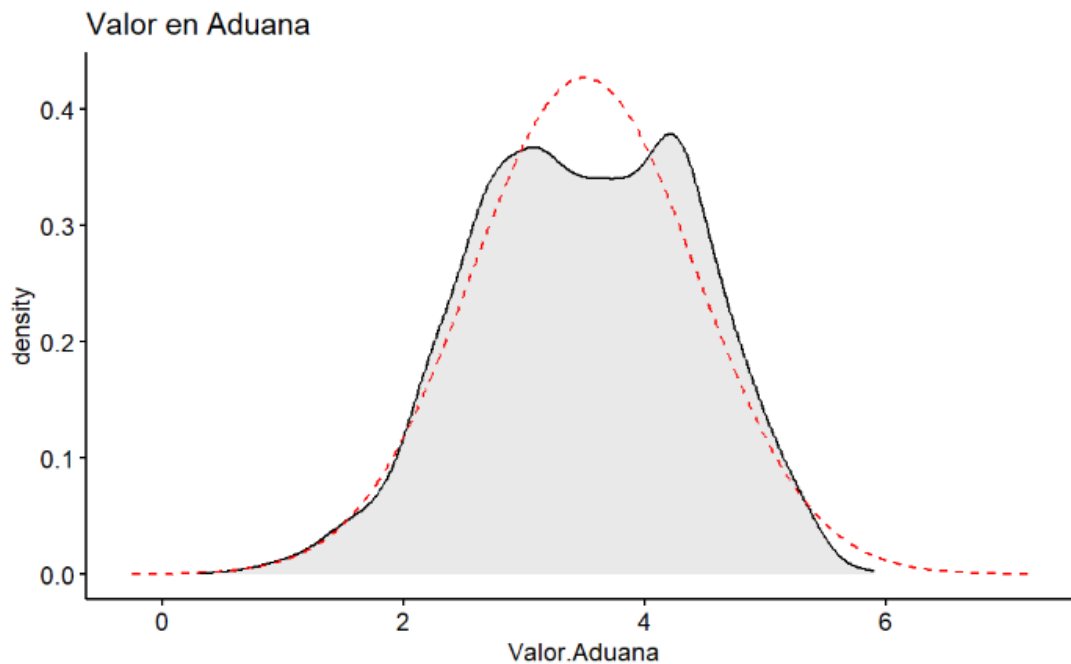
En la Ilustración 11 se puede apreciar el histograma de una de las variables a modo de ejemplo, visualizando la clara asimetría de la distribución.

Tras estos resultados se aplica una transformación logarítmica a todas aquellas variables numéricas con altos grados de asimetría y se vuelve a pasar el test. En este caso, muchas de las variables mejoran considerablemente, si bien un par de ellas siguen teniendo cierta asimetría en aparte central de la distribución tal y como podemos observar en la Ilustración 12.



*Ilustración 12 - Histograma de la variable "Total Divisas" tras una transformación logarítmica*

Así, aquellas variables con bajos niveles de asimetría quedaron definidas con una forma muy similar a la de una distribución normal (véase la Ilustración 13)

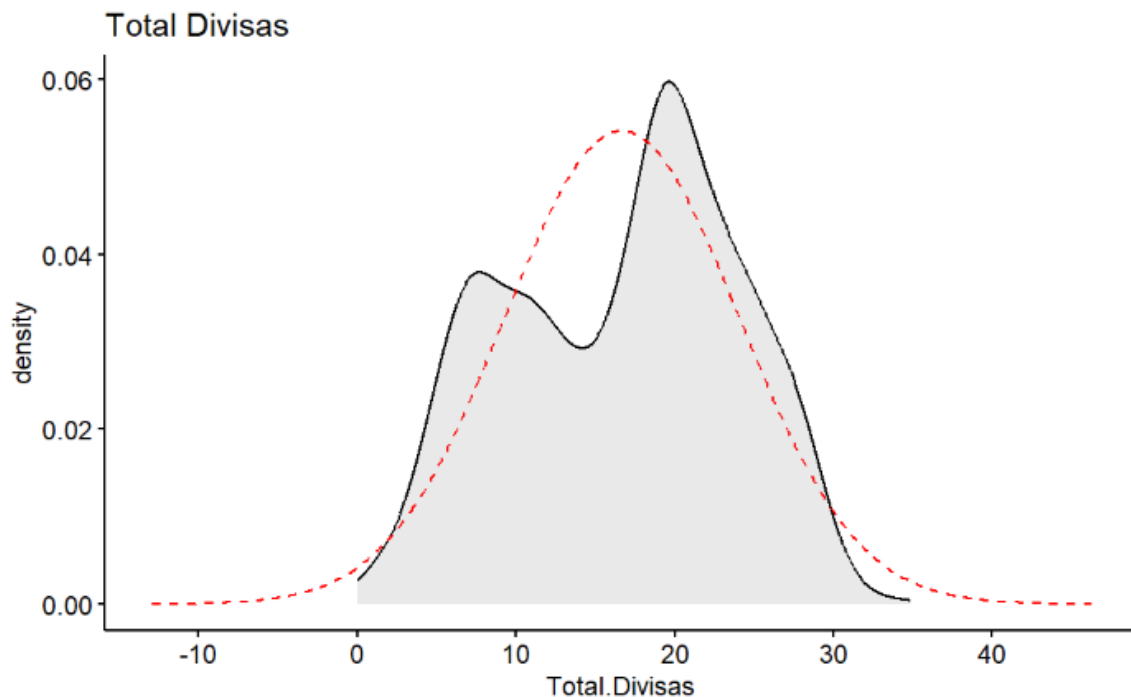


*Ilustración 13 - Histograma de la variable "Valor Aduana" tras una transformación logarítmica*

Para aquellas variables que seguían teniendo un grado de asimetría considerable se hizo una segunda transformación elevando al cuadrado dichas variables. Se vuelve a hacer la prueba de asimetría y los resultados mejoran aún más que en el paso anterior quedando los datos numéricos lo más ajustado posible a una distribución normal. Si bien, como se puede apreciar en la [Ilustración 14](#), la distribución no es perfectamente normal, para ello se hizo una nueva prueba para determinar la curtosis de estas variables transformadas y en todas de ellas los resultados fueron menores a 3 (valor estándar de curtosis para una distribución normal). Este resultado implica que las distribuciones tengan un exceso de curtosis negativa lo que se traduce en colas más finas que las de una distribución normal con menos valores extremos ya sean positivos o negativos. Esto tiene sentido ya que en la limpieza inicial de los datos se quitaron muchas variables que contenían valores atípicos.

Finalmente, se hicieron dos transformaciones más previas a la introducción de los datos en los modelos de *machine learning*. Por un lado, se normalizaron las variables numéricas después de todas sus transformaciones para que todas se movieran en el mismo rango y el modelo no diese más importancia a unas que otras. Por otro lado, para poder incluir las

variables categóricas en el modelo es necesario hacer una traducción numérica para que el modelo comprenda el sentido de los datos de la variable. Así, se hicieron dos tipos de transformación. Para aquellas variables en las que el orden de las categorías no influye y puede ser aleatorio se aplicó *one hot encoding*. Se trata de una transformación en la que desaparece la propia variable categórica y aparecen  $n-1$  variables nuevas, siendo  $n$  el número de categorías diferentes de la variable original. Estas nuevas columnas serán variables binarias que representan si el dato pertenece o no a una categoría concreta. Por otro lado, para aquellas variables en las categorías si deben tener algún tipo de orden o ranking, se aplicó *label encoding*. Esta transformación consiste en darle un valor numérico a cada categoría de la variable original en el orden que se estime oportuno.



*Ilustración 14 - Histograma de la variable "Total Divisas" tras una transformación cuadrada*

### **5.3 MODELOS DE MACHINE LEARNING**

Con los datos preprocesados se da paso a la elección del modelo de *machine learning* más adecuado para la predicción de la variable "Valor Aduana" que es la variable que más valor

aporta al cliente a la hora de realizar una transacción internacional ya que le sirve como referencia de cuáles serán los gastos estimados para una mercancía con unas características determinadas en la frontera aduanera.

A continuación, se va a exponer el proceso seguido hasta la elección final del modelo, sin embargo, el detalle de los resultados con el análisis individual de cada uno de los pasos se hará en el apartado 6. Análisis de los resultados.

La idea de esta fase del proyecto fue buscar el modelo más adecuado con las variables que realizasen la predicción más acertada posible de cuál sería el valor en aduana de una mercancía.

En primer lugar, se definen las variables de medición que fueron marcando la validez del modelo y las variables involucradas. Como variables de medición se usó en primer lugar el error cuadrático medio. En estadística, el error cuadrático medio (ECM) de un estimador mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima. El ECM es una función de riesgo, correspondiente al valor esperado de la pérdida del error al cuadrado o pérdida cuadrática. La diferencia se produce debido a la aleatoriedad o porque el estimador no tiene en cuenta la información que podría producir una estimación más precisa. (Lehmann & Casella, 1998)

Si bien el error cuadrático medio es la medida estándar a la hora de evaluar la eficacia de un modelo de *machine learning*, no siempre responde bien a los resultados esperados ya que perjudica mucho al modelo si existen errores muy notorios (por ejemplo, un modelo que predice perfectamente el 90% de la muestra pero falla en un 10% con errores muy alejados, tendrá un elevado error cuadrático medio desechando, a priori, la eficacia del modelo cuando la realidad es que es capaz de predecir con alta precisión el 90% de los datos). Para evitar este problema junto al error cuadrático medio se va a usar una variable que mida la distancia absoluta entre el dato real y su predicción. Esta información se mostrará a modo de histograma para cada una de las variables con el objetivo de que la gráfica tenga forma de “L” concluyendo así que el modelo tendrá una alta capacidad de predicción. En algunos casos, estos histogramas pueden acercarse a una forma de “U”, casos en los que el error

cuadrático medio castiga gravemente al modelo. Será en estos casos en los que habrá que evaluar las semejanzas y diferencias de los datos que se encuentran en un extremo u otro de la gráfica tratando de dar una explicación lógica y de negocio que nos permita seguir aceptando el modelo como buen predictor.

La iteración sobre modelos y variables irá de lo más sencillo a lo más complejo. Empezando en primer lugar con una regresión lineal que incluya únicamente aquellas variables numéricas que están mayormente correlacionadas con la variable a predecir. Se trata de ir ayudando al modelo al inicio para que prediga bien e ir incorporándole complejidad a medida que va aprendiendo y mejorando. Aunque se haya hecho un análisis exploratorio y se hayan transformado las variables en el apartado anterior para evitar introducirle complejidad al sistema se empezará por incluir las variables sin transformaciones

Con estos resultados, incluimos en el modelo únicamente las variables de interés y ejecutamos la regresión lineal. Al ser el primer acercamiento se espera un alto nivel de error cuadrático medio y una gráfica de “L” inversa en los histogramas de las distintas variables.

A partir de aquí, la idea es seguir incluyendo más variables en el modelo de regresión lineal y evaluando si mejoran o empeoran la capacidad de predicción del modelo. Se irán incluyendo también las variables categóricas usando, tal y como se ha explicado en el apartado 5.2 Limpieza y transformación de los datos, usando *label encoding* y *one hot encoding*.

Una vez se han evaluado las distintas combinaciones de variables con el modelo de regresión lineal y se obtiene la mejor combinación, se prueba a incluir transformaciones a las variables numéricas (transformaciones de logaritmo o potencia al cuadrado como las indicadas en el apartado anterior que favorezcan la normalidad de la distribución, simetría y carencia de curtosis)

Tras hacer distintas combinaciones de transformaciones se procede a normalizar las variables numéricas para evitar que unas variables obtengan mayor peso en el modelo que otras sin un motivo aparente. En este punto, se opta también por la creación de nuevas

variables que tenga un sentido de negocio como puede ser la masa específica de cada bulto de la mercancía (división entre la variable “Masa total” y “Número de bultos”) o el precio por kilogramo de la mercancía (dividiendo la variable “Factura total” y “Masa total”)

Una vez se han probado distintas combinaciones de variables, se han hecho transformaciones, se han escalado las variables numéricas y se han incluido nuevas variables de estudio, se escogerá la mejor combinación de todo lo anterior, aquella que tenga el menor error cuadrático medio o la que tenga un histograma de errores absolutos lo más cercano a una forma de “L”.

Así ya tenemos la combinación perfecta de variables en el modelo más simple de todos. Por ello, la idea que sigue es ir metiendo estas variables en diferentes modelos yendo del más simple (regresión lineal se ha tomado como punto de partida) hacia el más complejo y adoptar aquel que haga una mejor predicción de la variable “Valor Aduana” según los criterios de error cuadrático medio e histogramas de error absoluto.

En el capítulo 6 Análisis de los resultados se detallará más a fondo esta parte de iteración sobre modelos, pero se adelanta que el orden de las iteraciones será el siguiente:

1. Regresión lineal
2. Regresión polinómica de grado 2
3. Regresión polinómica de grado 3
4. Modelo de árbol de regresión
5. Modelo de *bagging* con n estimadores
6. Modelo de *gradient boosting*
7. Modelo de *random forest*

## **5.4 VISUALIZACIÓN DE LA INFORMACIÓN**

La visualización de la información se ha realizado a través de un reporte de *PowerBI* que el aporta al cliente información valiosa acerca de sus transacciones internacionales desde donde podrá filtrar la información según determinadas variables o características que le

proporcionan información y le permite realizar asociaciones, así como tener un control de todas sus transacciones.

Al mismo tiempo desde este mismo reporte el cliente será capaz de realizar una predicción introduciendo unos datos básicos con los que el modelo desarrollado será capaz de predecir el valor de la mercancía al llegar a la aduana teniendo en cuenta todas las variables que entran en juego.

El reporte está dividido en 8 apartados y una página exclusiva para la predicción del valor de la mercancía en la aduana. Los apartados son “Resumen”, “Circuitos”, “Representantes”, “Orígenes”, “Mercancías”, “Aranceles”; “Valor aduanas” y “Regímenes”.

En la página “Resumen” se hace una foto esquemática de las principales características de las importaciones/exportaciones. Se muestran KPI como la facturación total, el número total de declaraciones, el número de orígenes distintos de las mercancías, el número de aduanas de entrada, el importe total del IVA de importación y el importe total en derechos de aduana. Además, se muestra un mapa con los países de origen de las mercancías, así como una tabla resumen de todas las importaciones ordenadas por tipo de régimen aduanero consolidando datos como el importe de los derechos de importación, los derechos antidumping, el IVA de importación, el valor en aduana y el total de la factura. Todos los KPI y tablas de la página están, además, atados a unos filtros de país de origen o año de admisión que filtra toda la información mostrada según los filtros elegidos (véase [Ilustración 15](#))

La pestaña “Circuitos” muestra un *pie chart* con todas las transacciones distribuidas por el tipo de circuito (rojo, naranja, amarillo o verde), así como un KPI con el número total de DUA y un mapa con las aduanas de entrada. En la parte inferior se muestra una tabla con todos los DUA incluyendo información relevante tal como fecha de admisión, circuito, país de origen, aduana de entrada o modo de transporte en frontera, entre otros. Igual que en la página anterior el KPI, el gráfico, el mapa y la tabla están atadas a unos filtros por tipo de circuito, aduana de destino o año de admisión que filtran todos los datos de cada uno de los componentes (véase [Ilustración 16](#))



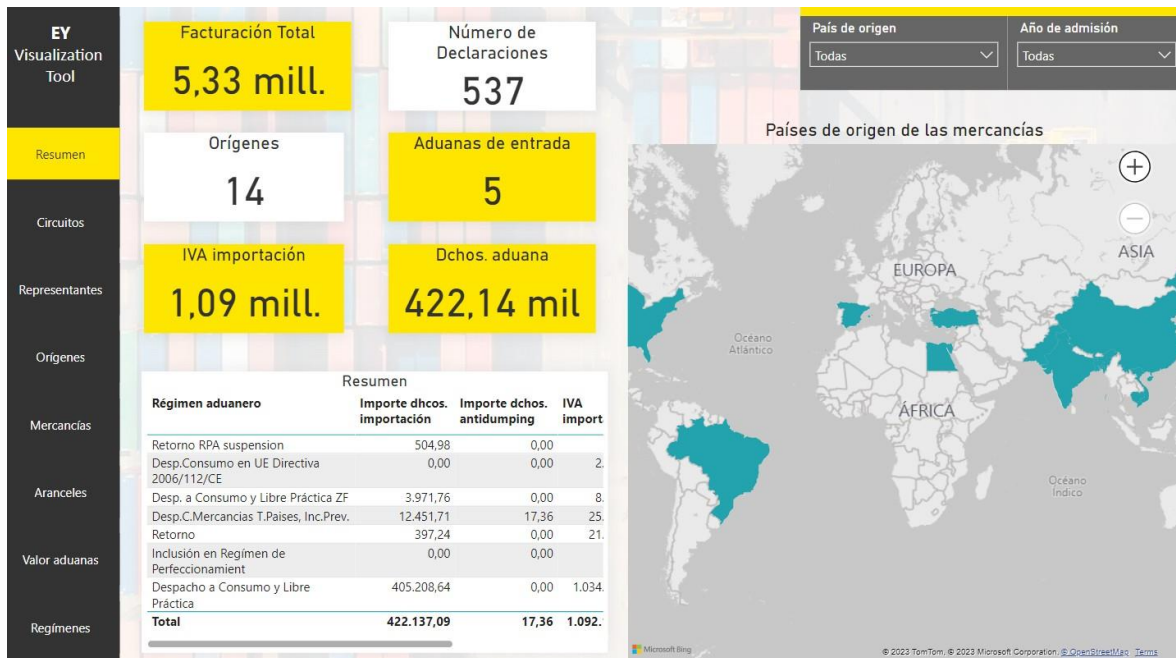


Ilustración 15 - Página "Resumen" del reporte de PowerBI

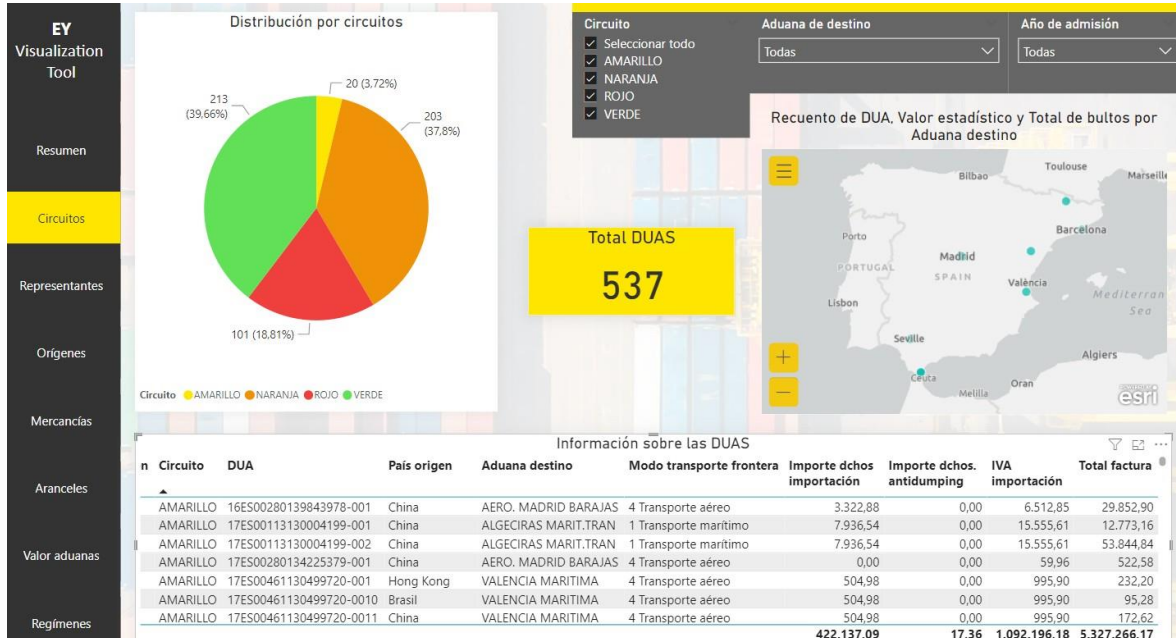
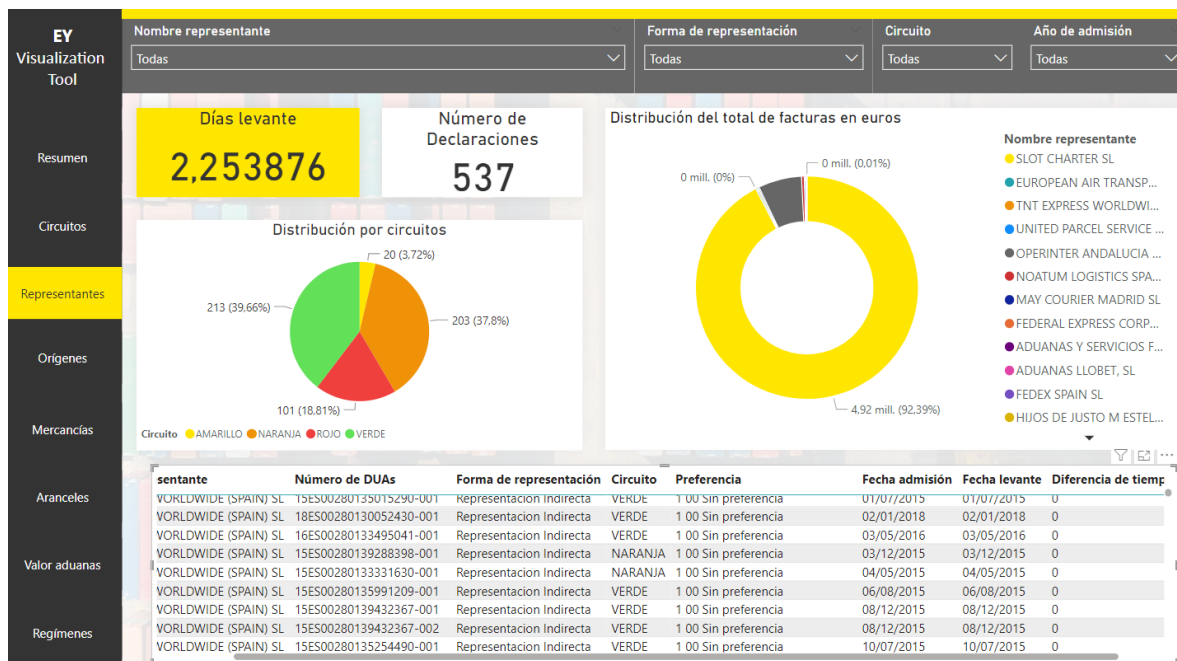


Ilustración 16 - Página "Circuitos" del reporte de PowerBI

La página "Representantes" contiene un KPI con el número de declaraciones y con los días de levante de la mercancía en media. El mismo *pie chart* que en la página anterior con los tipos de circuito y otro gráfico con la distribución del total de facturas en euros según el tipo

de representante. En la parte inferior se muestra una tabla con la información de los DUA con datos del representante, la forma de representación, el circuito, a preferencia, fechas de admisión y levante, así como la diferencia de tiempo entre ambas fechas en días. Como en las páginas anteriores todos los gráficos y tablas son dinámicos pudiendo filtrar en alguna sección del *pie chart* o en algún DUA de la tabla provocando que el resto de los componentes de la página filtra únicamente por el dato seleccionado. Además, existen en la parte superior filtros manuales tales como el nombre del representante, la forma de representación, el circuito o el año de admisión que filtran la información de toda la página (véase [Ilustración 17](#))



*Ilustración 17 - Página "Representantes" del reporte de PowerBI*

En la página “Orígenes” se muestran KPI como el valor total en aduana, el importe de los derechos de importación o el porcentaje de operaciones con preferencia. Aparecen también los KPI de DUA totales y el número de orígenes de las mercancías. Además, aparece un mapa que unen el país de destino (España para las importaciones) con el país de origen de las mercancías. Por último, se muestra una tabla con todos los DUA con información del país de origen, la preferencia de la mercancía, el porcentaje del valor en aduana respecto del total del valor en aduana de todas los DUA y el valor en aduana, entre otros. Se incluyen

también filtros al reporte dinámico como el país de origen, filtro por países FTA, preferencia de la mercancía o el año de admisión (véase [Ilustración 18](#))

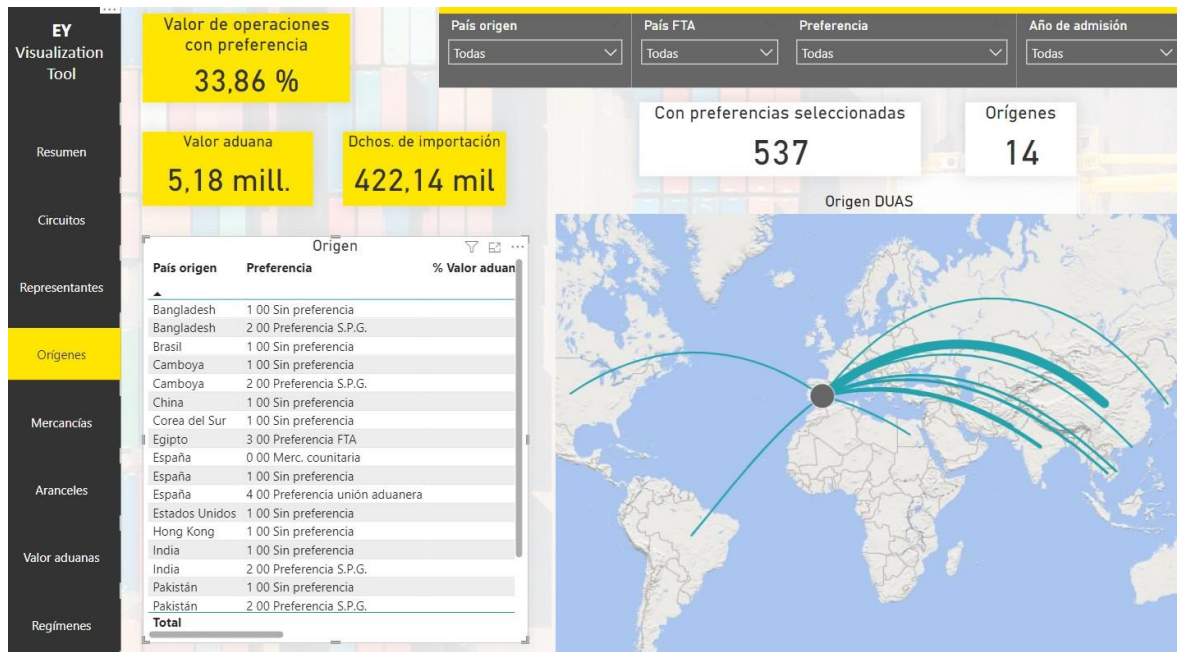
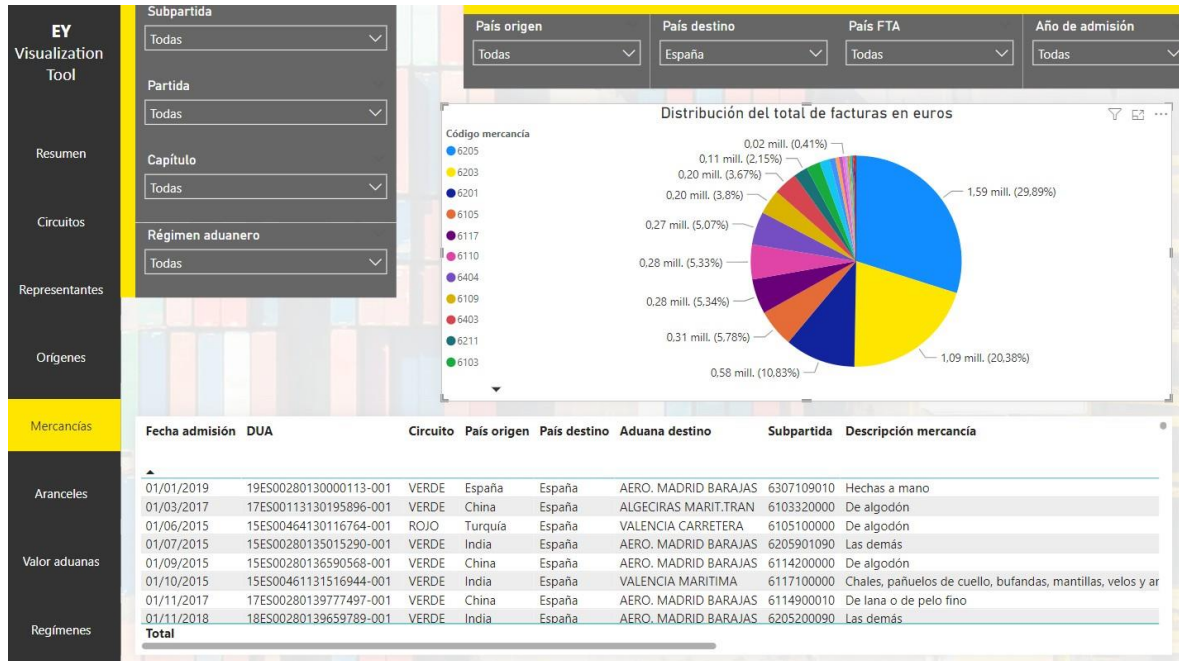


Ilustración 18 - Página "Orígenes" del reporte de PowerBI

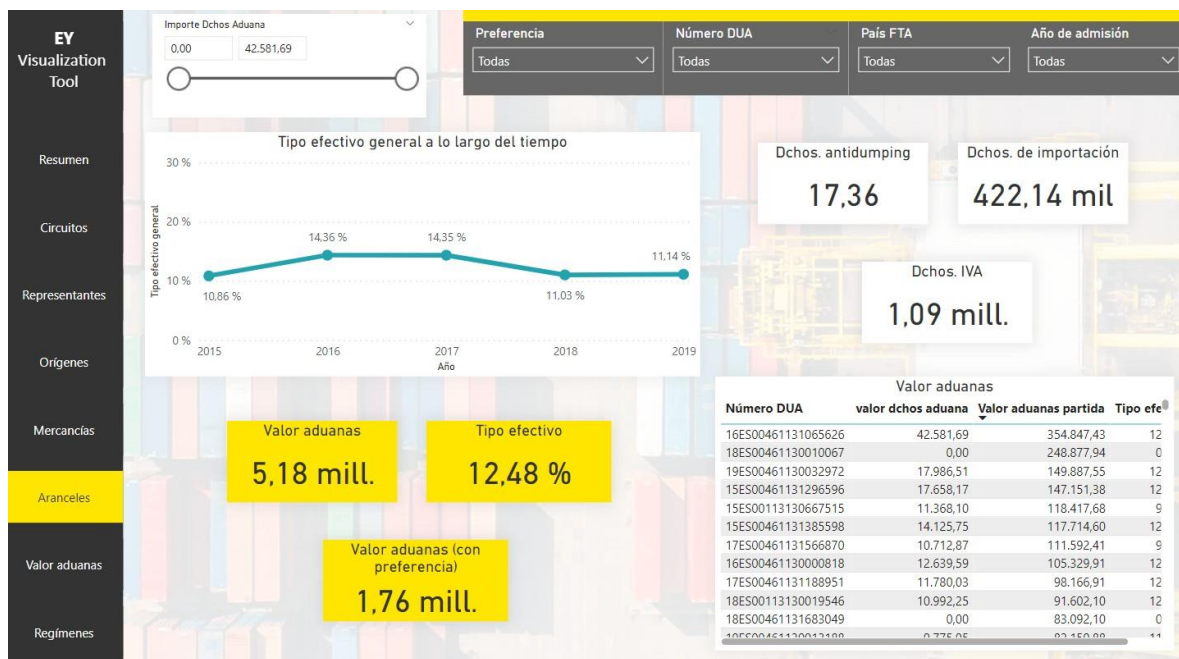
En la pestaña "Mercancías" se muestra un *pie chart* con la distribución del total de facturas en euros según el tipo de mercancía. En la parte inferior se muestra una tabla con todos los DUA con información acerca del circuito, el país de origen, la aduana de destino, descripción de la mercancía y valores numéricos como el valor en aduana, derechos de importación, antidumping o IVA de importación. En esta página existen filtros según la subpartida o la partida elegida, el capítulo del tipo de mercancía, el régimen aduanero, país de origen, país de destino, país FTA y el año de admisión (véase [Ilustración 19](#))

La página "Aranceles" muestra un gráfico con la evolución del tipo efectivo general a lo largo de los años de los datos de estudio. Se muestran KPI con los importes de los derechos antidumping, los derechos de importación y los derechos del IVA de importación. Se incluyen también el valor total en aduana, el tipo efectivo medio y el valor total en aduana de las mercancías con preferencia. Además, se incluye una tabla con los DUA e información como el importe de derechos en aduana, el valor en aduana o el tipo efectivo. Los filtros de

esta página serán la preferencia de las mercancías, el número de DUA, país FTA y el año de admisión (véase Ilustración 20)



*Ilustración 19 - Página "Mercancías" del reporte de PowerBI*



*Ilustración 20 - Página "Aranceles" del reporte de PowerBI*



La página “Valor Aduana” contiene KPI para los ajustes positivos y para los ajustes negativos. Se muestra a la derecha una gráfica con la proporción de ajustes según el tipo de *Incoterm*. En la parte inferior se muestra una tabla con los DUA, el *Incoterm* utilizado, ajustes positivos y negativos, el valor en aduana, el importe de derechos en aduana, el modo de transporte en la frontera y la naturaleza de la transacción. Los filtros de esta pestaña son la naturaleza de la transacción, el *Incoterm* utilizado, si se han realizado o no ajustes positivos o negativos y el año de admisión. (véase [Ilustración 21](#))

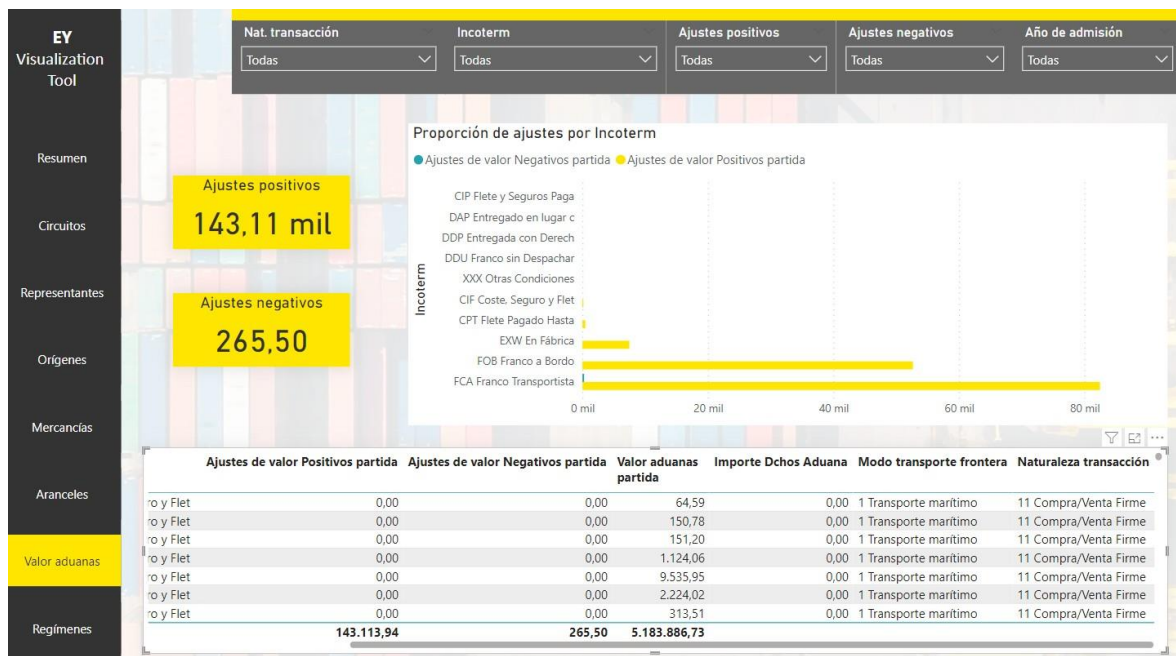


Ilustración 21 - Página "Valor Aduanas" del reporte de PowerBI

Por último, la pestaña “Regímenes” contiene un *pie chart* con la distribución del total de facturas en euros según el tipo de régimen aduanero. Además, se incluyen dos tablas, una con los códigos y nombre de régimen aduanero consolidando el valor en aduanas, y otra con información de los DUA, fecha de admisión, circuito, aduana de importación, país de origen y descripción de la mercancía (véase [Ilustración 22](#))

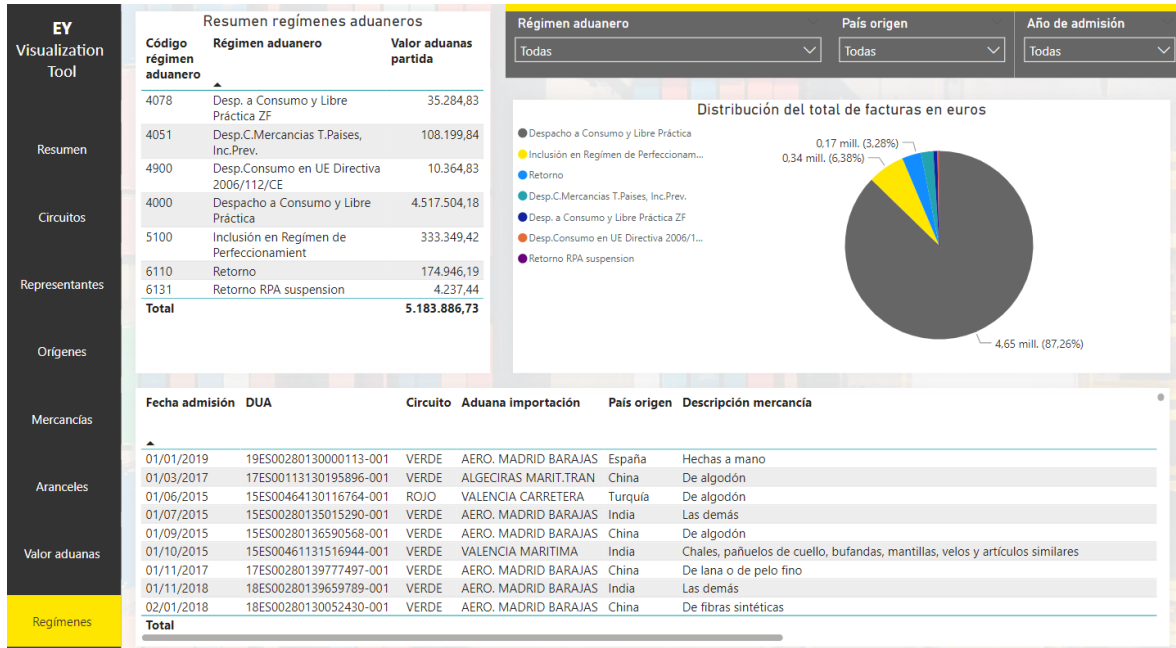


Ilustración 22 - Página "Regímenes" del reporte de PowerBI

## Capítulo 6. ANÁLISIS DE RESULTADOS

Este capítulo va a recorrer todas las iteraciones que se han ido haciendo a lo largo del desarrollo del sistema para la elección más adecuada del modelo de predicción de la variable “Valor Aduana”.

Como se ha explicado en el capítulo anterior, el proceso irá de lo más simple a lo más complejo. Es por esto por lo que se va a tratar de elegir la mejor combinación de variables usando en primer lugar únicamente un modelo de regresión lineal. Inicialmente incluiremos únicamente las variables numéricas. Para ello se hace un test de correlación entre las variables numéricas para elegir aquellas más correlacionadas con la variable a predecir con la finalidad de ayudar lo máximo posible al modelo.

	05. Número de Partidas:	06. Total de Bultos:	22.2 Total Divisas:	Total Factura en €:	35. Masa Bruta(Kg):	38. Masa Neta(kg):	41. Unidades Suplementarias :	Valor en Aduana:
05. Número de Partidas:	1							
06. Total de Bultos:	-0,096243417	1						
22.2 Total Divisas:	0,330609784	0,577905028	1					
Total Factura en €:	0,322835936	0,581189994	0,996699043	1				
35. Masa Bruta(Kg):	0,077006865	0,266666179	0,404625351	0,4003893	1			
38. Masa Neta(kg):	0,080756095	0,266816353	0,404445828	0,40077607	0,99835818	1		
41. Unidades Suplementarias :	-0,052384021	0,395213962	0,490991944	0,496057826	0,173925199	0,174058714	1	
Valor en Aduana:	-0,045505689	0,519364137	0,691236006	0,691913034	0,52200394	0,523222306	0,712879687	1

Ilustración 23 - Matriz de correlación de las variables numéricas

En la Ilustración 23 se puede apreciar como las variables “Total Divisas” y “Total Factura en €” están altamente correlacionadas entre ellas al aportar la misma información con un factor de conversión; por ello eliminaremos una de ellas del modelo. Lo mismo pasa con “Masa Bruta en Kg” y “Masa Neta en Kg”, la información es prácticamente idéntica de ahí su alta correlación, eliminaremos una de ellas también.

Por otro lado, se visualiza como las variables “Total Bultos”, “Total Factura en €”, “Masa bruta en Kg” y “Unidades Suplementarias” tienen una correlación superior al 50% con la variable a predecir “Valor en Aduana”, por lo que se comenzará por incluirlas en el modelo. Por su lado, la variable “Número Partidas” tiene apenas un 4% de correlación, por lo que se desechará. Así, se realiza la primera iteración con estas cuatro variables para un modelo de regresión lineal.

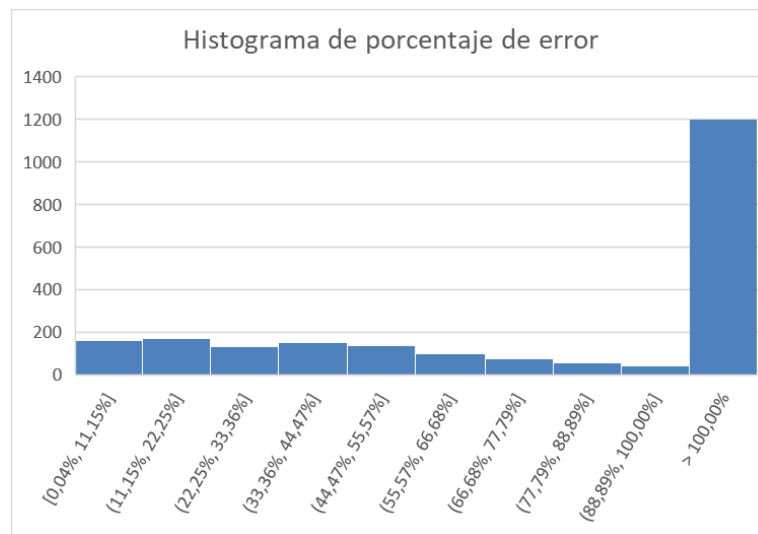


Ilustración 24 - Histograma de error de la 1ª iteración de regresión lineal

Tal y como esperábamos, los resultados de predicción del modelo fueron bastante pobres. Como se puede visualizar en la [Ilustración 24](#), el modelo predice con un porcentaje de error superior al cien por cien la mayor parte de los datos (forma de “L” invertida). El primer valor de la raíz del error cuadrático medio fue 21.959.

Así comenzamos a introducir variables en el modelo que mejoran su capacidad de predicción buscando un balance entre una mejora en el error cuadrático medio a la vez que se mejora el histograma del error absoluto. Por ejemplo, en la iteración número 7 se han introducido en el modelo las cuatro variables numéricas de la primera iteración junto con las variables “Declarante”, “Situación” y “Ciudad”. Al ser variables categóricas hubo que incluirlas con *one hot encoding* al no necesitar un orden lógico las categorías de cada una de las variables. En un primer momento la iteración iba por buen camino ya que se había conseguido reducir el error cuadrático medio a 20.915, sin embargo, al visualizar la gráfica se observó como los casos con error superior al 100% eran mayores que los de la iteración 3 (iteración con mejores resultados hasta el momento) y los casos con error inferior al 10% eran prácticamente los mismos. Por lo que se decidió no incluir la variable “Ciudad” en el modelo, a pesar de la mejora sustancial en el error cuadrático medio. (para ver la comparación de gráficas véase [Ilustración 25](#) e [Ilustración 26](#))



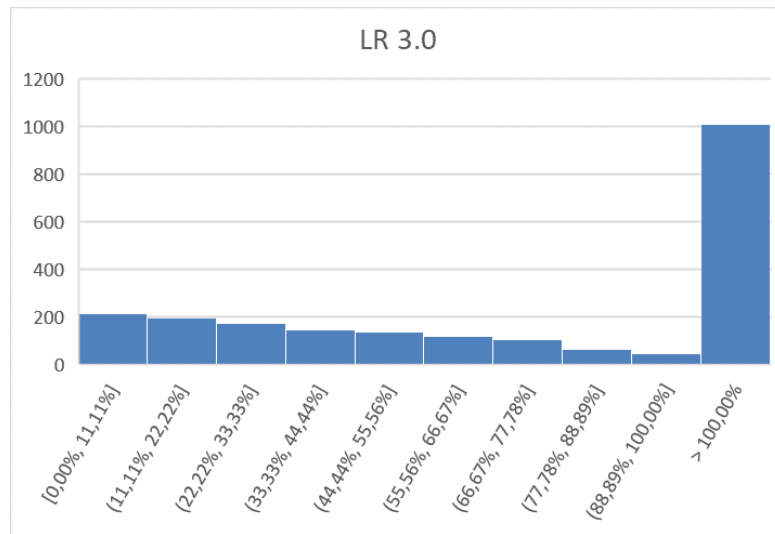


Ilustración 25 - Histograma de error de la 3ª iteración de regresión lineal

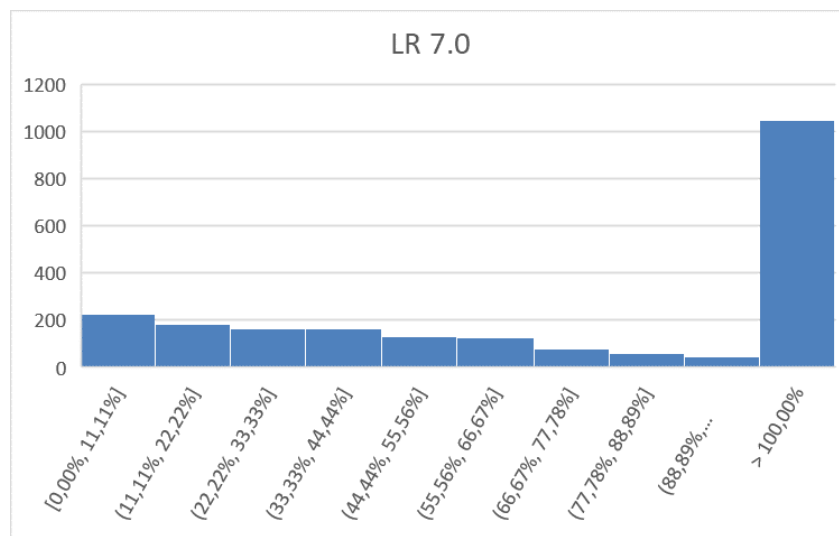


Ilustración 26 - Histograma de error de la 7ª iteración de regresión lineal

Todas las variables categóricas que se han ido probando para introducir en el modelo se han incluido con *one hot encoding*, excepto la variable “Circuito” que se introdujo con *label encoding* ya que las categorías “Verde”, “Amarillo”, “Naranja” y “Rojo” tenían un orden lógico que no podía perderse en la aleatoriedad.

También, muchas variables se desecharon previamente a cualquier iteración como “Aduana Destino” o “Transporte Interior” entre otros, por contener valores nulos o por ser variables categóricas en las que una de las categorías abarca más del 90% de los datos.

En la iteración 17 se consiguió la combinación con mejores resultados del modelo utilizando las variables existentes. En esta iteración se incluyeron las siguientes variables: “Total Bultos”, “Total Factura en €”, “Masa Bruta (Kg)”, “Unidades suplementarias”, “Declarante”, “Situación”, “Localización de las mercancías”, “Código TARIC”, “Circuito” y “Código Incoterms” (véase Ilustración 27)

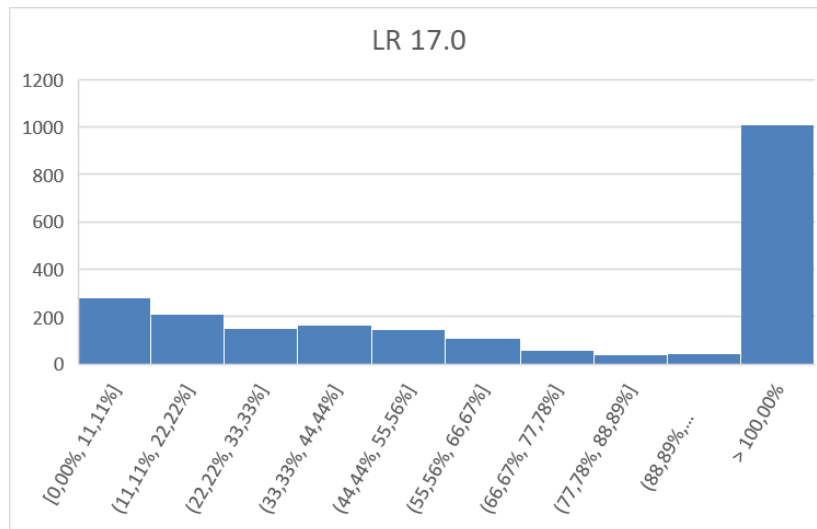
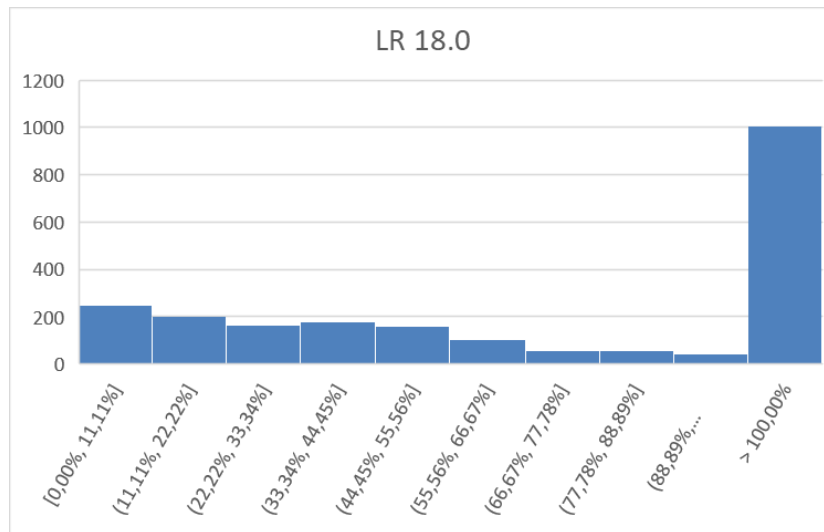


Ilustración 27 - Histograma de error de la 17ª iteración de regresión lineal

A partir de aquí, se incluyeron dos nuevas variables con sentido técnico y de negocio comentadas en el capítulo anterior, “Masa por bulto” y “€ por kg”. Tras la ejecución del modelo los resultados empeoraron en pequeña medida, por lo que se decidió no incluirlas en el mismo (véase Ilustración 28)



*Ilustración 28 - Histograma de error de la 18ª iteración de regresión lineal*

El siguiente proceso iterativo fue normalizar las variables independientes con el fin de que estuviesen todas evaluadas bajo la misma vara de medida. Los resultados del modelo ni mejoraron ni empeoraron, por lo que se decidió mantener la transformación de normalizar las variables.

Dado el análisis exploratorio de las variables explicado en el apartado 5.2 Limpieza y transformación de los datos se probó a aplicar transformaciones logarítmicas a las variables numéricas tratando de buscar la normalidad de las mismas. Los resultados resultaron ser contraproducentes con un error cuadrático medio de 23.582 y un histograma con mayor porcentaje de error. Se decidió no aplicar transformaciones sobre las variables independientes.

Sin embargo, se optó por aplicar una transformación logarítmica a la variable dependiente y los resultados fueron extraordinariamente buenos. El valor del error cuadrático medio descendió hasta un valor de 0.26 y el histograma de errores cogió una forma mucho más parecida a la de una “L” (véase Ilustración 29)

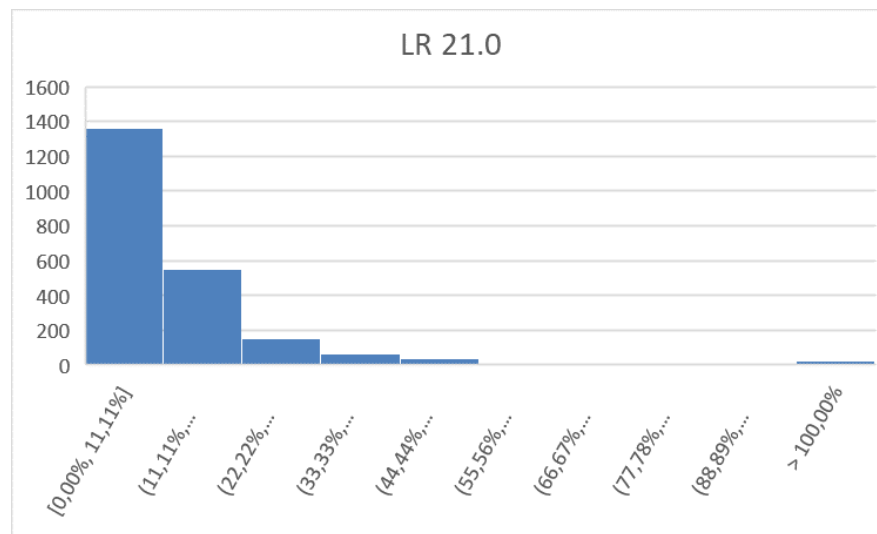


Ilustración 29 - Histograma de error de la 21ª iteración de regresión lineal

Ante este sorprendente resultado, se optó por empezar a dividir la muestra en entrenamiento y test para ver la capacidad real de predicción del modelo, ya que hasta ahora se estaban haciendo predicciones sobre los mismos datos de entrenamiento.

Al separar la muestra en entrenamiento y test, los resultados siguen siendo bastante convincentes. Si bien, el error cuadrático medio se ha disparado debido a varios errores muy notorios (19.709.324.891,2), aun así, el 84% de los valores se han predicho con un error menor al 23% y solo un 3% de los datos han tenido un error superior al 100% (véase [Ilustración 30](#)). De aquí en adelante se reducirá la escala del eje X de los gráficos (porcentaje de error) para poder apreciar más detalladamente las mejoras de unos modelos con otros.

Además, investigando un poco en el origen de los datos se ha descubierto uno de los posibles factores por los que el modelo no es capaz de predecir correctamente ese 3% de los datos. Por un lado, se ha visto que aquellas mercancías cuyo declarante es “*Transcoma*” o “*Cargored*” serán difícilmente evaluables debido al bajo número de registros que existen en la base de datos original de estos dos declarantes. Lo mismo pasa con la variable “Código TARIC”, al ser una variable bastante individual y propia de cada tipo de mercancía, aquellas cuyo código TARIC aparezca en muy pocos registros de la base de datos será difícilmente predecible su valor en aduana.

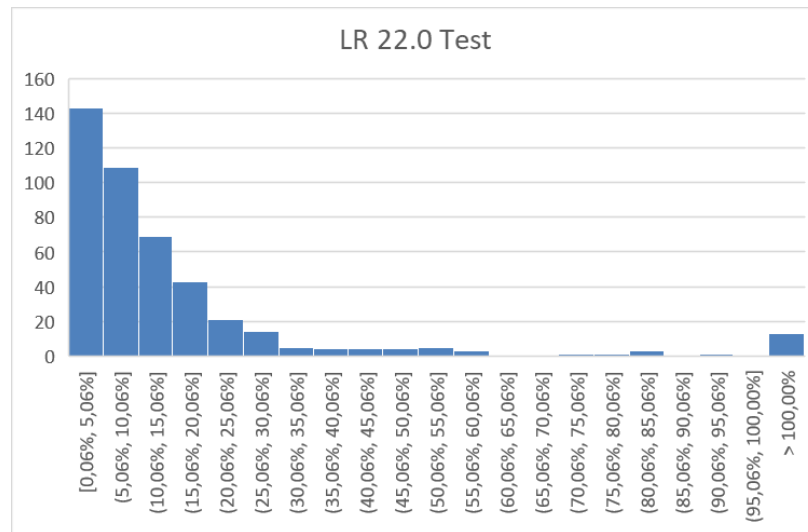


Ilustración 30 - Histograma de error de la iteración definitiva de regresión lineal con la muestra de test

El siguiente paso iterativo fue cambiar de tipo de modelo, de regresión lineal a regresión polinómica. Se empezó desarrollando un modelo de grado 2 cuyos resultados fueron bastante peores que los de la regresión lineal. Se optó por subirle el grado de la regresión polinómica a 3 con la esperanza de ver mejoras en la predicción. Sin embargo, los resultados fueron aún peores que con el modelo de grado 2 (véase [Ilustración 31](#) e [Ilustración 32](#))

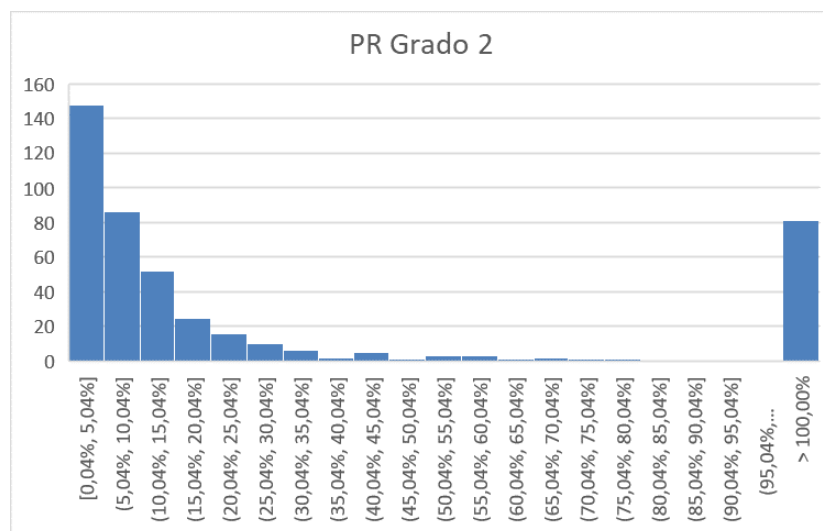


Ilustración 31 - Histograma de error para una regresión polinómica de grado 2

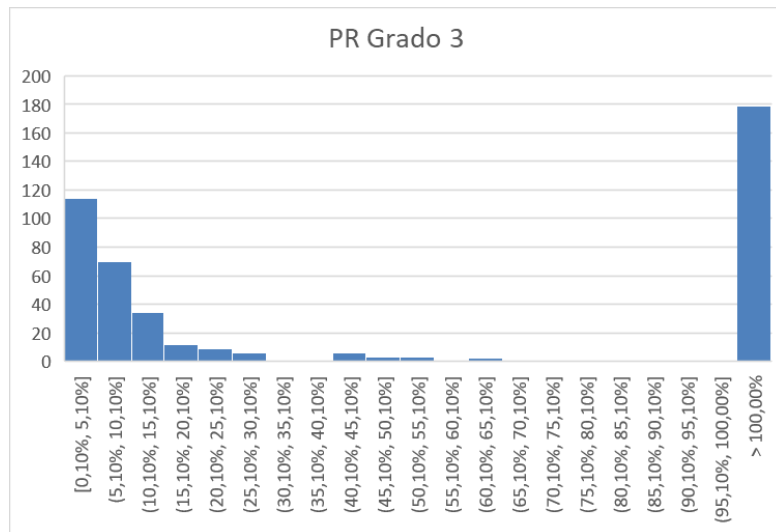


Ilustración 32 - Histograma de error para una regresión polinómica de grado 3

A continuación, se itera hacia un modelo de árbol de regresión. Los resultados mejoran en gran medida a los de la regresión lineal y consiguiendo eliminar los errores notorios (no se producen predicciones con un error superior al 77%). Esto se aprecia sobre todo en la mejora del error cuadrático medio que ahora se sitúa en 0,14. Además, los valores con un error menor al 11% han aumentado también respecto a la regresión lineal (véase Ilustración 33)

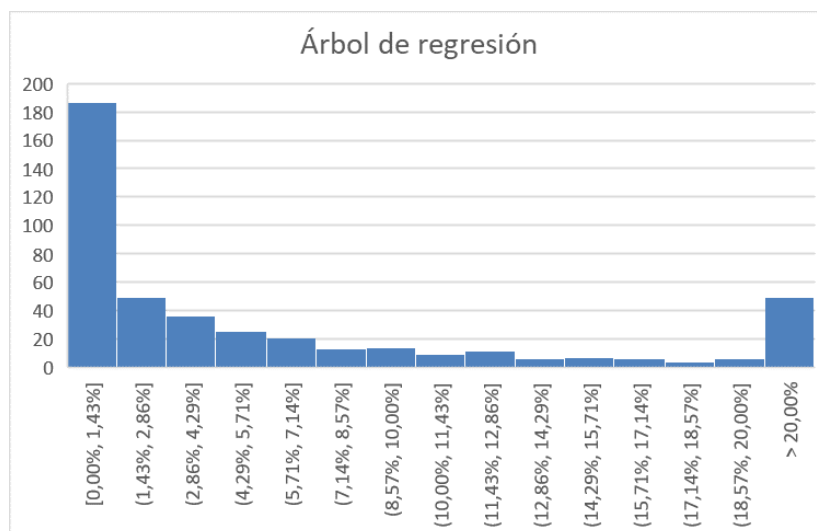


Ilustración 33 - Histograma de error para un árbol de regresión

La siguiente iteración se ha probado con un modelo de *bagging*. Para ello el primer paso es decidir cuál será el número óptimo de estimadores haciendo un balance entre el error cuadrático medio y la complejidad del sistema. Como se puede apreciar en la [Ilustración 34](#) referente a la gráfica de codo que representa el error cuadrático medio para cada número de estimadores, el número de estimadores óptimo podría ser 4 o 5. Dado que la diferencia entre ambos es mínima se optará por usar 4 estimadores quitándole así complejidad al modelo.

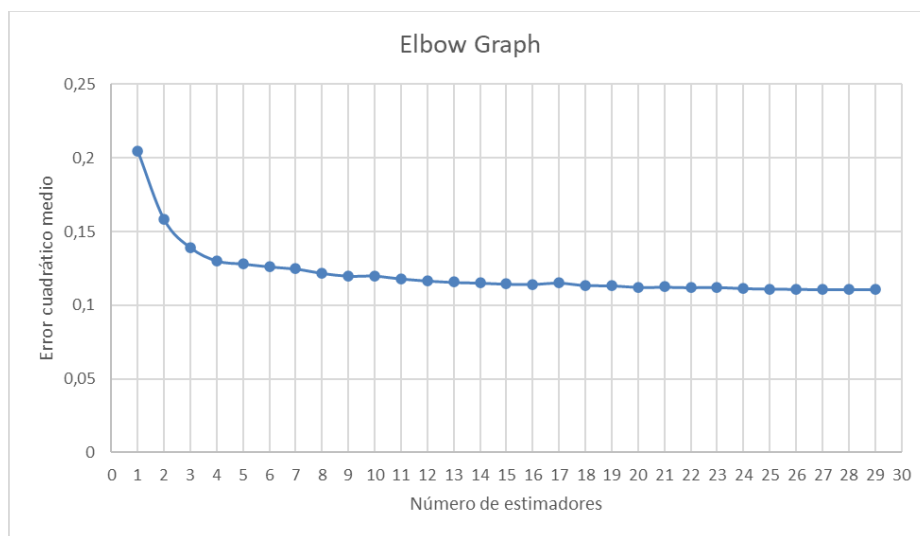
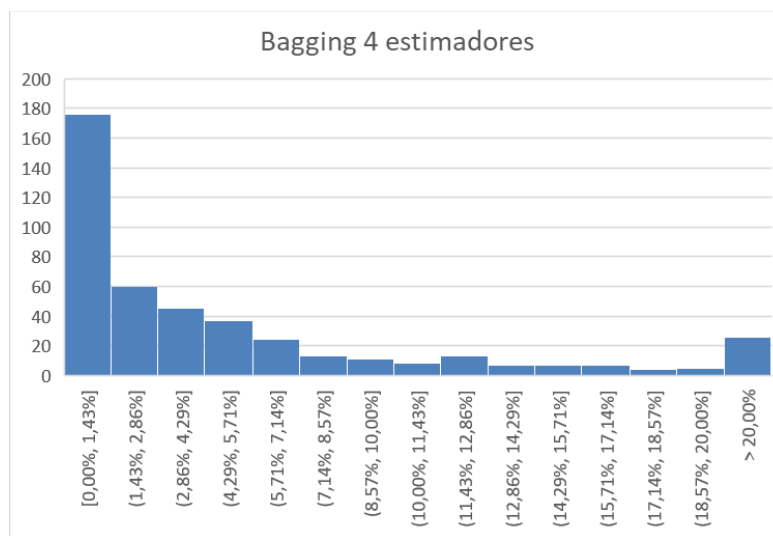


Ilustración 34 - Gráfica de codo para elegir el número óptimo de estimadores

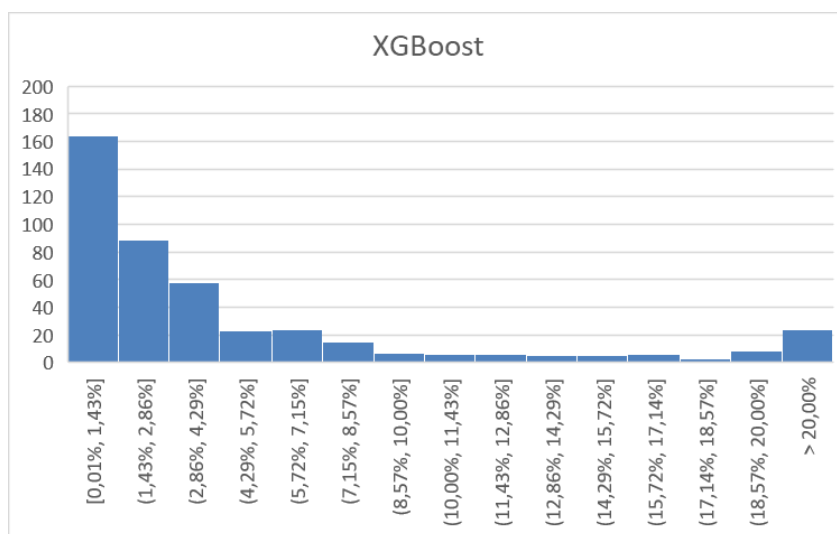
Con este resultado se lanza un modelo de *bagging* con 4 estimadores. El error cuadrático medio mejora ligeramente respecto al modelo de un árbol de regresión con un valor de 0,09, sin embargo, el histograma de errores no refleja una clara mejora. Como se puede apreciar en la [Ilustración 35](#) el número de predicciones con error menor al 1,5% es menor, aunque si es cierto que los dos rangos de error aumentan y el número de predicciones con un error mayor al 20% disminuye en un 5% de los casos. De momento, continuamos iterando el modelo antes de tomar una decisión.

El siguiente modelo que se va a desarrollar será un modelo específico de *boosting* denominado *gradient boosting* (potenciación del gradiente), en concreto se usará la librería de Python XGBoost que es una mejora de este tipo de modelos. El modelo mejora ligeramente el error cuadrático medio respecto al árbol de regresión o el modelo de *bagging*

con un valor de 0,06. Este caso es parecido al del modelo de *bagging* si hacemos la comparación con el modelo de árbol de regresión ya que las predicciones con error menor al 1,5% son menores, sin embargo, el número de predicciones con un error mayor al 20% es considerablemente menor. Se decide que XGBoost predice peor que *bagging* por el factor de que el número de predicciones con error menor a 1,5% disminuyen en un 11% respecto al árbol de decisión (véase [Ilustración 36](#))



*Ilustración 35 - Histograma de error para el modelo de bagging con 4 estimadores*



*Ilustración 36 - Histograma de error para el modelo de XGBoost*



Por último, la iteración final antes de decidir el modelo a utilizar será un modelo *random forest*. Para este caso, el error cuadrático medio se mantiene en 0,06, siendo un valor realmente bueno ya que lo viene a decir es que el modelo es capaz de predecir cualquier valor con un error en media del 6%. Además, si se observa la Ilustración 37, se aprecia como en comparación árbol de decisión, *bagging*, XGBoost y *random forest*, es este último junto con el árbol de regresión el que más predicciones con error inferior a 1,5% tiene. Además, tiene la ventaja sobre el árbol de decisión de que los valores con errores superiores al 20% son un 5% menor.

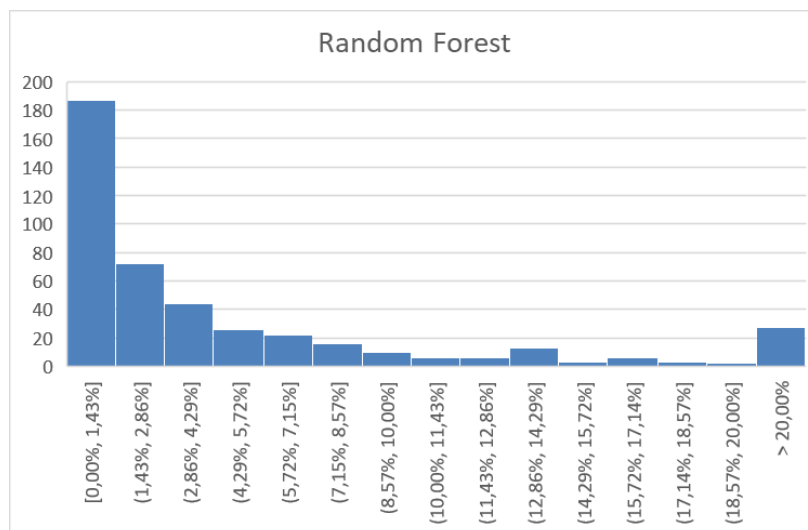


Ilustración 37 - Histograma de error para el modelo de Random Forest

Con esto, se decide usar un modelo de *random forest* para la predicción de la variable “Valor en Aduana” por tener un error cuadrático medio muy reducido (6,59%) y porque el histograma que representa la diferencia de error entre todos los valores a predecir en relación a los valores reales del conjunto de test tiene una forma muy cercana a una “L”. Este histograma acumula la mayor parte de predicciones (42% del conjunto) en un rango de error menor al 1,5%, además de que el 85% del conjunto es predicho por el modelo con un error menor al 10%. Al mismo tiempo, las predicciones con un error superior al 20% suponen únicamente un 6% del conjunto, que coinciden con aquellas muestras cuyas variables “Código TARIC” y “Declarante” son minoritarias (1 o 2 registros en toda la muestra) lo que provoca que el modelo no sea capaz de predecirlos con exactitud por la falta de datos de este

estilo (al igual que pasaba con el modelo de regresión lineal tras aplicar la separación de datos en entrenamiento y test)

## Capítulo 7. CONCLUSIONES Y TRABAJOS FUTUROS

El comercio internacional es una práctica que, en los tiempos que corren, es crucial para gran parte de las compañías y países del mundo debido al fenómeno de la globalización. Al ser una práctica tan extendida, su estudio y análisis puede significar un punto clave para las empresas a la hora de colocarse en una posición ventajosa en comparación con sus competidores. De esta idea se desprende la implementación del sistema desarrollado que busca aconsejar a las empresas para que con las herramientas de visualización, análisis y predicción del sistema sean capaces de tomar las decisiones más acertadas a la hora de posicionarse en el mercado.

A pesar del desarrollo del actual sistema, en el comercio exterior intervienen infinidad de variables y constantes que hacen que un sistema como el propuesto se haga extremadamente largo y complejo. Así, el sistema desarrollado se ha centrado en el ámbito fiscal de una mercancía en su llegada a una aduana de destino, más concretamente en la información provista por los DUA de las transacciones en las que se incluye la información más relevante acerca de las mercancías involucradas en la transacción.

El sistema desarrollado extrae automáticamente la información de los DUA de una compañía directamente de la página web de la AEAT desde que se tienen registro. Esta información es explorada, tratada, limpiada y transformada teniendo en cuenta tanto el punto de vista técnico (adecuando la información al modelo de *machine learning*) como desde el punto de vista de negocio o fiscal (adecuando la información a los objetivos de negocio propuestos) Con esta información se desarrolla un modelo de predicción con *machine learning* con un elevado grado de precisión al mismo tiempo que es capaz de generalizar para cualquier dato de entrada (sin errores y sin sobreajuste) Finalmente, los datos son expuestos en un cuadro de mandos desde que el que se puede visualizar y hacer un análisis de toda la información. Al mismo tiempo, se puede realizar una predicción en tiempo real de cuál va a ser el valor en aduana de una mercancía con unas características determinadas.

Con este sistema, la empresa es capaz de visualizar y analizar toda la información relativa a sus transacciones internacionales haciendo uso de KPI y filtros que le permiten simplificar el análisis de los datos. Este análisis de la información en conjunto con la herramienta de predicción en tiempo real proporciona a la compañía los útiles necesarios para tomar decisiones de negocio que les posicionen en un lugar de ventaja respecto de sus competidores. Así, por ejemplo, la empresa podrá analizar con que representantes o declarantes las mercancías reducen sus días de levante, qué regímenes aduaneros son los que más le convienen en cuanto a los costes de los derechos de importación o antidumping, qué países de origen o preferencias son los que reducen su valor en aduana, o analizar los circuitos que reciben las distintas mercancías en función de sus características.

En conclusión, el sistema desarrollado es capaz de identificar gastos que son potencialmente reducibles, al mismo tiempo que es capaz de realizar predicciones en tiempo real de los costes asociados creando una mayor eficiencia a la hora de realizar una transacción internacional, ya que la empresa es capaz de anticiparse al mercado. Todo esto se traduce en un aumento de la competitividad de la empresa al colocarse en una posición de ventaja respecto de sus competidores.

A pesar de todos los objetivos cubiertos por el sistema se propone como trabajos futuros un desarrollo más profundo tanto desde el lado de los datos (visión fiscal) como desde el lado del modelo (visión técnica) Desde el punto de vista de la información, sería interesante poder cruzar información con más fuentes de datos para aumentar la robustez del modelo. Como se ha comentado al inicio de esta sección, el comercio exterior involucra infinidad de variables por lo que las fuentes de datos podrían ir desde una introducción detallada de todos y cada uno de los códigos TARIC de las mercancías hasta la inclusión de las tasas de flete y transporte de mercancías según los distintos medios de transporte, distancias y mercancías en sí. Desde el punto de vista del modelo, sería interesante también realizar un modelo de clasificación o *clusters*. Con un modelo de este estilo se agruparían las distintas transacciones en grupos lógicos similares que servirían para aumentar la capacidad de análisis de la herramienta pudiendo analizar más en detalle las características propias de cada grupo.

Así, con la resolución de este sistema y los trabajos futuros se puede concluir que, adoptando el poder del aprendizaje automático en el comercio internacional, se puede desbloquear un mundo de posibilidades ilimitadas, donde los conocimientos basados en datos allanan el camino para tomar decisiones más inteligentes, mejorar la eficiencia y lograr un crecimiento sin precedentes en el mercado global.

## Capítulo 8. BIBLIOGRAFÍA

- Abdullaiev, O., & Ibrahimova, L. V. (2020). *Review of testing software: Selenium software*.  
Obtenido de Universidad Técnica Nacional de Vinnytsia:  
<http://ir.lib.vntu.edu.ua/bitstream/handle/123456789/29630/9920.pdf?sequence=3>
- Across Logistics. (Marzo de 2023). *DUA Documento Único Administrativo. Qué es y cuándo utilizarlo*. Obtenido de Across Logistics: <https://acrosslogistics.com/blog/dua-documento-unico-administrativo>
- Adeva, J. J., Cerviño, U., & Calvo, R. (2005). Accuracy and Diversity in Ensembles of Text Categorisers. *CLEI Journal*, 1-12.
- Agbonlahor, W. (Octubre de 2015). *OECD unveils measures against corporate tax avoidance*. Obtenido de Global Government Forum:  
<https://www.globalgovernmentforum.com/oecd-unveils-measures-against-corporate-tax-avoidance/>
- Arrondo, A. (Noviembre de 2019). *Los circuitos aduaneros*. Obtenido de Aduana Fácil:  
<https://www.aduanafacil.es/2019/11/21/circuitos-aduaneros-importacion/>
- Bertrand, T. J., Wonnacott, P., Allais, M., Balassa, B., & Robinson, R. (Junio de 2023). *International trade*. Obtenido de Encyclopedia Britannica:  
<https://www.britannica.com/money/topic/international-trade/additional-info#contributors>
- Breiman, L. (1996). *BIAS, VARIANCE, AND ARCING CLASSIFIERS*.
- Brown, G., Wyatt, J., Harris, R., & Yao, X. (2005). Diversity creation methods: a survey and categorisation. *Information Fusion*, 5-20.

- Cámaras de Comercio. Servicios de Estudios. (2003). *Fiscalidad del Comercio Exterior*. Madrid: Cámara de Comercio de España.
- Chang, Y.-W., Hsieh, C.-J., Chang, K.-W., Ringgaard, M., & Lin, C.-J. (2010). Training and testing low-degree polynomial data mappings via linear SVM. *Journal of Machine Learning Research*, 1471–1490.
- Cook, R. D., & Weisberg, S. (1982). Criticism and Influence Analysis in Regression. En R. D. Cook, & S. Weisberg, *Sociological Methodology* (págs. 313-361).
- Delaney, L. (Diciembre de 2004). *20 Factors to Consider Before Going Global*. Obtenido de Entrepreneur: <https://www.entrepreneur.com/growing-a-business/20-factors-to-consider-before-going-global/75138>
- DSV. (2020). *Tipos de Incoterms® 2020 - Incoterms en español*. Obtenido de DSV: <https://www.dsv.com/es-es/ayuda/faq/tipos-de-incoterms-2020>
- EAE Business School. (Abril de 2017). *Factors to consider before expanding your business internationally*. Obtenido de EAE: <https://www.eae.es/en/news/eae-news/factors-consider-expanding-your-business-internationally>
- European Commission. (s.f.). *TARIC*. Obtenido de Taxation and Customs Union: [https://taxation-customs.ec.europa.eu/customs-4/calculation-customs-duties/customs-tariff/eu-customs-tariff-taric\\_en](https://taxation-customs.ec.europa.eu/customs-4/calculation-customs-duties/customs-tariff/eu-customs-tariff-taric_en)
- Freedman, D. A. (2009). *Statistical Models: Theory and Practice*. Cambridge University Press.
- Fushiki, T. (2010). Bayesian bootstrap prediction. *Journal of Statistical Planning and Inference*, 65-74.
- Gartner, Inc. (s.f.). *Magic Quadrant for Business Intelligence and Analytics Platforms*". Obtenido de Gartner: <https://www.gartner.com/doc/reprints?id=1-2XYY9ZR&ct=160204&st=sb>

- Gashler, M., Giraud-Carrier, C., & Martinez, T. (2008). Decision Tree Ensemble: Small Heterogeneous Is Better Than Large Homogeneous. *The Seventh International Conference on Machine Learning and Applications*, 900–905.
- Gibb, L. (s.f.). *What to consider when expanding a business internationally*. Obtenido de Scaling Partners: <https://www.scaling.partners/resources/articles/what-to-consider-when-expanding-a-business-internationally/>
- Global HR Practices. (s.f.). *Understanding The Role Of Culture In International Business*. Obtenido de Multiplier: <https://www.usemultiplier.com/blog/culture-in-international-business>
- Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning*. Springer.
- Heakal, R. (Marzo de 2023). *International (Global) Trade: Definition, Benefits, Criticisms*. Obtenido de Investopedia: <https://www.investopedia.com/insights/what-is-international-trade/>
- Ho, T. (1995). Random Decision Forests. *Proceedings of the Third International Conference on Document Analysis and Recognition*, 278-282.
- Ho, T. K. (1995). Random Decision Forests. *3rd International Conference on Document Analysis and Recognition*, (págs. 14-16). Montreal.
- Ho, T. K. (1998). The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 832-844.
- Hunter, W. (Agosto de 2021). *What Is Web Scraping? A Complete Beginner's Guide*. Obtenido de Career Foundry: <https://careerfoundry.com/en/blog/data-analytics/web-scraping-guide/>
- Infoautónomos. (Septiembre de 2022). *Impuestos y trámites aduaneros para exportar e importar*. Obtenido de Infoautónomos:



<https://www.infoautonomos.com/fiscalidad/impuestos-y-tramites-aduaneros-para-exportar-e-importar/>

Kearns, M. (1988). *Thoughts on Hypothesis Boosting*.

Kearns, M., & Valiant, L. (1989). Cryptographic limitations on learning Boolean formulae and finite automata. Symposium on Theory of Computing. En M. Kearns, & L. Valiant, *Cryptographic limitations on learning Boolean formulae and finite automata. Symposium on Theory of Computing* (págs. 433-444). ACM.

Kuncheva, L. a. (2003). Measures of diversity in classifier ensembles. *Machine Learning*, 181-207.

Lehmann, E. L., & Casella, G. (1998). *Theory of Point Estimation*. Nueva York: Springer.

Liu, Y., & Yao, X. (1999). Ensemble learning via negative correlation. *Neural Networks*, 1399–1404.

Madeh, P. S., & E., E.-D. T. (2020). Role of Data Analytics in Infrastructure Asset Management: Overcoming Data Size and Quality Problems. *Journal of Transportation Engineering, Part B: Pavements*, 146.

Microsoft. (s.f.). *Bring your data to life with Microsoft Power BI*. Obtenido de Microsoft: <https://powerbi.microsoft.com/en-us/>

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of Machine Learning*. The MIT Press.

Morales, F. C. (Abril de 2020). *Doble imposición*. Obtenido de Economipedia: <https://economipedia.com/definiciones/doble-imposicion.html>

Mukewar, P. (Septiembre de 2022). *How Technology Can Help Global Trade Become More Efficient, Inclusive And Equitable*. Obtenido de Forbes:

- <https://www.forbes.com/sites/forbesfinancecouncil/2022/09/02/how-technology-can-help-global-trade-become-more-efficient-inclusive-and-equitable/>
- Núñez, E. A. (Mayo de 2016). *Qué es la Metodología Agile*. Obtenido de OpenWebinars: <https://openwebinars.net/blog/que-es-la-metodologia-agile/>
- Oftex International Sales. (Mayo de 2022). *Main barriers to international trade*. Obtenido de Oftex: <https://www.oftex.es/en/main-barriers-to-international-trade/>
- Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 169-198.
- Piryonesi, S. M., & El-Diraby, T. E. (2021). Using Machine Learning to Examine Impact of Type of Performance Indicator on Flexible Pavement Deterioration Modeling. *Journal of Infrastructure Systems*, 27.
- Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 21-45.
- Radcliffe, B. (Abril de 2023). *The Basics of Tariffs and Trade Barriers*. Obtenido de Investopedia: <https://www.investopedia.com/articles/economics/08/tariff-trade-barrier-basics.asp#toc-who-collects-a-tariff>
- Ramírez, C. M. (Octubre de 2021). *¿Qué es la fiscalidad internacional?* Obtenido de UNIR: <https://www.unir.net/derecho/revista/fiscalidad-internacional-que-es/>
- Robinson, R. A., Gómez, E. F., & Ruiz, I. P. (2015). *Economía Política Internacional*. Madrid.
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 1-39.
- Russell, S. J., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Prentice Hall.

- Sánchez, J. (Junio de 2015). *10 consejos para la internacionalización de la empresa*. Obtenido de Pymes y Autónomos: [vhttps://www.pymesyautonomos.com/vocacion-de-empresa/10-consejos-para-la-internacionalizacion-de-la-empresa](https://www.pymesyautonomos.com/vocacion-de-empresa/10-consejos-para-la-internacionalizacion-de-la-empresa)
- Santander Trade Markets. (2020). *Incoterms*. Obtenido de Santander Trade: [https://santandertrade.com/es/portal/analizar-mercados/incoterms?url\\_de\\_la\\_page=%2Fes%2Fportal%2Fanalizar-mercados%2Fincoterms&&actualiser\\_id\\_banque=oui&id\\_banque=0&memoriser\\_choix=memoriser](https://santandertrade.com/es/portal/analizar-mercados/incoterms?url_de_la_page=%2Fes%2Fportal%2Fanalizar-mercados%2Fincoterms&&actualiser_id_banque=oui&id_banque=0&memoriser_choix=memoriser)
- Seal, H. L. (1967). *The historical development of the Gauss linear mode*. *Biometrika*.
- Selenium. (Marzo de 2020). *The Selenium Browser Automation Project*. Obtenido de Selenium Dev: <https://www.selenium.dev/documentation/>
- Shoham, R., & Permuter, H. (2019). Amended Cross-Entropy Cost: An Approach for Encouraging Diversity in Classification Ensemble (Brief Announcement). *Cyber Security Cryptography and Machine Learning*, 202-207.
- Sollich, P. a. (1996). Learning with ensembles: How overfitting can be useful. *Advances in Neural Information Processing Systems*, 190-196.
- Stock logistic. (Octubre de 2018). *El despacho de aduanas: el canal verde, amarillo, naranja y rojo*. Obtenido de tock logistic: <https://www.stocklogistic.com/el-despacho-de-aduanas-el-canal-verde-amarillo-naranja-y-rojo/>
- Studer, M., Ritschard, G., Gabadinho, A., & Müller, N. S. (2011). Discrepancy Analysis of State Sequences. *Sociological Methods & Research*, 471–510.
- Tarazona, K. (2022). *El comercio internacional como motor del desarrollo sostenible*. Obtenido de Centro de Estudios Estratégicos de Relaciones Internacionales: <https://www.ceeriglobal.org/el-comercio-internacional-como-motor-del-desarrollo-sostenible/>

- Veigler Business School. (Octubre de 2022). *Código TARIC: qué es y cómo debe interpretarse*. Obtenido de Veigler Business School: <https://veigler.com/codigo-taric-interpretacion/#:~:text=%C2%BFQu%C3%A9%20es%20el%20c%C3%B3digo%20TARIC%3F%20La%20Tarifa%20Integrada,la%20factura%20proforma%20que%20acompa%C3%B1a%20la%20mercanc%C3%ADa>.
- Wikipedia. (Octubre de 2022). *Polynomial regression*. Obtenido de Wikipedia: [https://en.wikipedia.org/wiki/Polynomial\\_regression](https://en.wikipedia.org/wiki/Polynomial_regression)
- Wikipedia. (Mayo de 2023). *Boosting\_(machine\_learning)*. Obtenido de Wikipedia: [https://en.wikipedia.org/wiki/Boosting\\_\(machine\\_learning\)](https://en.wikipedia.org/wiki/Boosting_(machine_learning))
- Wikipedia. (Mayo de 2023). *Bootstrap aggregating*. Obtenido de Wikipedia: [https://en.wikipedia.org/wiki/Bootstrap\\_aggregating](https://en.wikipedia.org/wiki/Bootstrap_aggregating)
- Wikipedia. (Mayo de 2023). *Decision\_tree\_learning*. Obtenido de Wikipedia: [https://en.wikipedia.org/wiki/Decision\\_tree\\_learning](https://en.wikipedia.org/wiki/Decision_tree_learning)
- Wikipedia. (Junio de 2023). *Ensemble\_learning*. Obtenido de Wikipedia: [https://en.wikipedia.org/wiki/Ensemble\\_learning](https://en.wikipedia.org/wiki/Ensemble_learning)
- Wikipedia. (Mayo de 2023). *Linear regression*. Obtenido de Wikipedia: [https://en.wikipedia.org/wiki/Linear\\_regression](https://en.wikipedia.org/wiki/Linear_regression)
- Wikipedia. (Junio de 2023). *PowerBI*. Obtenido de Wikipedia: [https://en.wikipedia.org/wiki/Microsoft\\_Power\\_BI](https://en.wikipedia.org/wiki/Microsoft_Power_BI)
- Wikipedia. (Junio de 2023). *Random Forest*. Obtenido de Wikipedia: [https://en.wikipedia.org/wiki/Random\\_forest](https://en.wikipedia.org/wiki/Random_forest)
- Wikipedia. (Junio de 2023). *Regression analysis*. Obtenido de Wikipedia: [https://en.wikipedia.org/wiki/Regression\\_analysis](https://en.wikipedia.org/wiki/Regression_analysis)

- Wikipedia. (Marzo de 2023). *Supervised learning*. Obtenido de Wikipedia:  
[https://en.wikipedia.org/wiki/Supervised\\_learning](https://en.wikipedia.org/wiki/Supervised_learning)
- Wrike. (2023). *Fundamentos de la metodología Agile*. Obtenido de Wrike:  
<https://www.wrike.com/es/project-management-guide/fundamentos-de-la-metodologia-agile/>
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., . . . Zhou, Z.-H. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 1–37.
- Zhi-Hua, Z. (2012). *Ensemble Methods: Foundations and Algorithms*. *Chapman and Hall/CRC*, 23.

## ANEXO I

### 8.1 TIPOS DE INCOTERMS 2020 (SANTANDER TRADE MARKETS, 2020)

#### 8.1.1 EXW (EX WORKS / EN FÁBRICA)

##### Vendedor

La única responsabilidad del vendedor es poner la mercancía a disposición del comprador en sus locales en un embalaje adaptado al tipo de transporte (en general, el precio incluye la ubicación de la mercancía en *palets*).

##### Comprador

El comprador asume todos los gastos y riesgos inherentes al transporte, desde la salida de la fábrica hasta el destino. El término EXW representa la menor cantidad de obligaciones para el vendedor. Sin embargo, si las partes desean que la carga de la mercancía al salir sea asumida por el vendedor "EXW Loaded", gastos y riesgos, deberán para ello señalarlo claramente a través de una cláusula explícita que se añadirá en el contrato de venta (ej.: EXW París cargado, ICC 2020). Se asume que el vendedor presta al comprador, según el pedido de éste y asumiendo gastos y riesgos, toda la asistencia necesaria para la obtención de una licencia de exportación, un seguro, dándole toda la información útil a su disposición para permitirle al comprador exportar su mercancía de manera segura.

##### Variante

"EXW Loaded", o en español "en fábrica, cargado en". La revisión 2000 de los Incoterms introdujo este concepto de "EXW loaded", que reconoce una práctica muy utilizada: el vendedor es responsable de la carga de la mercancía en el vehículo del comprador.

Formulaciones EXW	Gastos	Riesgos
Embalaje	V	V
Antes de la entrega de la mercancía al transportista	C	C
Aduana exportación	C	C
Manutención transporte principal (carga)	C	C
Transporte principal	C	C
Seguro transporte	C	C
Manutención transporte principal (descarga)	C	C
Aduana importación	C	C
Al finalizar la ruta	C	C

*Ilustración 38 - Formulaciones EXW (Santander Trade Markets, 2020)*

### 8.1.2 FCA (*FREE CARRIER* / FRANCO-TRANSPORTISTA)

#### Vendedor

Si la entrega se efectúa en los locales del vendedor, es el vendedor quien efectúa la carga de la mercancía, embalada convenientemente en el vehículo proporcionado por el comprador (señálese “FCA locales del vendedor”). El pago de aranceles de exportación es responsabilidad del vendedor.

#### Comprador

El comprador escoge el medio de transporte y el transportista con quien desea efectuar el contrato de transporte (en su caso), y cancela el transporte principal. El traspaso de los gastos y riesgos ocurre cuando el transportista se hace cargo de la mercancía. Las partes deben convenir un lugar para la entrega de las mercancías (terminal del transportista o locales del vendedor). El vendedor debe, si fuese necesario, entregar la mercancía al comprador en tiempo útil, e incluso darle asistencia para obtener todo documento o información relacionados con la seguridad necesaria para la exportación y/o la importación de sus mercancías y/o para su transporte al destino final. Los documentos proporcionados y/o el apoyo proporcionado son asumidos costes y riesgos por el comprador.

Las reglas Incoterms 2020 permiten explícitamente que el transporte necesario sea realizado en virtud de un contrato de transporte o bien organizado con medios propios (sin implicación de un transportista que actúe como tercera parte).

### **Variante**

"FCA locales del vendedor": Este Incoterm se volvió oficial en la revisión 2000 de los Incoterms: el vendedor es responsable de cargar las mercancías.

### **Precisión geográfica**

Más aún que en los otros Incoterms, en FCA se precisará con cuidado el "punto convenido". FCA (*Le Havre*) no es suficiente si el exportador se encuentra en *Le Havre*. ¿Se trata de FCA (fábrica *Le Havre*) o FCA (bodega de agrupamiento del consignatario X *Le Havre*), o aun FCA (muelle N° X del puerto de *Le Havre*)? Si la entrega se realiza en otro lugar que los locales del vendedor, por ejemplo, una entrega en un terminal de transporte -de carreteras, ferroviario, aéreo, marítimo-, el vendedor encaminará la mercancía hasta ese terminal, pero no será responsable de la descarga del vehículo. La descarga incumbirá a quien reciba la mercancía en ese terminal de transporte. Prefiérase FCA a FOB, si el transporte se realiza en un contenedor o en un buque de carga de transbordo rodado.

### **Conocimiento con mención de embarque en las ventas FCA**

Cuando las mercancías se venden FCA, tanto vendedores como compradores (o su banco si se ha empleado una carta de crédito) pueden desear obtener un conocimiento con mención de embarque.

Sin embargo, la entrega en virtud del Incoterm FCA es efectiva antes de la carga de las mercancías a bordo de un barco. No resulta de ninguna manera seguro que el vendedor pueda obtener un conocimiento por parte del transportista marítimo. Según el contrato de transporte, este transportista solo está obligado y autorizado a producir un conocimiento cuando las mercancías se han cargado correctamente.



Para solucionar esta situación, los Incoterms 2020 han previsto una opción complementaria. El comprador y el vendedor pueden acordar que el comprador debe dar la instrucción a su transportista de proporcionar al vendedor el conocimiento que lleve la mención de embarque después de la carga de las mercancías, y el vendedor después está obligado a presentar este conocimiento al comprador, normalmente a través de bancos. No se le exigirá al vendedor ninguna obligación de cara al comprador en lo que concierne a términos del contrato de transporte.

Formulaciones FCA	Gastos	Riesgos
Embalaje	V	V
Antes de la entrega de la mercancía al transportista	V	V
Aduana exportación	V	V
Manutención transporte principal (carga)	C	C
Transporte principal	C	C
Seguro transporte	C	C
Manutención transporte principal (descarga)	C	C
Aduana importe	C	C
Al finalizar la ruta	C	C

*Ilustración 39 - Formulaciones FCA (Santander Trade Markets, 2020)*

### 8.1.3 FAS (*FREE ALONGSIDE SHIP* / FRANCO AL COSTADO DEL BUQUE)

#### Vendedor

Las obligaciones del vendedor han sido cumplidas cuando la mercancía, ya pagados los aranceles, es ubicada al costado del buque en el muelle o en alijadores del puerto de embarque convenido.

## **Comprador**

El comprador asumirá desde ese momento todos los gastos y riesgos de pérdida o deterioro, tan pronto la mercancía haya sido entregada junto al buque, en especial en caso de retraso de la embarcación o anulación de la escala. El comprador designa al transportista, concreta el contrato de transporte y paga el flete.

## **Obligación de lugar y de momento**

El vendedor sólo entrega FAS si entrega al costado del buque, cuando el buque está en el muelle. Es una obligación de lugar y momento (por ejemplo, de Marsella a Anvers, donde cada compañía propone al menos un viaje por semana, entregar ocho días antes de la fecha del buque escogida por el comprador es un acto prematuro).

## **Obtención de una licencia**

La obtención de una licencia de exportación o de otra autorización oficial corre por cuenta y riesgo del vendedor. Así ocurre también por parte del comprador en importación. Este último debe dar al vendedor toda la información sobre el nombre del buque, el lugar de carga y el momento de entrega escogido en el periodo acordado.

## **Gastos de documentación**

El vendedor debe, si fuese necesario y en un plazo apropiado, proporcionar al comprador (o ayudarlo para su obtención) todo documento o información en relación con la seguridad necesaria para la exportación y/o importación de sus mercancías y/o para su transporte al destino final. Los documentos proporcionados y/o la ayuda proporcionada corren por cuenta y riesgo del comprador.

Formulaciones FAS	Gastos	Riesgos
Embalaje	V	V
Antes de la entrega de la mercancía al transportista	V	V
Aduana exportación	V	V
Manutención transporte principal (carga)	C	C
Transporte principal	C	C
Seguro transporte	C	C
Manutención transporte principal (descarga)	C	C
Aduana importación	C	C
Al finalizar la ruta	C	C

*Ilustración 40 - Formulaciones FAS (Santander Trade Markets, 2020)*

#### **8.1.4 FOB (FREE ON BOARD / FRANCO A BORDO)**

##### **Vendedor**

Debe poner a disposición la mercancía, en el puerto de embarque designado, a bordo del buque escogido por el comprador, y efectuar los trámites de aduana para la exportación, si fuesen necesarias. En un contrato de tipo FOB, el vendedor cumple con su obligación de entrega cuando la mercancía está a bordo del buque en el puerto de embarque designado o, en el caso de las ventas sucesivas, el vendedor consigue las mercancías así entregadas para su transporte hasta su destino designado e indicado en el contrato de venta.

##### **Comprador**

Escoge el buque, paga el flete marítimo y el seguro, y se encarga de las formalidades a la llegada de la mercancía. Asume así todos los gastos y riesgos de pérdida o de deterioro que pueden ocurrirle a la mercancía desde el momento en que fue entregada.

## Variante

Nótese que la “puesta FOB” es la terminología utilizada por los transitarios para señalar que las operaciones anteriores al embarque fueron efectuadas, incluyendo, si fuese necesario, el pago de aranceles a la exportación. El conjunto de estas operaciones representa un coste que pagará el vendedor, llamado a veces “gastos de puesta en FOB”. El “FOB STOWED” y/o el “FOB STOWED and TRIMMED” (FOB estibado y equilibrado) es una variante. El vendedor asume la totalidad de los gastos de la mercancía en el puerto de embarque. Hay que precisar sin embargo en el contrato dónde se encuentra la transferencia de riesgos.

El vendedor debe, llegado el caso, proporcionar al comprador (o incluso ayudarlo a obtenerlos) todo documento o información relacionados con la seguridad necesaria para la exportación y/o la importación de sus mercancías y/o para su transporte a su destino final. Los documentos entregados y/o la ayuda prestada corren por cuenta y riesgo del comprador.

## Los FOB estadounidenses

Los FOB estadounidenses son diferentes. En Estados Unidos, el Incoterm FOB (*Free on Board*) no señala un envío en barco o un puerto: señala un destino en Estados Unidos, en la frontera. Para Estados Unidos, pueden existir principalmente cuatro tipos de FOB:

- ▶ FOB/Punto de partida: El comprador paga todo.
- ▶ FOB/Frontera: El fabricante paga hasta la frontera, sin pagar los aranceles de la mercancía.
- ▶ FOB/Punto de venta: La mercancía es enviada hasta un destino (una ciudad estadounidense) determinado. Es por ello que el proveedor paga los aranceles de la mercancía. Hay que señalar siempre el punto franco escogido, por lo general la ciudad.
- ▶ FOB/Destino con aranceles cancelados: En este caso, el fabricante se encarga de todo, sin la ayuda del comprador. Se dice también DDP/Entregado con los impuestos de aduana pagados. La mayoría de las ventas en Estados Unidos se efectúan bajo esta premisa.

Formulaciones FOB	Charges	Risks
Embalaje	V	V
Antes de la entrega de la mercancía al Aduana exportación	V	V
Aduana exportación	V	V
Manutención transporte principal (carga)	V	V
Transporte principal	C	C
Seguro transporte	C	C
Manutención transporte principal (descarga)	C	C
Aduana importe	C	C
Al finalizar la ruta	C	C

*Ilustración 41 - Formulaciones FOB (Santander Trade Markets, 2020)*

### 8.1.5 CFR (*COST AND FREIGHT / COSTE Y FLETE*)

#### **Vendedor**

Escoge al transportista, cierra y asume los gastos pagando el flete hasta el puerto de llegada convenido, descarga no incluida. La descarga de las mercancías con aranceles cancelados en el buque es su responsabilidad, al igual que las formalidades de despacho. Por otra parte, la transferencia de los riesgos es la misma que en FOB.

#### **Comprador**

Asume el riesgo de transporte cuando la mercancía ha sido entregada a bordo del buque en el puerto de embarque; debe recibirla al transportista y recibir la entrega de la mercancía en el puerto de destino convenido.

#### **Gastos de documentación**

El vendedor, corriendo con los gastos, debe entregar al comprador un documento de transporte válido hasta el puerto de destino convenido cubriendo las mercancías contractuales, para hacer valer sus derechos (ej.: reclamo de la mercancía al transportista, venta de la mercancía en tránsito, etc.). Además, debe darle todas las informaciones

necesarias para tomar las medidas idóneas para recibir las mercancías. Las informaciones y documentos relacionados con la seguridad que necesita el comprador para la exportación y/o la importación y/o el transporte hasta el destino final deben ser entregados por el vendedor al comprador, habiéndolos solicitado y asumiendo los gastos y riesgos.

Formulaciones CFR	Gastos	Riesgos
Embalaje	V	V
Antes de la entrega de la mercancía al transportista	V	V
Aduana exportación	V	V
Manutención transporte principal (carga)	V	V
Transporte principal	V	C
Seguro transporte	C	C
Manutención transporte principal (descarga)	C	C
Aduana importación	C	C
Al finalizar la ruta	C	C

*Ilustración 42 - Formulaciones CFR (Santander Trade Markets, 2020)*

### 8.1.6 CIF (*COST INSURANCE AND FREIGHT / COSTE, SEGURO Y FLETE*)

#### Vendedor

Término idéntico al CFR, con la obligación adicional para el vendedor de proporcionar un seguro marítimo contra el riesgo de pérdida o de daños de las mercancías. El vendedor paga la prima del seguro. En virtud de la regla CIF de los Incoterms 2020, el vendedor está obligado a obtener una cobertura de seguro limitada de conformidad con las Cláusulas C de las *Institute Cargo Clauses* o de cualquier otra serie de cláusulas similares. Las partes son sin embargo libres para acordar un nivel de cobertura superior.

## Comprador

Asume el riesgo de transporte cuando la mercancía ha sido entregada a bordo del buque en el puerto de embarque. Verificar y recibir la mercancía del transportista en el puerto de destino convenido.

Los compradores aprecian este Incoterm, porque se liberan de las formalidades logísticas.

## Gastos de documentación

Las informaciones y documentos relacionados con la seguridad que requiere el comprador para la exportación y/o la importación y/o el transporte hasta el destino final deben ser entregados por el vendedor al comprador, habiéndolos solicitado, y asume los gastos y riesgos.

Formulaciones CIF	Charges	Risks
Embalaje	V	V
Antes de la entrega de la mercancía al transportista	V	V
Aduana exportación	V	V
Manutención transporte principal (carga)	V	V/C
Transporte principal	V	C
Seguro transporte	V	C
Manutención transporte principal (descarga)	C	C
Aduana importación	C	C
Al finalizar la ruta	C	C

*Ilustración 43 - Formulaciones CIF (Santander Trade Markets, 2020)*

### 8.1.7 CPT (CARRIAGE PAID TO / TRANSPORTE PAGADO HASTA)

#### Vendedor

El vendedor coordina la cadena logística. Después de haberse encargado de los aranceles de exportación, escoge los transportistas y paga los gastos hasta el lugar convenido.

## **Comprador**

Los riesgos de avería o pérdida son asumidos por el comprador desde el momento en que las mercancías han sido entregadas al primer transportista. Luego, el comprador se encarga de pagar los aranceles de importación y los gastos de descarga.

## **Los gastos de descarga**

Es importante esclarecer correctamente la noción de gastos de descarga en el contexto del contrato de transporte. El comprador normalmente debe asumir los gastos de descarga, salvo si éstos son incorporados en el precio del transporte. En ese caso, son asumidos por el vendedor. Por ello, el vendedor debe clarificar con cuidado este punto con su comprador, para evitar encontrarse en una situación en que el destinatario rehúse pagar y el transportista contacte a quien le ha encargado el trabajo (el vendedor) y le exija el pago de los gastos de descarga y los eventuales gastos de estancamiento del vehículo en espera de una resolución del problema.

## **Precisiones geográficas**

En la regla CPT hay transferencia de riesgos y gastos en lugares distintos. Se recomienda entonces que las partes indiquen con precisión, en su contrato, tanto el lugar de entrega en que el riesgo pasa al comprador como el lugar de destino convenido donde el vendedor concluirá un contrato de transporte.

## **Gastos de documentación**

La información y documentos requeridos por el comprador para la seguridad en la exportación y/o la importación y/o el transporte hasta el destino final deben ser proporcionados por el vendedor al comprador, tras haberlos solicitado, y asumiendo gastos y riesgos.



Formulaciones CPT	Gastos	Riesgos
Embalaje	V	V
Antes de la entrega de la mercancía al transportista	V	V
Aduana exportación	V	V
Manutención transporte principal (carga)	V	V
Transporte principal	V	C
Seguro transporte	C	C
Manutención transporte principal (descarga)	C	C
Aduana importación	C	C
Al finalizar la ruta	C	C

*Ilustración 44 - Formulaciones CPT (Santander Trade Markets, 2020)*

### **8.1.8 CIP (CARRIAGE AND INSURANCE PAID TO / TRANSPORTE Y SEGURO PAGADOS HASTA)**

#### **Vendedor**

CIP es idéntico al CPT, pero el vendedor debe proporcionar además un seguro de transporte. El vendedor cierra el contrato de transporte, paga el flete y la prima de seguro. En virtud de la regla CIP de los Incoterms 2020, el vendedor está obligado a obtener una cobertura de seguro limitada de conformidad con las Cláusulas A de las *Institute Cargo Clauses* o de cualquier otra serie de cláusulas similares. Las partes son sin embargo libres para acordar un nivel de cobertura inferior.

#### **Comprador**

Los riesgos de daños o pérdida son asumidos por el comprador desde el momento en que las mercancías han sido entregadas al primer transportista. Luego, el comprador se encarga del pago de aranceles e impuestos de importación y los gastos de descarga.

## Cobertura del seguro

Según el término CIP, el vendedor no está obligado a contratar el seguro, salvo con una cobertura mínima. Si el comprador desea estar protegido con una cobertura más amplia, deberá en esas condiciones obtener la autorización del vendedor, o contratar él mismo un seguro complementario.

## Gastos de documentación

Las informaciones y documentos que corresponden a la seguridad que requiere el comprador para la exportación y/o la importación y/o el transporte hasta el destino final deben ser proporcionados por el vendedor al comprador, habiéndolos solicitado y asumiendo costes y riesgos.

Formulaciones CIP	Gastos	Riesgos
Embalaje	V	V
Antes de la entrega de la mercancía al transportista	V	V
Aduana exportación	V	V
Manutención transporte principal (carga)	V	V
Transporte principal	V	C
Seguro transporte	V	C
Manutención transporte principal (descarga)	C	C
Aduana importación	C	C
Al finalizar la ruta	C	C

*Ilustración 45 - Formulaciones CIP (Santander Trade Markets, 2020)*

### 8.1.9 DAP (*DELIVERED AT PLACE* / ENTREGADO EN PUNTO DE DESTINO)

#### Vendedor

Debe entregar las mercancías, poniéndolas a disposición del comprador en el medio de transporte de aproximación listo para la descarga en el destino convenido. Debe pagar los aranceles de exportación, pero en cambio no tiene ninguna obligación de hacer lo mismo

con los aranceles. El vendedor debe tomar a su cargo el transporte de las mercancías al destino convenido, así como descargarlas del medio de transporte al llegar. Por lo demás, el vendedor no está obligado, ante el comprador, a cerrar un contrato de seguro. Asimismo, debe proporcionar al comprador, por cuenta propia, el documento que le permite recibir las mercancías. Las reglas Incoterms 2020 permiten explícitamente que el transporte necesario sea realizado en virtud de un contrato de transporte o bien organizado con medios propios (sin implicación de un transportista que actúe como tercera parte).

### Comprador

Debe pagar el precio de las mercancías, como previsto en el contrato de venta, y recibir estas mercancías tan pronto hayan sido entregadas.

### Seguridad

El comprador debe además comunicar al vendedor que es necesario proporcionarle toda la información relacionada con la seguridad que pueda necesitar para la exportación, la importación, el transporte de las mercancías a su destino final. Esta nueva regla reemplaza el DDU. Se aconseja utilizarla únicamente en los países en que los medios de transporte al destino son manejables.

Formulaciones DAP	Gastos	Riesgos
Embalaje	V	V
Antes de la entrega de la mercancía al transportista	V	V
Aduana exportación	V	V
Manutención transporte principal (carga)	V	V
Transporte principal	V	V
Seguro transporte	V	V
Manutención transporte principal (descarga)	V	V
Aduana importación	C	C
Al finalizar la ruta	V	V

*Ilustración 46 - Formulaciones DAP (Santander Trade Markets, 2020)*

### **8.1.10 DPU (*DELIVERED AT PLACE UNLOADED* / ENTREGADO Y DESCARGADO EN LUGAR ACORDADO)**

El Incoterm DPU sustituye al DAT (*Delivered At Terminal* - Entregado en terminal) para señalar el hecho de que el lugar de destino no es forzosamente un «terminal». Sin embargo, si el lugar de destino no es un terminal, el vendedor debe asegurarse de que es completamente capaz de descargar las mercancías en el lugar donde ha previsto entregarlas.

#### **Vendedor**

Debe entregar las mercancías poniéndolas a disposición del comprador en el lugar acordado en el destino, fecha o plazos acordados. El vendedor debe celebrar un contrato cuyos gastos correrán a su cargo para el transporte de mercancías hasta este lugar y descargar las mercancías del medio de transporte que llegue. El vendedor no tiene la obligación de cara al comprador de celebrar un contrato de seguro. Debe sin embargo proporcionar al comprador, corriendo con sus gastos, el documento que le permita aceptar la entrega de las mercancías. El Incoterm DPU obliga al vendedor a despachar en la aduana las mercancías en la exportación. No obstante, no existe ninguna obligación de hacer el despacho aduanero en la importación.

Las reglas Incoterms 2020 permiten explícitamente que el transporte necesario sea realizado en virtud de un contrato de transporte o bien organizado con medios propios (sin implicación de un transportista que actúe como tercera parte).

#### **Comprador**

Debe aceptar la entrega de las mercancías desde que se hayan enviado y pagar su precio como se haya previsto en el contrato de venta. El comprador debe además avisar al vendedor de la necesidad de proporcionarle toda la información relativa a la seguridad que pudiera necesitar para la exportación, la importación, el transporte de las mercancías a su destino final. Esta regla Incoterms se ha creado específicamente para el transporte que se realiza mediante contenedores. También está adaptada al transporte marítimo convencional cuando

el vendedor quiere conservar los riesgos de la descarga del barco en el puerto de destino. Conviene en este caso especificar el lugar de entrega (muelle, punto de amarre...).

Formulaciones DPU	Gastos	Riesgos
Embalaje	V	V
Antes de la entrega de la mercancía al transportista	V	V
Douane export	V	V
Manutention transport principal (chargement)	V	V
Transporte principal	V	V
Seguro transporte	V*	V*
Manutención transporte principal (descarga)	V	V
Aduana importación	C	C
Después del transporte	C	C

\* no obligatorio

*Ilustración 47 - Formulaciones DPU (Santander Trade Markets, 2020)*

### 8.1.11 DDP (*DELIVERED DUTY PAID* / ENTREGADO CON DERECHOS PAGADOS)

#### Vendedor

El vendedor, en este caso, tiene las máximas obligaciones: la transferencia de los gastos y riesgos se efectúa en el momento de la entrega donde el comprador. El pago de derechos de aduana de importación también le incumbe.

Las reglas Incoterms 2020 permiten explícitamente que el transporte necesario sea realizado en virtud de un contrato de transporte o bien organizado con medios propios (sin implicación de un transportista que actúe como tercera parte).

## Comprador

Recibir en el lugar de destino convenido y pagar los gastos de descarga. Además, debe comunicar al vendedor la necesidad de proporcionarle toda la información relacionada con la seguridad que fuese necesaria para la exportación, la importación, el transporte de las mercancías a su destino final.

## DDP versus EXW

El término DDP es exactamente lo contrario de EXW.

## Los gastos relativos a la importación de mercancías

Si las partes desean excluir de las obligaciones del vendedor el pago de ciertos cargos ligados a la importación de la mercancía, se especifica: por ejemplo "*Delivered Duty Paid*", IVA no pagado (DDP, *VAT unpaid*).

Formulaciones DDP	Gastos	Riesgos
Embalaje	V	V
Antes de la entrega de la mercancía al transportista	V	V
Aduana exportación	V	V
Manutención transporte principal (carga)	V	V
Transporte principal	V	V
Seguro transporte	V	V
Manutención transporte principal (descarga)	V	V
Aduana importación	V	V
Al finalizar la ruta	V	V

*Ilustración 48 - Formulaciones DDP (Santander Trade Markets, 2020)*

## 8.2 DIAGRAMA DE GANTT

### MODELO DE AHORRO DE COSTES PARA EL COMERCIO INTERNACIONAL

EY  
 Jose Antonio Leiva

Fecha de inicio:

**Legenda:**

☎ Reunión con Product Owner 🗣 Reunión con equipo EY 📅 Reunión con tutor 📖 Formación 🏢 Reunión con ICAI ★ Final del proyecto

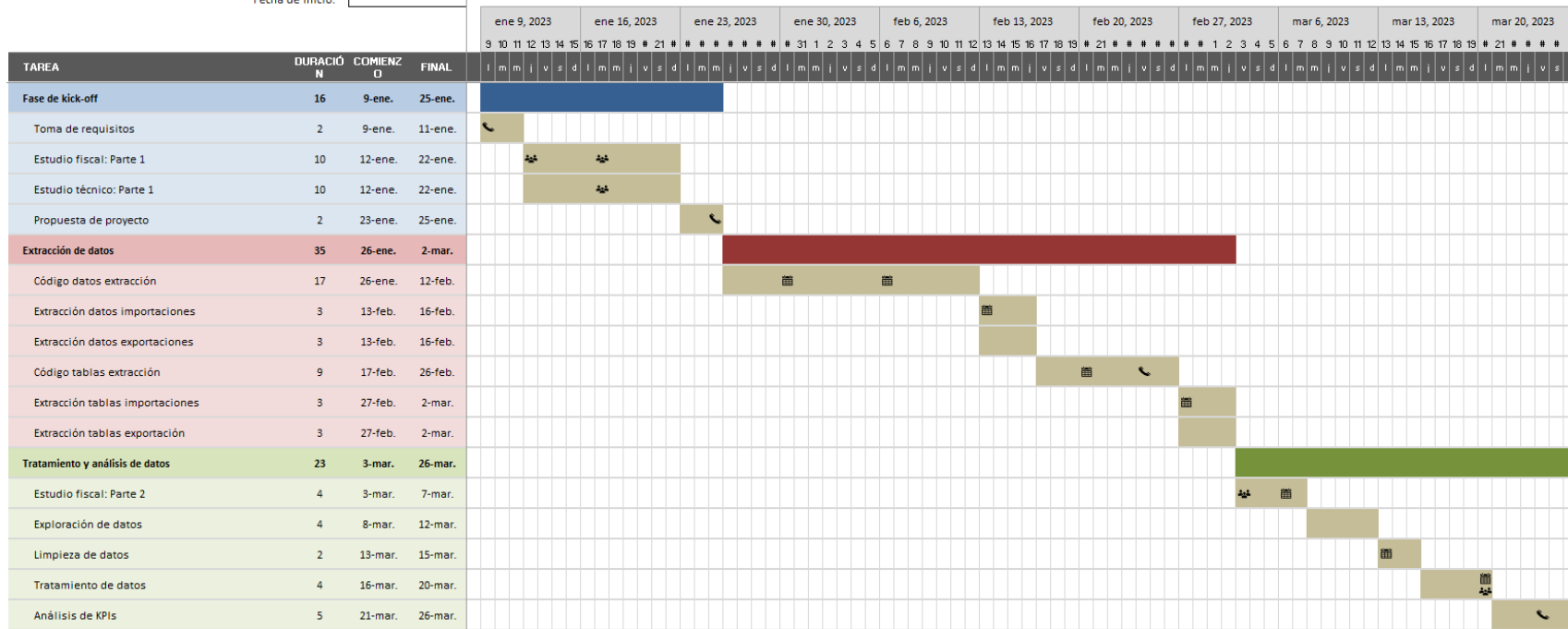
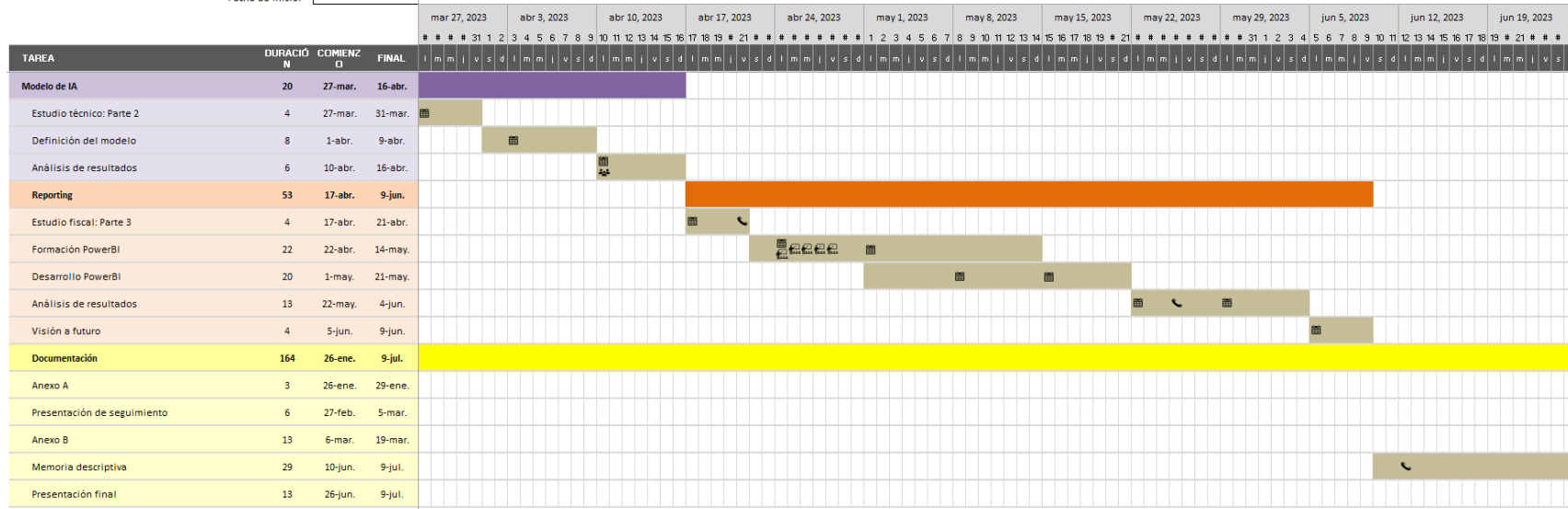


Ilustración 49 - Diagrama de Gantt Pt. 1

**MODELO DE AHORRO DE COSTES PARA EL COMER**

EY  
 Jose Antonio Leiva

Fecha de inicio:



*Ilustración 50 - Diagrama de Gantt Pt. 2*



### 8.3 CÓDIGO PYTHON EXTRACCIÓN DUA

```
# -----IMPORTS-----
from selenium import webdriver # Import SeleniumT
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.service import Service
# from webdriver_manager.chrome import ChromeDriverManager

# Imports for excel files
import xlswriter
from xlrld import open_workbook
from openpyxl import load_workbook
from openpyxl.utils import get_column_letter

# Other imports
import time
import os
from math import floor, log10
import shutil
import pandas as pd
import numpy as np

# -----MAIN-----
if __name__ == '__main__':

    curr_time = time.strftime("%H:%M:%S", time.localtime())
    print("\n\nHora de inicio: ", curr_time)
    # Set WebDriver options
    chromeOptions = webdriver.ChromeOptions()

    chromeOptions.add_experimental_option('useAutomationExtension',
False)
    chromeOptions.add_argument('--ignore-certificate-errors')
    chromeOptions.add_experimental_option('excludeSwitches',
['enable-logging'])
    chromeOptions.add_experimental_option('prefs', {
        "download.default_directory": "C:\\Users\\RJ227FZ\\OneDrive -
EY\\Desktop\\ICAI\\TFM",
        # Change default directory for downloads
        "download.prompt_for_download": False, # To auto download
the file
        "download.directory_upgrade": True,
        "plugins.always_open_pdf_externally": True # It will not
show PDF directly in chrome
    })

    # driver = webdriver.Chrome(options=chromeOptions,
executable_path='C:\\chromedriver_win32\\chromedriver.exe',
desired_capabilities=chromeOptions.to_capabilities())
```

```
driver =
webdriver.Chrome(service=Service(r"C:\chromedriver_win32\chromedriver
.exe"), options=chromeOptions)

# Go to AEAT website

driver.get("https://ww1.agenciatributaria.gob.es/wlpl/inwinvoc/es.ae
at.dit.adu.adip.inter.qry.QImporInternet")

# Date ranges taken from excel file 'Fechas.xlsx'
FechaDesde = [] # Max 2 months interval
FechasHasta = []

# Import date ranges
# Para borrar valores hay que eliminar las filas no vale con
borrar el contenido (otherwise coge los blancos)
wb = load_workbook('Fechas.xlsx')
ws = wb.worksheets

for sheet in ws:
    number_of_rows = sheet.max_row
    for row in range(2, number_of_rows + 1):
        value = (sheet.cell(row, 1).value)
        value2 = (sheet.cell(row, 2).value)
        FechaDesde.append(value)
        FechasHasta.append(value2)

wb.close()

# List to gather all DUAs
DUAS = []

# Get DUA number
#operador =
driver.find_element(By.XPATH,"//*[@class='cua_cols']/input")
operador =
driver.find_element(By.XPATH,"//*[@class='cua_cols']/big/strong")
#path = 'DUA ' + operador.get_attribute('value')
path = 'DUA ' + operador.text

# Delete existing excel file
try:
    os.remove(path + '.xls')
except:
    print("Archivo excel eliminado")

# Make the search and download excel file
for i in range(0, len(FechaDesde)):
    # Find date input fields
    inputElementFechaDesde =
driver.find_element(By.XPATH,"//*[@id='wFecOpeDDia']")
```

```
inputElementFechaHasta =
driver.find_element(By.XPATH, "//*[@id='wFecOpeHDia']")

# Find importador input field and check it
inputElementImportador =
driver.find_element(By.XPATH, "//*[@id='wTipoBusqueda0']").click()

# Clear the date inputs
inputElementFechaDesde.clear()
inputElementFechaHasta.clear()

# Clear the DUA value input to avoid previous cache values
duaField1 = driver.find_element(By.ID, "wAnio").clear()
duaField2 = driver.find_element(By.ID, "wPais").clear()
duaField3 =
driver.find_element(By.ID, "CABECERA_recinto").clear()
duaField4 = driver.find_element(By.ID, "wPrefijo").clear()
duaField5 = driver.find_element(By.ID, "wNum").clear()
duaField6 = driver.find_element(By.ID, "wDigCtrl").clear()

# Enter a date range
inputElementFechaDesde.send_keys(FechaDesde[i])
inputElementFechaHasta.send_keys(FechasHasta[i])

# Click enter and wait
driver.find_element(By.XPATH, "//*[@id='Query']/div/input[1]").click()
driver.implicitly_wait(2)

# Click to download Excel file
driver.find_element(By.XPATH, "//*[@class='aplicacionAcciones
lista_espaciada']/li/a[1]").click()
print("Descargando ...")

# Comprobar que ha terminado de descargarse el archivo
while not os.path.exists(path + '.xls'):
    time.sleep(1)

# Open downloaded excel file to take all DUAs
wb_downloaded = open_workbook(path + '.xls')
number_of_rows_duas = wb_downloaded.sheets()[0].nrows

for row in range(1, number_of_rows_duas):
    DUAS.append(wb_downloaded.sheets()[0].cell(row, 0).value)

# Remove excel file to avoid overwriting
os.remove(path + '.xls')

# [BORRAR] Me quedo solo con 5 DUAs para agilizar el proceso
lenDUAS = len(DUAS)

#for i in range(lenDUAS - 5):
```

```
# del DUAS[0]

for f in os.listdir('Requerimientos'):
    split_tup = os.path.splitext(f)
    if split_tup[1] == '.pdf':
        os.remove('Requerimientos/' + f)

wb = load_workbook('DUAS Importacion 25-04-2023.xlsx')
ws = wb.worksheets
DUAS = []

for sheet in ws:
    number_of_rows = sheet.max_row
    for row in range(2, number_of_rows + 1):
        value = sheet.cell(row, 2).value
        DUAS.append(value)

# DUAS=['20ES00280133143300']
# DUAS = ['20ES00280131236828']
# DUAS = ['20ES00280130362617', '20ES00280131236828']

# Array to get the max number of headers for all DUAs (and in the
right order)
arrayHeaders = []

# Array of info of each DUA
arrayInfo = []

# DataFrame with all the data
df = pd.DataFrame()

# Collection of tables (collection of dataframes)
tables = {}

# Indice para ir almacenando las tablas de las distintas
secciones y las distintas duas
globalIndex = 0

# Indice para ir almacenando la info de distitnas duas
globalIndexDUA = 0

indexSumario0_prev = 0
indexSumario1_prev = 0
indexSumario2_prev = 0
indexSumario3_prev = 0
indexDUA0_prev = 0
indexDUA1_prev = 0
indexPartida0_prev = 0
indexPartida1_prev = 0
indexPartida2_prev = 0
indexPartida3_prev = 0
indexPartida4_prev = 0
```

```
indexPartida5_prev = 0

indexSumario0_after = 0
indexSumario1_after = 0
indexSumario2_after = 0
indexSumario3_after = 0
indexDUA0_after = 0
indexDUA1_after = 0
indexPartida0_after = 0
indexPartida1_after = 0
indexPartida2_after = 0
indexPartida3_after = 0
indexPartida4_after = 0
indexPartida5_after = 0

writer = pd.ExcelWriter('Tablas DUAS Importacion 04-05-
2023.xlsx', mode='a', engine='openpyxl',
                      if_sheet_exists="overlay")

# Iterate all DUAS
for dua in DUAS:
    """if DUAS.index(dua) != 0:
        df.loc[len(df)] = np.zeros(len(df.columns), dtype=int)"""
    # Looking for DUA details (individual web page for each DUA)
    # Get DUA details from AEAT

    url =
    "https://ww1.agenciatributaria.gob.es/wlpl/inwinvoc/es.aeat.dit.adu.
    adip.inter.cnt.CImporInternet?retorno=es.aeat.dit.adu.adip.inter.gry.
    QImporInternet&operacion=2250&CABECERA_MRN=" + dua.replace(
        " ", "")
    driver.get(url)
    # Click on See All for each section
    driver.find_element(By.CSS_SELECTOR, "a[href='" + url +
    "#verTodoSumario']").click()
    driver.find_element(By.CSS_SELECTOR, "a[href='" + url +
    "#verTodoCabecera']").click()

    # Get Number of Partidas
    numero_partidas = driver.find_element(By.XPATH,
    "//*[@id='verTodoSumario']/ul[2]/li[2]/span").text.strip()
    # Expand each Partida section
    for i in range(1, int(numero_partidas) + 1):
        text = str(0) * (5 - (floor(log10(i)) + 1)) + str(i)
        try:
            driver.find_element(By.CSS_SELECTOR, "a[href='" + url
    + "#verTodoPartida" + text + "'").click()
        except:
            print("'ver Todo' button not found in Partida " +
    str(i))

    # Download existing requirements if so
```

```
numbReq = 0
try:
    # Requerimientos download pdfs
    driver.find_element(By.XPATH,
        "//*[ @id='requerimientos' ]")
    requerimientos = driver.find_elements(By.XPATH,
        "//*[contains(@href,
        'es.aeat.dit.adu.ites.escritos.EscritosVerPDF')]")
    for requerimiento in requerimientos:
        try:
            if requerimiento.text != '':
                numbReq += 1
                requerimiento.click()
                driver.implicitly_wait(2)
                time.sleep(2)
                print("Descargando pdf...")
                name = str(dua).replace(" ", "") + '
Requerimiento ' + str(numbReq) + '.pdf'
os.rename('es.aeat.dit.adu.ites.escritos.pdf', name)
                driver.implicitly_wait(1)
                time.sleep(1)
                shutil.move(name, 'Requerimientos/' + name)
            try:
os.remove('es.aeat.dit.adu.ites.escritos.pdf')
            except:
                print("Se ha hecho bien el rename")
        except:
            print("No hay link")
    except:
        print("No hay requerimientos")

# SUMARIO
resultSetSumario = driver.find_element(By.XPATH,
    "//*[ @id='verTodoSumario' ]")
headerSumario = resultSetSumario.find_elements(By.CLASS_NAME,
    "col_ancha")

if (DUAS.index(dua) < 1):
    df["Partida ID"] = ['']
    df["Requerimientos:" ] = ['']

for i in range(len(headerSumario)):
    # Avoid including requerimientos headers
    parent = headerSumario[i].find_element(By.XPATH, "..")
    parent = parent.find_element(By.XPATH, "..")
    parent = parent.find_element(By.XPATH, "..")
    if parent.tag_name != "fieldset" or
headerSumario[i].find_element(By.TAG_NAME,
```

```

"strong").text == "      Requerimientos:": # me aseguro de coger el
header "requerimiento" pero evitar los numeros
      header = headerSumario[i].find_element(By.TAG_NAME,
"strong").text.strip()
      info = headerSumario[i].find_elements(By.CLASS_NAME,
"notraducir")
      # Rellenar dataframe si es la primera dua que leo
      if DUAS.index(dua) == 0:
          df[header] = ['']
      # Rellenar numero de requerimientos
      if header == "Requerimientos:":
          df[header][globalIndexDUA] = numbReq
      # Rellenar headers e informacion de manera simultanea
      if header in df.columns:
          if header != "Requerimientos:":
              df[header][globalIndexDUA] =
info[0].text.strip()
          else:
              df[header] = ''
              df[header][globalIndexDUA] = info[0].text.strip()

      else:
          continue

      # DUA
      resultSetDUA = driver.find_element(By.XPATH,
"//*[ @id='verTodoCabecera']")
      headerDUA = resultSetDUA.find_elements(By.CLASS_NAME,
"col_ancha")

      for i in range(len(headerDUA)):
          header = headerDUA[i].find_element(By.TAG_NAME,
"strong").text.strip()
          info = headerDUA[i].find_elements(By.CLASS_NAME,
"notraducir")
          # Rellenar headers e informacion de manera simultanea
          if header in df.columns:
              df[header][globalIndexDUA] = info[0].text.strip()
          else:
              df[header] = ''
              df[header][globalIndexDUA] = info[0].text.strip()

      # PARTIDAS
      # Relleno primero la info general para todas las partidas
      for i in range(2, int(numero_partidas) + 1):
          """df.loc[i - 1 + globalIndexDUA] =
df.loc[globalIndexDUA]"""
          for j in range(1, int(numero_partidas) + 1):
              text = str(0) * (5 - (floor(log10(j)) + 1)) + str(j)
              # Partidas con botón de Ver Todo
          try:

```

```

        resultSetPartida = driver.find_element(By.XPATH,
        "//*[@id='verTodoPartida" + text + "']")
        df["Partida ID"][j - 1 + globalIndexDUA] = j
        headerPartida =
resultSetPartida.find_elements(By.CLASS_NAME, "col_ancha")
        for i in range(len(headerPartida)):
            header =
headerPartida[i].find_element(By.TAG_NAME, "strong").text.strip()
            info =
headerPartida[i].find_elements(By.CLASS_NAME, "notraducir")
                if header in df.columns:
                    df[header][j - 1 + globalIndexDUA] =
info[0].text.strip()
                else:
                    df[header] = ''
                    df[header][j - 1 + globalIndexDUA] =
info[0].text.strip()
            # TABLAS Partidas con boton
            resultSetTables =
resultSetPartida.find_elements(By.TAG_NAME, "table")
            for table in resultSetTables:
                parent = table.find_element(By.XPATH, "..")
                parent = parent.find_element(By.XPATH, "..")
                parent2 = parent.find_element(By.XPATH, "..")
                # me aseguro de coger solo las tablas visibles en
la web
                if parent.get_attribute("class") == 'conborde '
or parent2.get_attribute("class") == 'conborde ':
                    dfTable = pd.DataFrame()
                    headers = table.find_elements(By.TAG_NAME,
"th")

                    for header in headers:
                        if header.text != ' ':
                            dfTable[header.text.strip()] = ['']
                    rows = table.find_elements(By.TAG_NAME, "tr")
                    for row in rows:
                        info = []
                        if row.find_element(By.XPATH,
"..").tag_name == "tbody":
                            allData =
row.find_elements(By.CLASS_NAME, "notraducir")
                            for singleData in allData:
                                info.append(singleData.text.strip())
                                if rows.index(row) > 1:
                                    info.append(dua)
                                    info.append('Partida ' + str(j))
                                    dfTable.loc[rows.index(row) - 1]
= info
                                else:
                                    dfTable.loc[rows.index(row) - 1]
= info

```



```

dfTable.loc[rows.index(row) - 1,
'DUA_COL'] = dua
dfTable.loc[rows.index(row) - 1,
'Partida_COL'] = 'Partida ' + str(j)

    if resultSetTables.index(table) == 0:
        indexPartida0_after += 1
    elif resultSetTables.index(table) ==
1:
        indexPartida1_after += 1
    elif resultSetTables.index(table) ==
2:
        indexPartida2_after += 1
    elif resultSetTables.index(table) ==
3:
        indexPartida3_after += 1
    elif resultSetTables.index(table) ==
4:
        indexPartida4_after += 1
    elif resultSetTables.index(table) ==
5:
        indexPartida5_after += 1

        sheetName = "Partida" +
str(resultSetTables.index(table))

        if DUAS.index(dua) == 0 and j == 1:
            if resultSetTables.index(table) == 0:
                dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida0_prev, index=False)
                indexPartida0_after += 1
            elif resultSetTables.index(table) == 1:
                dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida1_prev, index=False)
                indexPartida1_after += 1
            elif resultSetTables.index(table) == 2:
                dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida2_prev, index=False)
                indexPartida2_after += 1
            elif resultSetTables.index(table) == 3:
                dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida3_prev, index=False)
                indexPartida3_after += 1
            elif resultSetTables.index(table) == 4:
                dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida4_prev, index=False)
                indexPartida4_after += 1
            elif resultSetTables.index(table) == 5:
                dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida5_prev, index=False)
                indexPartida5_after += 1
        else:

```

```

        if resultSetTables.index(table) == 0:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida0_prev,
                                header=False,
index=False)
            elif resultSetTables.index(table) == 1:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida1_prev,
                                header=False,
index=False)
            elif resultSetTables.index(table) == 2:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida2_prev,
                                header=False,
index=False)
            elif resultSetTables.index(table) == 3:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida3_prev,
                                header=False,
index=False)
            elif resultSetTables.index(table) == 4:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida4_prev,
                                header=False,
index=False)
            elif resultSetTables.index(table) == 5:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida5_prev,
                                header=False,
index=False)

            indexPartida0_prev = indexPartida0_after
            indexPartida1_prev = indexPartida1_after
            indexPartida2_prev = indexPartida2_after
            indexPartida3_prev = indexPartida3_after
            indexPartida4_prev = indexPartida4_after
            indexPartida5_prev = indexPartida5_after

            tables[resultSetTables.index(table) +
globalIndex + 1] = dfTable
            globalIndex += resultSetTables.index(table) + 1

            # Partidas sin botón de Ver Todo
            except:
                resultSetPartidaBis = driver.find_elements(By.XPATH,
"//*[@class='conborde bloque_100']")
                for partida in resultSetPartidaBis:
                    headerPartidaBis =
resultSetPartidaBis[resultSetPartidaBis.index(partida)].find_elements
(
                    By.CLASS_NAME, "col_ancha")
                    df["Partida ID"][

```

```

                                globalIndexDUA + j +
resultSetPartidaBis.index(partida) - 1] = j +
resultSetPartidaBis.index(
    partida)
    for i in range(len(headerPartidaBis)):
        header =
headerPartidaBis[i].find_element(By.TAG_NAME, "strong").text.strip()
        info =
headerPartidaBis[i].find_elements(By.CLASS_NAME, "notraducir")
        if header in df.columns:
            df[header][j +
resultSetPartidaBis.index(partida) - 1 + globalIndexDUA] = info[
                0].text.strip()
        else:
            df[header] = ''
            df[header][j +
resultSetPartidaBis.index(partida) - 1 + globalIndexDUA] = info[
                0].text.strip()
        # TABLAS Partidas sin boton
resultSetTables =
partida.find_elements(By.TAG_NAME, "table")
    for table in resultSetTables:
        parent = table.find_element(By.XPATH, "..")
        parent = parent.find_element(By.XPATH, "..")
        parent2 = parent.find_element(By.XPATH, "..")
        # me aseguro de coger solo las tablas
visibles en la web
        if parent.get_attribute("class") == 'conborde
bloque_100' or parent2.get_attribute(
            "class") == 'conborde bloque_100':
            dfTable = pd.DataFrame()
            headers =
table.find_elements(By.TAG_NAME, "th")
            for header in headers:
                if header.text != ' ':
                    dfTable[header.text.strip()] =
['']
            rows = table.find_elements(By.TAG_NAME,
"tr")
            for row in rows:
                info = []
                if row.find_element(By.XPATH,
"..").tag_name == "tbody":
                    allData =
row.find_elements(By.CLASS_NAME, "notraducir")
                    for singleData in allData:
info.append(singleData.text.strip())
                if rows.index(row) > 1:
                    info.append(dua)
                    info.append('Partida ' +
str(j + resultSetPartidaBis.index(partida)))

```

```

dfTable.loc[rows.index(row) -
1] = info
else:
dfTable.loc[rows.index(row) -
1] = info
dfTable.loc[rows.index(row) -
1, 'DUA_COL'] = dua
dfTable.loc[rows.index(row) -
1, 'Partida_COL'] = 'Partida ' + str(
j +
resultSetPartidaBis.index(partida))

if resultSetTables.index(table)
== 0:
indexPartida0_after += 1
elif resultSetTables.index(table)
== 1:
indexPartida1_after += 1
elif resultSetTables.index(table)
== 2:
indexPartida2_after += 1
elif resultSetTables.index(table)
== 3:
indexPartida3_after += 1
elif resultSetTables.index(table)
== 4:
indexPartida4_after += 1
elif resultSetTables.index(table)
== 5:
indexPartida5_after += 1

sheetName = "Partida" +
str(resultSetTables.index(table))

if DUAS.index(dua) == 0 and j == 1:
if resultSetTables.index(table) == 0:
dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida0_prev,
index=False)
indexPartida0_after += 1
elif resultSetTables.index(table) ==
1:
dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida1_prev,
index=False)
indexPartida1_after += 1
elif resultSetTables.index(table) ==
2:
dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida2_prev,
index=False)
indexPartida2_after += 1

```

```
elif resultSetTables.index(table) ==
3:
    dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida3_prev,
                        index=False)
    indexPartida3_after += 1
elif resultSetTables.index(table) ==
4:
    dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida4_prev,
                        index=False)
    indexPartida4_after += 1
elif resultSetTables.index(table) ==
5:
    dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida5_prev,
                        index=False)
    indexPartida5_after += 1
else:
    if resultSetTables.index(table) == 0:
sheet_name=sheetName, startrow=indexPartida0_prev,
                        header=False,
index=False)
    elif resultSetTables.index(table) ==
1:
    dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida1_prev,
                        header=False,
index=False)
    elif resultSetTables.index(table) ==
2:
    dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida2_prev,
                        header=False,
index=False)
    elif resultSetTables.index(table) ==
3:
    dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida3_prev,
                        header=False,
index=False)
    elif resultSetTables.index(table) ==
4:
    dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida4_prev,
                        header=False,
index=False)
    elif resultSetTables.index(table) ==
5:
    dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexPartida5_prev,
```

```

header=False,
index=False)

        indexPartida0_prev = indexPartida0_after
        indexPartida1_prev = indexPartida1_after
        indexPartida2_prev = indexPartida2_after
        indexPartida3_prev = indexPartida3_after
        indexPartida4_prev = indexPartida4_after
        indexPartida5_prev = indexPartida5_after

        tables[resultSetTables.index(table) +
globalIndex + 1] = dfTable
        globalIndex += resultSetTables.index(table) + 1
        break
    globalIndexDUA += int(numero_partidas)

    # TABLAS Sumario
    resultSetTables = resultSetSumario.find_elements(By.TAG_NAME,
"table")
    for table in resultSetTables:
        parent = table.find_element(By.XPATH, "..")
        parent = parent.find_element(By.XPATH, "..")
        parent2 = parent.find_element(By.XPATH, "..")
        # me aseguro de coger solo las tablas visibles en la web
        if parent.get_attribute("class") == 'conborde ' or
parent2.get_attribute("class") == 'conborde ':
            dfTable = pd.DataFrame()
            headers = table.find_elements(By.TAG_NAME, "th")
            for header in headers:
                if header.text != ' ':
                    dfTable[header.text.strip()] = []
            rows = table.find_elements(By.TAG_NAME, "tr")
            for row in rows:
                info = []
                if row.find_element(By.XPATH, "..").tag_name ==
"tbody":
                    allData = row.find_elements(By.CLASS_NAME,
"notraducir")

                    for singleData in allData:
                        info.append(singleData.text.strip())
                    if rows.index(row) > 1:
                        info.append(dua)
                        info.append('Sumario')
                        dfTable.loc[rows.index(row) - 1] = info
                    else:
                        dfTable.loc[rows.index(row) - 1] = info
                        dfTable.loc[rows.index(row) - 1,
'DUA_COL'] = dua
                        dfTable.loc[rows.index(row) - 1,
'Partida_COL'] = 'Sumario'

            if resultSetTables.index(table) == 0:

```

```
        indexSumario0_after += 1
    elif resultSetTables.index(table) == 1:
        indexSumario1_after += 1
    elif resultSetTables.index(table) == 2:
        indexSumario2_after += 1
    elif resultSetTables.index(table) == 3:
        indexSumario3_after += 1

    sheetName = "Sumario" +
str(resultSetTables.index(table))

    if DUAS.index(dua) == 0:
        if resultSetTables.index(table) == 0:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexSumario0_prev, index=False)
            indexSumario0_after += 1
        elif resultSetTables.index(table) == 1:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexSumario1_prev, index=False)
            indexSumario1_after += 1
        elif resultSetTables.index(table) == 2:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexSumario2_prev, index=False)
            indexSumario2_after += 1
        elif resultSetTables.index(table) == 3:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexSumario3_prev, index=False)
            indexSumario3_after += 1
    else:
        if resultSetTables.index(table) == 0:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexSumario0_prev, header=False,
            index=False)
        elif resultSetTables.index(table) == 1:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexSumario1_prev, header=False,
            index=False)
        elif resultSetTables.index(table) == 2:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexSumario2_prev, header=False,
            index=False)
        elif resultSetTables.index(table) == 3:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexSumario3_prev, header=False,
            index=False)

    indexSumario0_prev = indexSumario0_after
    indexSumario1_prev = indexSumario1_after
    indexSumario2_prev = indexSumario2_after
    indexSumario3_prev = indexSumario3_after
```

```

        # dfTable.to_excel(excel_writer="Tablas DUAS
Importacion 04-05-2023.xlsx", sheet_name=sheetName, mode="a")
        tables[resultSetTables.index(table) + globalIndex +
1] = dfTable
        globalIndex += resultSetTables.index(table) + 1

    # TABLAS DUA
    resultSetTables = resultSetDUA.find_elements(By.TAG_NAME,
"table")
    for table in resultSetTables:
        parent = table.find_element(By.XPATH, "..")
        parent = parent.find_element(By.XPATH, "..")
        parent2 = parent.find_element(By.XPATH, "..")
        # me aseguro de coger solo las tablas visibles en la web
        if parent.get_attribute("class") == 'conborde ' or
parent2.get_attribute("class") == 'conborde ':
            dfTable = pd.DataFrame()
            headers = table.find_elements(By.TAG_NAME, "th")
            for header in headers:
                if header.text != ' ':
                    dfTable[header.text.strip()] = []
            rows = table.find_elements(By.TAG_NAME, "tr")
            for row in rows:
                info = []
                if row.find_element(By.XPATH, "..").tag_name ==


```



```
        indexDUA0_after += 1
        elif resultSetTables.index(table) == 1:
            dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexDUA1_prev, index=False)
            indexDUA0_after += 1
        else:
            if resultSetTables.index(table) == 0:
                dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexDUA0_prev, header=False,
                    index=False)

            elif resultSetTables.index(table) == 1:
                dfTable.to_excel(writer,
sheet_name=sheetName, startrow=indexDUA1_prev, header=False,
                    index=False)

        indexDUA0_prev = indexDUA0_after
        indexDUA1_prev = indexDUA1_after

        tables[resultSetTables.index(table) + globalIndex +
1] = dfTable
        if len(resultSetTables) > 0:
            globalIndex += resultSetTables.index(table) + 1

        # df.to_excel(excel_writer="Informacion DUAS Importacion 25-
04-2023.xlsx")
        print("Fin DUA: " + str(DUAS.index(dua) + 1) + "/" +
str(len(DUAS)))

    writer.close()
    print("FIN")
```

## 8.4 CÓDIGO PYTHON MODELO DE REGRESIÓN LINEAL

```
import pandas as pd
from sklearn.linear_model import LinearRegression
import numpy as np
from sklearn.metrics import mean_squared_error
import openpyxl
from sklearn import preprocessing
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split

label_encoder = preprocessing.LabelEncoder()

# Read data from Excel file
data = pd.read_excel('Datos MODELOS.xlsx', sheet_name="LR 17.0")

# Create a MinMaxScaler object
scaler = MinMaxScaler()

#X = pd.DataFrame(data) # Used without categorical variables

# Create dataframe including numerical variables and hot encoding
categorical variables
X = pd.get_dummies(data, columns=['Declarante:', 'Situación:', '30.
Localización de las mercancías:', '33.1-2 Código
TARIC:', 'Circuito:', '20.1 Código incoterms:'])

# Apply log transformation to some numerical variables
#X['06. Total de Bultos:'] = np.log10(X['06. Total de Bultos:'])
#X['Total Factura en €:'] = np.log10(X['Total Factura en €:'])
#X['35. Masa Bruta(Kg):'] = np.log10(X['35. Masa Bruta(Kg):'])

#X['Total Factura en €:'] = X['Total Factura en €:']**2

#X['Masa/Bultos'] = X['35. Masa Bruta(Kg):']/X['06. Total de
Bultos:']
#X['Factura/Masa'] = X['Total Factura en €:']/X['35. Masa
Bruta(Kg):']

# Normalize the columns of the DataFrame
X = pd.DataFrame(scaler.fit_transform(X), columns=X.columns)

# Create a column with a label encoding for categorical variables
needing a ranking order
X['Circuito:'] = label_encoder.fit_transform(data['Circuito:'])

variable_estudio = pd.read_excel('Datos raw.xlsx',
sheet_name="Sheet1")
y = np.log10(variable_estudio['Valor en Aduana:'])
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Create a linear regression model
model = LinearRegression()

# Train the model
model.fit(X_train, y_train)

# Predict the target values
y_pred = model.predict(X_test)

# Calculate mean squared error (MSE)
mse = mean_squared_error(y_test, y_pred)

# Calculate root mean squared error (RMSE)
rmse = np.sqrt(mse)

# Print the RMSE
print("Mean Squared Error (MSE):", mse)
print("Root Mean Squared Error (RMSE):", rmse)

# Calculate the absolute error
absolute_error = np.abs(y_test - y_pred)

# Calculate the error percentage
error_percentage = (absolute_error / y_test)

# Create a DataFrame from the array
df = pd.DataFrame(error_percentage)

# Save DataFrame to Excel
df.to_excel('Histograma Regresion Lineal.xlsx', index=False,
sheet_name="Histograma")
```

## 8.5 CÓDIGO PYTHON MODELO DE REGRESIÓN POLINÓMICA

```
import pandas as pd
from sklearn.linear_model import LinearRegression
import numpy as np
from sklearn.metrics import mean_squared_error
import openpyxl
from sklearn import preprocessing
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures

label_encoder = preprocessing.LabelEncoder()

# Read data from Excel file
data = pd.read_excel('Datos MODELOS.xlsx', sheet_name="LR 17.0")

# Create a MinMaxScaler object
scaler = MinMaxScaler()

#X = pd.DataFrame(data) # Used without categorical variables

# Create dataframe including numerical variables and hot encoding
categorical variables
X = pd.get_dummies(data, columns=['Declarante:', 'Situación:', '30.
Localización de las mercancías:', '33.1-2 Código
TARIC:', 'Circuito:', '20.1 Código incoterms:'])

# Apply log transformation to some numerical variables
#X['06. Total de Bultos:'] = np.log10(X['06. Total de Bultos:'])
#X['Total Factura en €:'] = np.log10(X['Total Factura en €:'])
#X['35. Masa Bruta(Kg):'] = np.log10(X['35. Masa Bruta(Kg):'])

#X['Total Factura en €:'] = X['Total Factura en €:']**2

#X['Masa/Bultos'] = X['35. Masa Bruta(Kg):']/X['06. Total de
Bultos:']
#X['Factura/Masa'] = X['Total Factura en €:']/X['35. Masa
Bruta(Kg):']

# Normalize the columns of the DataFrame
X = pd.DataFrame(scaler.fit_transform(X), columns=X.columns)

# Create a column with a label encoding for categorical variables
needing a ranking order
X['Circuito:'] = label_encoder.fit_transform(data['Circuito:'])

variable_estudio = pd.read_excel('Datos raw.xlsx',
sheet_name="Sheet1")
y = np.log10(variable_estudio['Valor en Aduana:'])
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Create polynomial features
degree = 3 # Specify the degree of the polynomial
poly_features = PolynomialFeatures(degree=degree, include_bias=False)
X_poly = poly_features.fit_transform(X_train)

# Fit the polynomial regression model
model = LinearRegression()
model.fit(X_poly, y_train)

# Generate predictions
X_poly_pred = poly_features.transform(X_test)
y_pred = model.predict(X_poly_pred)

# Calculate mean squared error (MSE)
mse = mean_squared_error(y_test, y_pred)

# Calculate root mean squared error (RMSE)
rmse = np.sqrt(mse)

# Print the RMSE
print("Mean Squared Error (MSE):", mse)
print("Root Mean Squared Error (RMSE):", rmse)

# Calculate the absolute error
absolute_error = np.abs(y_test - y_pred)

# Calculate the error percentage
error_percentage = (absolute_error / y_test)

# Create a DataFrame from the array
df = pd.DataFrame(error_percentage)

# Save DataFrame to Excel
df.to_excel('Histograma PR Grado 3.xlsx', index=False,
sheet_name="Histograma")
```

## 8.6 CÓDIGO PYTHON MODELO DE ÁRBOL DE DECISIÓN

```
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn import preprocessing
from sklearn.preprocessing import MinMaxScaler
import numpy as np
from sklearn.metrics import mean_squared_error
from sklearn.tree import DecisionTreeRegressor

label_encoder = preprocessing.LabelEncoder()

# Read data from Excel file
data = pd.read_excel('Datos MODELOS.xlsx', sheet_name="LR 17.0")

# Create a MinMaxScaler object
scaler = MinMaxScaler()

#X = pd.DataFrame(data) # Used without categorical variables

# Create dataframe including numerical variables and hot encoding
categorical variables
X = pd.get_dummies(data, columns=['Declarante:', 'Situación:', '30.
Localización de las mercancías:', '33.1-2 Código
TARIC:', 'Circuito:', '20.1 Código incoterms:'])

# Apply log transformation to some numerical variables
#X['06. Total de Bultos:'] = np.log10(X['06. Total de Bultos:'])
#X['Total Factura en €:'] = np.log10(X['Total Factura en €:'])
#X['35. Masa Bruta(Kg):'] = np.log10(X['35. Masa Bruta(Kg):'])

#X['Total Factura en €:'] = X['Total Factura en €:']**2

#X['Masa/Bultos'] = X['35. Masa Bruta(Kg):']/X['06. Total de
Bultos:']
#X['Factura/Masa'] = X['Total Factura en €:']/X['35. Masa
Bruta(Kg):']

# Normalize the columns of the DataFrame
X = pd.DataFrame(scaler.fit_transform(X), columns=X.columns)

# Create a column with a label encoding for categorical variables
needing a ranking order
X['Circuito:'] = label_encoder.fit_transform(data['Circuito:'])

variable_estudio = pd.read_excel('Datos raw.xlsx',
sheet_name="Sheet1")
y = np.log10(variable_estudio['Valor en Aduana:'])
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Create a decision tree regressor
model = DecisionTreeRegressor()

# Fit the model on the training data
model.fit(X_train, y_train)

# Make predictions on the test data
y_pred = model.predict(X_test)

# Calculate mean squared error
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)

# Calculate the absolute error
absolute_error = np.abs(y_test - y_pred)

# Calculate the error percentage
error_percentage = (absolute_error / y_test)

# Create a DataFrame from the array
df = pd.DataFrame(error_percentage)

# Save DataFrame to Excel
df.to_excel('Histograma Arbol Decision.xlsx', index=False,
sheet_name="Histograma")
```

## 8.7 CÓDIGO PYTHON MODELO DE BAGGING

```
# ----- Optimal Number of Estimators -----  
----- #  
  
import pandas as pd  
import numpy as np  
from sklearn.ensemble import BaggingRegressor  
from sklearn.tree import DecisionTreeRegressor  
from sklearn.model_selection import train_test_split, cross_val_score  
from sklearn.metrics import mean_squared_error  
from sklearn import preprocessing  
from sklearn.preprocessing import MinMaxScaler  
  
label_encoder = preprocessing.LabelEncoder()  
  
# Read data from Excel file  
data = pd.read_excel('Datos MODELOS.xlsx', sheet_name="LR 17.0")  
  
# Create a MinMaxScaler object  
scaler = MinMaxScaler()  
  
#X = pd.DataFrame(data) # Used without categorical variables  
  
# Create dataframe including numerical variables and hot encoding  
categorical variables  
X = pd.get_dummies(data, columns=['Declarante:', 'Situación:', '30.  
Localización de las mercancías:', '33.1-2 Código TARIC:', 'Circuito:', '20.1  
Código incoterms:'])  
  
# Apply log transformation to some numerical variables  
#X['06. Total de Bultos:'] = np.log10(X['06. Total de Bultos:'])  
#X['Total Factura en €:'] = np.log10(X['Total Factura en €:'])  
#X['35. Masa Bruta (Kg):'] = np.log10(X['35. Masa Bruta (Kg):'])  
  
#X['Total Factura en €:'] = X['Total Factura en €:']**2  
  
#X['Masa/Bultos'] = X['35. Masa Bruta (Kg):']/X['06. Total de Bultos:']  
#X['Factura/Masa'] = X['Total Factura en €:']/X['35. Masa Bruta (Kg):']  
  
# Normalize the columns of the DataFrame  
X = pd.DataFrame(scaler.fit_transform(X), columns=X.columns)  
  
# Create a column with a label encoding for categorical variables needing  
a ranking order  
X['Circuito:'] = label_encoder.fit_transform(data['Circuito:'])  
  
variable_estudio = pd.read_excel('Datos raw.xlsx', sheet_name="Sheet1")  
y = np.log10(variable_estudio['Valor en Aduana:'])  
  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,  
random_state=42)
```



```
# Create a decision tree regressor as the base model
base_model = DecisionTreeRegressor()

# Initialize lists to store the results
estimators = []
mse_scores = []

# Test different numbers of estimators
for n_estimators in range(1, 21):
    # Create a bagging regressor with the current number of estimators
    model = BaggingRegressor(base_estimator=base_model,
                             n_estimators=n_estimators, random_state=42)

    # Perform cross-validation
    scores = cross_val_score(model, X_train, y_train, cv=5,
                              scoring='neg_mean_squared_error')

    # Calculate mean squared error
    mse = -np.mean(scores)

    # Store the results
    estimators.append(n_estimators)
    mse_scores.append(mse)

# Find the index of the minimum mean squared error
best_index = np.argmin(mse_scores)

# Get the optimal number of estimators and corresponding mean squared
error
optimal_estimators = estimators[best_index]
optimal_mse = mse_scores[best_index]

print("Optimal Number of Estimators:", optimal_estimators)
print("Mean Squared Error with Optimal Estimators:", optimal_mse)

# ----- BAGGING MODEL ----- #

import pandas as pd
from sklearn.ensemble import BaggingRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
from sklearn import preprocessing
from sklearn.preprocessing import MinMaxScaler
import numpy as np

label_encoder = preprocessing.LabelEncoder()

# Read data from Excel file
data = pd.read_excel('Datos MODELOS.xlsx', sheet_name="LR 17.0")

# Create a MinMaxScaler object
scaler = MinMaxScaler()

#X = pd.DataFrame(data) # Used without categorical variables
```

```
# Create dataframe including numerical variables and hot encoding
categorical variables
X = pd.get_dummies(data, columns=['Declarante:', 'Situación:', '30.
Localización de las mercancías:', '33.1-2 Código TARIC:', 'Circuito:', '20.1
Código incoterms:'])

# Apply log transformation to some numerical variables
#X['06. Total de Bultos:'] = np.log10(X['06. Total de Bultos:'])
#X['Total Factura en €:'] = np.log10(X['Total Factura en €:'])
#X['35. Masa Bruta (Kg):'] = np.log10(X['35. Masa Bruta (Kg):'])

#X['Total Factura en €:'] = X['Total Factura en €:']**2

#X['Masa/Bultos'] = X['35. Masa Bruta (Kg):']/X['06. Total de Bultos:']
#X['Factura/Masa'] = X['Total Factura en €:']/X['35. Masa Bruta (Kg):']

# Normalize the columns of the DataFrame
X = pd.DataFrame(scaler.fit_transform(X), columns=X.columns)

# Create a column with a label encoding for categorical variables needing
a ranking order
X['Circuito:'] = label_encoder.fit_transform(data['Circuito:'])

variable_estudio = pd.read_excel('Datos raw.xlsx', sheet_name="Sheet1")
y = np.log10(variable_estudio['Valor en Aduana:'])

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Create a decision tree regressor
base_model = DecisionTreeRegressor()

# Create a bagging regressor
model = BaggingRegressor(base_estimator=base_model, n_estimators=4,
random_state=42)

# Fit the model on the training data
model.fit(X_train, y_train)

# Make predictions on the test data
y_pred = model.predict(X_test)

# Calculate mean squared error
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)

# Calculate the absolute error
absolute_error = np.abs(y_test - y_pred)

# Calculate the error percentage
error_percentage = (absolute_error / y_test)

# Create a DataFrame from the array
df = pd.DataFrame(error_percentage)
```

```
# Save DataFrame to Excel
df.to_excel('Histograma Bagging 4 estimadores.xlsx', index=False,
sheet_name="Histograma")
```

## 8.8 CÓDIGO PYTHON MODELO DE XGBOOST

```
import pandas as pd
from sklearn.linear_model import LinearRegression
import numpy as np
from sklearn.metrics import mean_squared_error
import openpyxl
from sklearn import preprocessing
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from xgboost import XGBRegressor
from sklearn.preprocessing import PolynomialFeatures

label_encoder = preprocessing.LabelEncoder()

# Read data from Excel file
data = pd.read_excel('Datos MODELOS.xlsx', sheet_name="LR 17.0")

# Create a MinMaxScaler object
scaler = MinMaxScaler()

#X = pd.DataFrame(data) # Used without categorical variables

# Create dataframe including numerical variables and hot encoding
categorical variables
X = pd.get_dummies(data, columns=['Declarante:', 'Situación:', '30.
Localización de las mercancías:', '33.1-2 Código TARIC:', 'Circuito:', '20.1
Código incoterms:'])

# Apply log transformation to some numerical variables
#X['06. Total de Bultos:'] = np.log10(X['06. Total de Bultos:'])
#X['Total Factura en €:'] = np.log10(X['Total Factura en €:'])
#X['35. Masa Bruta(Kg):'] = np.log10(X['35. Masa Bruta(Kg):'])

#X['Total Factura en €:'] = X['Total Factura en €:']**2

#X['Masa/Bultos'] = X['35. Masa Bruta(Kg):']/X['06. Total de Bultos:']
#X['Factura/Masa'] = X['Total Factura en €:']/X['35. Masa Bruta(Kg):']

# Normalize the columns of the DataFrame
X = pd.DataFrame(scaler.fit_transform(X), columns=X.columns)

# Create a column with a label encoding for categorical variables needing
a ranking order
X['Circuito:'] = label_encoder.fit_transform(data['Circuito:'])

variable_estudio = pd.read_excel('Datos raw.xlsx', sheet_name="Sheet1")
y = np.log10(variable_estudio['Valor en Aduana:'])

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
```

```
# Create an XGBoost regression model
model = XGBRegressor()

# Fit the model on the training data
model.fit(X_train, y_train)

# Make predictions on the test data
y_pred = model.predict(X_test)

# Calculate mean squared error (MSE)
mse = mean_squared_error(y_test, y_pred)

# Calculate root mean squared error (RMSE)
rmse = np.sqrt(mse)

# Print the RMSE
print("Mean Squared Error (MSE):", mse)

# Calculate the absolute error
absolute_error = np.abs(y_test - y_pred)

# Calculate the error percentage
error_percentage = (absolute_error / y_test)

# Create a DataFrame from the array
df = pd.DataFrame(error_percentage)

# Save DataFrame to Excel
df.to_excel('Histograma XGBoost.xlsx', index=False,
sheet_name="Histograma")
```

## 8.9 CÓDIGO PYTHON MODELO DE RANDOM FOREST

```
import pandas as pd
from sklearn.linear_model import LinearRegression
import numpy as np
from sklearn.metrics import mean_squared_error
import openpyxl
from sklearn import preprocessing
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import PolynomialFeatures

label_encoder = preprocessing.LabelEncoder()

# Read data from Excel file
data = pd.read_excel('Datos MODELOS.xlsx', sheet_name="LR 17.0")

# Create a MinMaxScaler object
scaler = MinMaxScaler()

#X = pd.DataFrame(data) # Used without categorical variables

# Create dataframe including numerical variables and hot encoding
categorical variables
X = pd.get_dummies(data, columns=['Declarante:', 'Situación:', '30.
Localización de las mercancías:', '33.1-2 Código TARIC:', 'Circuito:', '20.1
Código incoterms:'])

# Apply log transformation to some numerical variables
#X['06. Total de Bultos:'] = np.log10(X['06. Total de Bultos:'])
#X['Total Factura en €:'] = np.log10(X['Total Factura en €:'])
#X['35. Masa Bruta(Kg):'] = np.log10(X['35. Masa Bruta(Kg):'])

#X['Total Factura en €:'] = X['Total Factura en €:']**2

#X['Masa/Bultos'] = X['35. Masa Bruta(Kg):']/X['06. Total de Bultos:']
#X['Factura/Masa'] = X['Total Factura en €:']/X['35. Masa Bruta(Kg):']

# Normalize the columns of the DataFrame
X = pd.DataFrame(scaler.fit_transform(X), columns=X.columns)

# Create a column with a label encoding for categorical variables needing
a ranking order
X['Circuito:'] = label_encoder.fit_transform(data['Circuito:'])

variable_estudio = pd.read_excel('Datos raw.xlsx', sheet_name="Sheet1")
y = np.log10(variable_estudio['Valor en Aduana:'])

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Create a random forest regression model
```

```
model = RandomForestRegressor(n_estimators=100, random_state=42)

# Fit the model on the training data
model.fit(X_train, y_train)

# Get the feature importance
feature_importance = model.feature_importances_

# Create a DataFrame to hold the feature importance values
importance_df = pd.DataFrame({'Feature': X.columns, 'Importance':
feature_importance})

# Sort the DataFrame by importance in descending order
importance_df = importance_df.sort_values('Importance', ascending=False)

# Make predictions on the test data
y_pred = model.predict(X_test)

# Calculate mean squared error (MSE)
mse = mean_squared_error(y_test, y_pred)

# Calculate root mean squared error (RMSE)
rmse = np.sqrt(mse)

# Print the RMSE
print("Root Mean Squared Error (RMSE):", rmse)

# Print the MSE
print("Mean Squared Error (MSE):", mse)

# Calculate the absolute error
absolute_error = np.abs(y_test - y_pred)

# Calculate the error percentage
error_percentage = (absolute_error / y_test)

# Create a DataFrame from the array
df = pd.DataFrame(error_percentage)

# Save DataFrame to Excel
df.to_excel('Histograma Random Forest.xlsx', index=False,
sheet_name="Histograma")
```

## ANEXO II

### 8.10 ALINEACIÓN DEL PROYECTO CON LOS ODS

“Los objetivos de desarrollo sostenible son el plan maestro para conseguir un futuro sostenible para todos. Se interrelacionan entre sí e incorporan los desafíos globales a los que nos enfrentamos día a día, como la pobreza, la desigualdad, el clima, la degradación ambiental, la prosperidad, la paz y la justicia. Para no dejar a nadie atrás, es importante que logremos cumplir con cada uno de estos objetivos para 2030.” (Naciones Unidas, s.f.)

Así, el proyecto propuesto, tal y como se especifica en el apartado 4.1 Justificación, tiene como objetivo promover el desarrollo sostenible. En relación con los objetivos de desarrollo sostenible incide directamente sobre 4 de ellos: “Objetivo 7: Energía asequible y no contaminante”, “Objetivo 8: Trabajo decente y crecimiento económico”, “Objetivo 12: Producción y consumo responsables” y “Objetivo 13: Acción por el clima”.

#### 8.10.1 OBJETIVO 7: ENERGÍA ASEQUIBLE Y NO CONTAMINANTE

Uno de los objetivos que se desprenden de la utilización de la herramienta propuesta es una mayor utilización de energías renovables y limpias en el transporte de las mercancías. La búsqueda de rutas y modos de transporte mas eficientes y económicos puede llevar a la utilización de medios de transporte mas sostenibles, como el transporte ferroviario o el transporte marítimo, que pueden utilizar energías renovables para su funcionamiento.

#### 8.10.2 OBJETIVO 8: TRABAJO DECENTE Y CRECIMIENTO ECONÓMICO

“Un crecimiento económico inclusivo y sostenido puede impulsar el progreso, crear empleos decentes para todos y mejorar los estándares de vida” (Naciones Unidas, s.f.):

Así, el uso de la herramienta propuesta impulsará a las empresas a aventurarse en el mundo del comercio internacional al tener una optimización de los costes aduaneros que le permitan acceder al mercado. Cuanto mayor sea el número de empresas que comercien



internacionalmente, mayor será la inversión extranjera en aquellos países del tercer mundo o en vías de desarrollo, y por tanto mayor será la creación de empleo y consecuentemente el crecimiento económico del país al que se accede.

### **8.10.3 OBJETIVO 12: PRODUCCIÓN Y CONSUMO RESPONSABLES**

“El consumo y la producción mundiales (fuerzas impulsoras de la economía mundial) dependen del uso del medio ambiente natural y de los recursos de una manera que continúa teniendo efectos destructivos sobre el planeta.

El progreso económico y social conseguido durante el último siglo ha estado acompañado de una degradación medioambiental que está poniendo en peligro los mismos sistemas de los que depende nuestro desarrollo futuro (y ciertamente, nuestra supervivencia).” (Naciones Unidas, s.f.)

De esta manera, el uso de la herramienta conlleva una optimización de los costes aduaneros, que como se ha comentado anteriormente, implica indirectamente la inversión en otras partes del modelo productivo. Así, la herramienta ayuda a encontrar soluciones más eficientes para a gestión de la cadena de suministros, lo que reduce la cantidad de productos que se desperdician o se descartan en el proceso de transporte. Así como reducir el consumo de recursos y a minimizar el impacto ambiental del transporte de mercancías.

### **8.10.4 OBJETIVO 13: ACCIÓN POR EL CLIMA**

“El cambio climático está afectando a todos los países de todos los continentes. Está alterando las economías nacionales y afectando a distintas vidas. Los sistemas meteorológicos están cambiando, los niveles del mar están subiendo y los fenómenos meteorológicos son cada vez más extremos.

A pesar de que se estima que las emisiones de gases de efecto invernadero caigan alrededor de un 6 % en 2020 debido a las restricciones de movimiento y las recesiones económicas derivadas de la pandemia de la COVID-19, esta mejora es solo temporal. El cambio climático

no se va a pausar. Una vez que la economía mundial comience a recuperarse de la pandemia, se espera que las emisiones vuelvan a niveles mayores.” (Naciones Unidas, s.f.)

Por ello, el uso de la herramienta propuesta se enfoca hacia una menor emisión de gases de efecto invernadero en el transporte de mercancías. La optimización de los costes ayuda a encontrar rutas mas eficientes y reducir el numero de viajes necesarios, lo que puede tener un impacto positivo en el medio ambiente y en la huella de carbono del cliente.