



Universidad de San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería Ciencias y Sistemas

**DISEÑO DE INVESTIGACIÓN DE UNA PROPUESTA DE MODELO PARA PRONÓSTICO  
DEL COMPORTAMIENTO DE IMPORTACIONES DE TUBERÍA DE ACERO PROVENIENTE  
DE CHINA PARA ANALIZAR VENTAJAS COMPETITIVAS**

**Edwin Gustavo Lima Dubon**

Asesorado por el Msc. Marco Tulio Aldana Prillwitz

Guatemala, abril de 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**DISEÑO DE INVESTIGACIÓN DE UNA PROPUESTA DE MODELO PARA PRONÓSTICO  
DEL COMPORTAMIENTO DE IMPORTACIONES DE TUBERÍA DE ACERO PROVENIENTE  
DE CHINA PARA ANALIZAR VENTAJAS COMPETITIVAS**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA  
FACULTAD DE INGENIERÍA  
POR

**EDWIN GUSTAVO LIMA DUBON**

ASESORADO POR EL MSC. MARCO TULIO ALDANA PRILLWITZ

AL CONFERÍRSELE EL TÍTULO DE

**INGENIERO EN CIENCIAS Y SISTEMAS**

GUATEMALA, ABRIL DE 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA  
FACULTAD DE INGENIERÍA



**NÓMINA DE JUNTA DIRECTIVA**

DECANA	Ing. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

**TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO**

DECANO	Ing. Pedro Antonio Aguilar Polanco
EXAMINADOR	Ing. Pedro Pablo Hernández Ramírez
EXAMINADOR	Ing. Luis Fernando Espino Barrios
EXAMINADOR	Ing. Herman Igor Véliz Linares
SECRETARIA	Inga. Lesbia Magalí Herrera López

## **HONORABLE TRIBUNAL EXAMINADOR**

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**DISEÑO DE INVESTIGACIÓN DE UNA PROPUESTA DE MODELO PARA PRONÓSTICO  
DEL COMPORTAMIENTO DE IMPORTACIONES DE TUBERÍA DE ACERO PROVENIENTE  
DE CHINA PARA ANALIZAR VENTAJAS COMPETITIVAS**

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha enero 2022.

**Edwin Gustavo Lima Dubon**



**EEPFI-PP-0145-2022**

Guatemala, 12 de enero de 2022

Director  
null  
Escuela De Ingenieria En Sistemas  
Presente.

Estimado Ing. Alonzo

Reciba un cordial saludo de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería.

El propósito de la presente es para informarle que se ha revisado y aprobado el Diseño de Investigación titulado: **PROPUESTA DE MODELO PARA PRONOSTICÓ DEL COMPORTAMIENTO DE IMPORTACIONES DE TUBERIA DE ACERO PROVENIENTE DE CHINA PARA ANALIZAR VENTAJAS COMPETITIVAS**, el cual se enmarca en la línea de investigación: **Análisis de datos - Análisis de datos**, presentado por el estudiante **Edwin Gustavo Lima Dubon** carné número **200611383**, quien optó por la modalidad del "PROCESO DE GRADUACIÓN DE LOS ESTUDIANTES DE LA FACULTAD DE INGENIERÍA OPCIÓN ESTUDIOS DE POSTGRADO". Previa a culminar sus estudios en la Maestría en ARTES en Ingeniería Para La Industria Con Especialidad En Ciencias De La Computación.

Y habiendo cumplido y aprobado con los requisitos establecidos en el normativo de este Proceso de Graduación en el Punto 6.2, aprobado por la Junta Directiva de la Facultad de Ingeniería en el Punto Décimo, Inciso 10.2 del Acta 28-2011 de fecha 19 de septiembre de 2011, firmo y sello la presente para el trámite correspondiente de graduación de Pregrado.

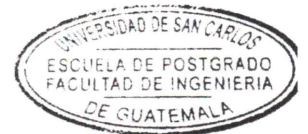
Atentamente,

*"Id y Enseñad a Todos"*

Mtro. Marco Tulio Aldana Prillwitz  
Asesor(a)

**Marco Tulio Aldana Prillwitz**  
MSc. BI y Data Analytics  
Colegio Humanidades No. 3076

Mtro. Mario Renato Escobedo Martinez  
Coordinador(a) de Maestría



Mtro. Edgar Darío Álvarez Cotí  
Director  
Escuela de Estudios de Postgrado  
Facultad de Ingeniería





EEP-EICS-0145-2022

El Director de la Escuela De Ingenieria En Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del Asesor, el visto bueno del Coordinador y Director de la Escuela de Estudios de Postgrado, del Diseño de Investigación en la modalidad Estudios de Pregrado y Postgrado titulado: **PROPUESTA DE MODELO PARA PRONOSTICÓ DEL COMPORTAMIENTO DE IMPORTACIONES DE TUBERIA DE ACERO PROVENIENTE DE CHINA PARA ANALIZAR VENTAJAS COMPETITIVAS**, presentado por el estudiante universitario **Edwin Gustavo Lima Dubon**, procedo con el Aval del mismo, ya que cumple con los requisitos normados por la Facultad de Ingeniería en esta modalidad.

ID Y ENSEÑAD A TODOS

Ing. Carlos Gustavo Alonzo  
Director  
Escuela De Ingenieria En Sistemas

Guatemala, enero de 2022



Decanato  
Facultad de Ingeniería  
24189101- 24189102  
secretariadecanato@ingenieria.usac.edu.gt

LNG.DECANATO.OI.318.2022

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al Trabajo de Graduación titulado: **DISEÑO DE INVESTIGACIÓN DE UNA PROPUESTA DE MODELO PARA PRONÓSTICO DEL COMPORTAMIENTO DE IMPORTACIONES DE TUBERÍA DE ACERO PROVENIENTE DE CHINA PARA ANALIZAR VENTAJAS COMPETITIVAS**, presentado por: **Edwin Gustavo Lima Dubon**, después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

Inga. Aurelia Anabela Cordova Estrada

Decana



Guatemala, abril de 2022

AACE/gaoc

## **ACTO QUE DEDICO A:**

- Mis padres** Por el apoyo incondicional durante toda mi carrera y ser inspiración para seguir adelante.
- Mis hermanas** Liza, Luisa y Grecia Lima por impulsarme a no darme por vencido.
- Mi primo** Por motivarme a que hiciera mi trabajo de graduación.
- Mis amigos** Por ayudarme durante y después de mis estudios.



## **AGRADECIMIENTOS A:**

**Universidad de San  
Carlos de Guatemala**

Por brindarme las herramientas y la oportunidad para seguir creciendo como profesional.

**Ingeniero**

Marco Aldana por apoyarme para lograr terminar mi trabajo de graduación.

## ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES .....	V
LISTA DE SÍMBOLOS .....	VII
GLOSARIO .....	IX
RESUMEN .....	XI
1. INTRODUCCIÓN .....	1
2. ANTECEDENTES .....	3
3. PLANTEAMIENTO DEL PROBLEMA .....	11
3.1. Contexto general .....	11
3.2. Descripción del problema .....	11
3.3. Formulación del problema .....	11
3.4. Delimitación del problema .....	12
4. JUSTIFICACIÓN .....	13
5. OBJETIVOS .....	15
6. NECESIDADES POR CUBRIR Y ESQUEMA DE LA SOLUCIÓN .....	17
7. MARCO TEÓRICO.....	19
7.1. Variables económicas.....	19
7.1.1. Producto interno bruto .....	19
7.1.2. Tipo de cambio .....	19

7.1.3.	Inflación .....	19
7.1.4.	Balanza de pagos.....	20
7.2.	Estadística.....	20
7.2.1.	Estadística Descriptiva .....	20
7.2.2.	Estadística Inferencial .....	20
7.2.3.	P-Value.....	21
7.3.	Pronósticos .....	21
7.3.1.	Métodos de pronóstico .....	21
7.3.2.	Métodos cualitativos .....	21
7.3.3.	Métodos de proyección histórica .....	22
7.3.4.	Métodos causales .....	22
7.4.	Series de tiempo .....	22
7.4.1.	Prueba Dickey Fuller .....	22
7.5.	Análisis predictivo con series de tiempo.....	23
7.5.1.	Método ARMA .....	23
7.5.2.	Método ARIMA .....	24
7.5.3.	Método SARIMA.....	25
7.5.4.	Método de suavizado exponencial .....	26
7.6.	Ciencia de datos .....	26
7.6.1.	Flujo de proceso de ciencia de datos .....	26
7.6.1.1.	Obtener .....	27
7.6.1.2.	Limpiar la información .....	27
7.6.1.3.	Explorar la información.....	28
7.6.1.4.	Modelar la información .....	28
7.6.1.5.	Interpretar la información.....	28
7.7.	Inteligencia artificial .....	29
7.8.	Machine Learning.....	29
7.8.1.	Supervised learning.....	30
7.8.2.	Unsupervised learning.....	30

7.8.3.	Reinforcement learning.....	30
7.8.4.	Métodos de pronóstico de machine learning .....	31
7.8.4.1.	XGBoost .....	31
7.9.	Herramientas .....	31
7.9.1.	Python .....	31
7.9.2.	NumPy.....	32
7.9.3.	Pandas .....	32
7.9.4.	Statsmodels.....	32
7.9.5.	Scikit-learn .....	32
7.9.6.	KNIME .....	32
7.9.7.	Jupyter Notebook.....	33
7.9.8.	Visual Studio Code .....	33
8.	PROPUESTA DE ÍNDICE DE CONTENIDOS .....	35
9.	METODOLOGÍA.....	41
9.1.	Características del estudio .....	41
9.2.	Unidades de análisis .....	41
9.3.	Variables.....	42
9.4.	Fases del estudio .....	43
9.4.1.	Fase 1. Revisión de literatura .....	43
9.4.2.	Fase 2. Diseño del modelo de pronóstico.....	44
9.4.3.	Fase 3. Gestión o recolección de la información ....	44
9.4.4.	Fase 4. Configuración de ambientes .....	44
9.4.5.	Fase 5. Análisis de información .....	44
9.4.6.	Fase 6. Modelar la información con series de tiempo.....	45
9.4.7.	Fase 7. Modelar la información con machine learning.....	45

9.4.8.	Fase 8. Interpretación de información .....	46
9.4.9.	Fase 9. Selección de modelo de pronóstico adecuado.....	46
10.	TÉCNICAS DE ANÁLISIS DE INFORMACIÓN .....	47
10.1.	Preparación de la información.....	47
10.2.	Análisis de información .....	47
10.3.	Presentación de los resultados .....	48
11.	CRONOGRAMA .....	49
12.	FACTIBILIDAD DEL ESTUDIO .....	51
12.1.	Recurso humano necesario .....	51
12.2.	Acceso a la información .....	51
12.3.	Recursos tecnológicos .....	51
12.4.	Tiempo necesario para el desarrollo de la solución .....	52
12.5.	Recursos financieros.....	52
	REFERENCIAS .....	55

## ÍNDICE DE ILUSTRACIONES

### FIGURAS

1. Diagrama de arquitectura.....	18
2. Imagen flujo de proceso de ciencia de datos.....	27

### TABLAS

I. Tipos de variables.....	42
II. Cronograma de actividades.....	49
III. Recursos financieros.....	52



## LISTA DE SÍMBOLOS

<b>Símbolo</b>	<b>Significado</b>
<b>h</b>	Horas
<b>Q</b>	Quetzales





## GLOSARIO

<b>AI</b>	Artificial intelligence
<b>ANN</b>	Artificial neural network
<b>API</b>	Application programming interface
<b>AR</b>	Auto regressive
<b>ARIMA</b>	Autoregressive integrated moving average
<b>ARMA</b>	Autoregressive moving aerge
<b>AR-NET</b>	Auto Regressive Neural Network
<b>CSV</b>	Comma separated values
<b>EPI</b>	El profesional de la información
<b>Gb</b>	GigaByte
<b>GBOOSTING</b>	Gradient boosting
<b>LIGHTGBM</b>	Light gradient boosting machine

<b>LSTM</b>	Log structured merge-tree
<b>MB</b>	MegaByte
<b>ML</b>	Machine Learning
<b>OSEM</b>	Obtain Scrubbing Exploring Modeling Interpreting
<b>PIB</b>	Producto interno bruto
<b>RNN</b>	Recurrent neural network
<b>SARIMA</b>	Seasonal Autoregressive integrated moving average
<b>SAT</b>	Superintendencia de administracion tributaria
<b>SVM</b>	Support vector machine
<b>Tb</b>	TeraByte
<b>TRADE-DSM</b>	Trade decision support model
<b>Vram</b>	Video random access memory

## **RESUMEN**

En Guatemala las empresas en diferentes sectores carecen de modelos de pronóstico de importaciones, tal es el caso de las empresas que se dedican a la importación de tubería de acero, la carencia de esta información limita su habilidad de crecimiento y poder de negociación.

Los modelos de predicción del comportamiento de importaciones de tubería proveniente de China permitirían contar con conocimiento, el cual le brindaría una ventaja competitiva a la hora de realizar toma de decisiones.

Esta investigación busca realizar un modelo de pronóstico haciendo uso de la ciencia de datos, con técnicas como: las series de tiempos, detección de patrones y aprendizaje automático. Por lo que el diseño de un modelo de predicción como herramienta de análisis predictivo puede ayudar interpretar patrones de comportamiento en la información recolectada por distintas fuentes.



# 1. INTRODUCCIÓN

El problema que se enfrentan las empresas que se dedican a la comercialización de tubería de acero proveniente de China, es que no cuentan con pronósticos sobre el comportamiento de importaciones de tubería de acero que le proporcionen ventajas a la hora de realizar negociaciones.

El diseño de un modelo de predicción del comportamiento de importaciones de tubería de acero proveniente de China permitirá contar con conocimiento que las empresas pueden utilizar para la toma de decisiones, por medio de identificar los métodos que se pueden emplear para el análisis predictivo, así como su relación con variables económicas y determinar posibles optimizaciones por métodos de machine learning.

En el primer capítulo se recopilan los antecedentes de trabajos que en los cuales se ha realizado pronósticos, recolecciones de datos utilizando distintas técnicas, herramientas y metodologías, esto con el fin de analizar cómo abordar el modelo propuesto.

En el segundo capítulo se encuentra la justificación acerca la cual se realiza este trabajo de investigación.

En el tercer capítulo se determina el alcance en el que se encuadra el estudio siendo de diseño de tendencia.

En el cuarto capítulo se presenta el marco teórico en donde se encuentra la información recolectada que da soporte al trabajo de investigación, el cual contiene las herramientas, metodologías y la arquitectura a utilizar.

En el quinto capítulo se lleva a cabo la elaboración del modelo de pronóstico propuesto.

En el sexto y último capítulo se discuten los resultados obtenidos y se selecciona el modelo más adecuado.

## 2. ANTECEDENTES

El estudio modelo de ARIMA para pronosticar el precio del arroz de calidad media para anticipar las fluctuaciones de precios (Ohyver y Pudjihastuti, 2018). En la cual se desarrolló una investigación experimental para desarrollar un modelo de series de tiempo, el cual permita hacer previsión del precio del arroz utilizando ARIMA. Analizar la previsión del precio es una acción que se realiza con el fin de ayudar al gobierno a monitorearlo y controlarlo.

El arroz es uno de los principales productos en indonesia, debido a esto el gobierno necesita garantizar el abastecimiento adecuado del mismo, por lo que debe realizar monitoreo de los precios para poder controlarlo. El uso de métodos de series como de tiempo como el método ARIMA permite realizar previsión del precio con la finalidad de poder analizar los patrones de comportamiento del arroz.

La utilización de métodos estadísticos como el ARIMA para poder predecir resultados con base a patrones detectados en colecciones de datos, permite analizar datos pasados y generar una previsión del comportamiento de los datos a futuro.

Cerrillo (2018) datos masivos y datos abiertos para una gobernanza inteligente. El profesional de la información (EPI), investigación explicativa. Determina la importancia del uso de grandes cantidades de información pública. Tener acceso a grandes cantidades de información pública permite contar con información adecuada para análisis de datos.



La gobernanza inteligente se basa en el uso de las ingentes cantidades de datos que la administración pública genera y recaba en el desarrollo de sus actividades y en sus relaciones con la ciudadanía y las empresas. Es importante conocer como el uso de grandes volúmenes de datos a disposición de las administraciones públicas, de la ciudadanía y de las empresas está generando un nuevo modelo de gestión pública conocido como gobernanza inteligente. La exploración de la incidencia del análisis de datos masivos en las administraciones públicas y la exposición de los canales a través de los cuales la ciudadanía y las empresas pueden colaborar activamente con las administraciones permiten un nuevo modelo de gobernanza.

La recolección de información pública para hacer uso de ella de manera inteligente, por medio de la tecnología, con el fin de analizarla y tomar decisiones. El uso de información pública por medio de minería de datos, big data, data cleaning, entre otros, permite realizar análisis, el cual brinda una ventaja competitiva.

El estudio predicción de series de tiempo económicas y financieras: ARIMA vs LSTM llevado a cabo por Siami y Namin (2018) en el que realizaron una investigación experimental con el fin de comparar la precisión de los métodos de análisis predictivos basados en series de tiempo contra métodos de Deep learning, determinar si el uso de métodos de Deep learning como LSTM muestra una ventaja de precisión y cantidad de errores en relación con el método ARIMA.

La volatilidad del mercado dificulta la realización de forecasting para economía y finanzas. Debido a esto, lograr un forecast preciso se hace necesario, para lo cual es importante recurrir a varios métodos, entre ellos los métodos de pronóstico utilizando análisis de regresión los cuales tienen muchas limitaciones en su aplicación. Por lo cual es necesario investigar cual método de forecasting

ofrece las mejores predicciones respecto a su precisión y menor cantidad de errores.

El método ARIMA es uno de los métodos más populares para forecasting por el cual se buscará comparar su precisión contra métodos de deep learning como LSTM los cuales han ganado atención en los últimos años con sus aplicaciones en muchas disciplinas incluyendo finanzas.

Comparar que existe ventaja al utilizar métodos de Deep learning contra métodos de series de tiempo en el análisis de información para forecasting. El uso de métodos alternativos utilizando técnicas avanzadas de machine learning como lo son los algoritmos de Deep learning provee beneficios para el análisis de problemas de predicción económica.

La investigación experimental Pronóstico de precios de productos de comercio electrónico mediante la combinación de un modelo de promedio móvil integrado autorregresivo (ARIMA) y datos de tendencias de Google (Carta, Medda, Pili, Reforgiato y Saia, 2019). Profundiza en el análisis predictivo de precios de productos utilizando métodos de series de tiempo como ARIMA. Utilizar datos recolectados de internet por medio de crawlers.

El uso de crawlers para recolectar información de productos y sus cambios de precios a través del tiempo, así mismo recolectar información de redes sociales y Google trends con el fin de analizar las variaciones en el tiempo de información, utilizando un modelo de series de tiempo ARIMA para poder predecir precios futuros.

La recolección de datos de internet, tratamiento de datos, análisis de datos y utilización de modelos estadísticos pueden ser utilizados con la finalidad de

predecir resultados. Es posible utilizar información recolectada de internet y clasificarla para poder analizarla y predecir valores futuros sin necesidad de utilizar machine learning, aplicando únicamente series de tiempo en su lugar.

El análisis de la aplicación del aprendizaje automático para pronosticar las tendencias del comercio internacional como investigación experimental (Batarseh, Gopinath, Nalluru y Beckman, 2019). Utiliza técnicas de machine learning para realizar mejores predicciones de mercado, por medio del análisis de información pública. Analizar como variables como el PIB afectan el mercado.

El análisis de la información pública por medio de técnicas de machine learning predictivas por medio de modelos como ARIMA, GBoosting, XGBoosting y LightGBM para predecir patrones futuros de comercio y K-Means clustering de países de acuerdo con factores económicos. Analizar cómo afectan las políticas externas al mercado por ejemplo la inclusión de nuevas tarifas y la detención de tratados comerciales.

Para diseñar un modelo predictivo se tiene que contar con información histórica de variables económicas y diferentes costos, que serán interpretados por técnicas de análisis estadístico, por lo que el uso de técnicas de machine learning para el análisis de datos y variables económicas permite determinar el comportamiento de los mercados.

Nikou, Mansourfar y Bagherzadeh (2019). Predicción del precio de las acciones mediante el algoritmo de aprendizaje DEEP y su comparación con los algoritmos de aprendizaje automático, es una investigación comparativa que utiliza múltiples técnicas de machine learning para pronóstico de acciones en combinación, para analizar si existe una ventaja comparada al uso de estas por separado.

La inversión en acciones es una de las mayores actividades de inversión, sin embargo, el desconocimiento de este debido a la carencia información y conocimiento, puede causar la perdida de dicha inversión. Esto hace necesaria la predicción del precio de las acciones de manera precisa, para poder reducir el desperdicio de recursos.

Comúnmente se hace uso de técnicas de machine learning como artificial neural network (ANN), las cuales se ha confirmado su efectividad para el forecast de precio de acciones, al igual que support vector machine (SVM), sin embargo el uso de ANN por sí solo dificulta la comprensión, debido a esto con el fin de realizar una predicción más exitosa y precisa, se analiza la combinación de diferentes métodos predictivos. Se analiza la comparación de métodos de machine learning tradicionales como SVM contra métodos de Deep learning como Long short-term memory (LSTM). Esto hace más fácil el análisis de la información. La combinación de estos métodos apunta a hacer una mejor predicción que utilizar cada uno separado.

Usar distintos métodos para realizar forecasting de análisis económico como es el mercado de acciones, la utilización de modelos de Deep learning proporcionan un aumento en la precisión del forecasting permitiendo un mayor apoyo en la toma de decisiones de inversión.

Calof y Viviers (2020) Análisis de Big Data y selección de mercados internacionales: un estudio exploratorio. Explora la selección de mercados internacionales por medio del análisis de información disponible de comercio internacional y el impacto análisis de big-data en las decisiones de selección de mercado internacional por medio de TRADE-DSM (Modelo de soporte de decisión).

Una gran cantidad de información del flujo de comercio internacional y mercados potenciales está disponible, sin embargo muchos exportadores no saben cómo identificarla, con precisión adecuada, estos mercados mantienen el potencial más grande. Incluso si ellos tienen acceso a información relevante, el gran volumen de información hace complicado el proceso de análisis de forma frecuente, además de ser costoso y tardado. Un reto adicional es que la mayoría de los exportadores realizan el proceso de forma empírica, por lo que carecen de una metodología de toma de decisiones apropiada, lo cual no les permite adoptar una aproximación sistemática para elegir mercados extranjeros.

El uso de métodos estadísticos puede ser aplicado a grandes colecciones de datos con el fin de analizar mercados, por lo cual la utilización de métodos estadísticos para el análisis de información pasada permite probar hipótesis o plantear métodos de predicción del estado del mercado.

Darapaneni, Reddy, Paduri, Acharya, y Nithin (2020). Pronóstico de COVID-19 en India utilizando el modelo ARIMA. Establece un modelo predictivo para detectar posibles puntos de concentración de COVID-19, utilizando modelos basados en estadística como lo es ARIMA.

El estado de pandemia COVID-19 ha provocado el cierre de los países, esto ha tenido efectos positivos para evitar la propagación de este, sin embargo esto no es sostenible a largo plazo debido a problemas económicos que esto genera más en países, en vías de desarrollo como India. Para detectar posibles puntos de concentración para el COVID-19, se hace necesario realizar forecasting el cual puede guiar a las administraciones a tomar decisiones. El uso de modelos basados en estadística como lo es ARIMA puede ayudar a realizar predicciones precisas, ya que tiene un rendimiento más prometedor que métodos como el support vector machine y wavelet neural network.

El forecasting ayuda a la toma de decisiones de manera precisa, haciendo uso de métodos basados en estadística, los métodos como ARIMA permiten determinar patrones de comportamiento y predecir resultados.

Karingula *et al.* (2021). Incorporaciones mejoradas para el pronóstico de series de tiempo, investigación demostrativa. Combina lo mejor de los métodos estadísticos tradicionales y redes neuronales por medio del uso de AR-Net. La cual es una red que es interpretable como un modelo clásico AR, pero también escala a dependencias de largo alcance.

Los métodos de forecasting juegan un gran rol en la toma de decisiones de muchos negocios, particularmente relacionados con problemas de predicción de data financiera, los cuales pueden ser modelados como un problema de series de tiempo. Los métodos de series de tiempo pueden ser modelados utilizando varios métodos como SARIMA, ARIMA, siendo este uno de los más utilizados para forecasting.

Es posible realizar forecasting con métodos de machine learning como Recurrent neural networks (RNNs) o principalmente short term memory (LSTM), los cuales han tenido éxito en forecasting.

Usar métodos de series de tiempo en conjunto con métodos de machine learning como gradient boosting permite realizar análisis predictivo con un mejor desempeño. Esto abre la posibilidad de realizar análisis multi variable para forecasting combinando dichos métodos.

El estudio experimental previsión del comercio agrícola internacional mediante aprendizaje automático (Gopinath, Batarseh, Beckman, Kulkarni y Jeong, 2021). Busca predecir los patrones de comportamiento del comercio

agricultor, por medio de la utilización de métodos de machine learning para periodos de tiempo extensos con gran precisión.

En los últimos años la agricultura ha tenido muchos acontecimientos mundiales tales como desastres naturales, guerras de comercio, pandemias. Esto ha generado incertidumbre, la cual ha afectado la toma de decisiones. Dichas tendencias no convencionales necesitan grandes cantidades de datos. El análisis de estos datos requiere de técnicas que vayan más allá de los modelos lineales, por consiguiente, el uso de machine learning se ha visto como una alternativa para estos retos, el uso de modelos de machine learning para determinar forecast de comercio internacional se hace una necesidad.

Usar métodos de machine learning para la predicción del comportamiento de comercio internacional, lo cual permite realizar mejor planeamiento del presupuesto basado en los flujos de comercio.

### **3. PLANTEAMIENTO DEL PROBLEMA**

#### **3.1. Contexto general**

El problema que se enfrentan las empresas que se dedican a la comercialización de tubería de acero proveniente de China, es que no cuentan con pronósticos sobre el comportamiento de importaciones de tubería de acero que le proporcionen ventajas a la hora de realizar negociaciones.

#### **3.2. Descripción del problema**

En la actualidad no se cuenta con pronósticos del comportamiento de importaciones de tubería de acero proveniente de China, el desconocimiento de métodos de predicción y variables económicas dificulta el realizar pronósticos de montos de importaciones. La carencia de un modelo de predicción (forecasting) para el comportamiento de importación de tubería de acero proveniente de China limita el panorama general del estado del mercado de tubería de acero, lo que dificulta la toma de decisiones.

#### **3.3. Formulación del problema**

Pregunta central

- ¿De qué forma se puede predecir el comportamiento de importaciones de tubería de acero proveniente de China hacia Guatemala?



#### Preguntas auxiliares

- ¿Qué métodos se pueden emplear para análisis predictivo del comportamiento de importaciones de tubería de acero proveniente de China hacia Guatemala?
- ¿Qué variables económicas influyen principalmente en el comportamiento de importaciones de tubería de acero proveniente de China hacia Guatemala?
- ¿Contribuye el uso de machine learning al análisis predictivo del comportamiento de importaciones de tubería de acero proveniente de China hacia Guatemala?

#### **3.4. Delimitación del problema**

Se analizará información histórica comprendida en los años 2013 al 2019 de montos de importaciones de tubería de acero proveniente de China hacia Guatemala brindada por la SAT, indicadores económicos brindadas por el banco de Guatemala y por el banco mundial. El contexto que bajo el cual se realizará el estudio es de *data science*.

## 4. JUSTIFICACIÓN

La realización de la presente investigación se justifica en la línea de investigación en ciencia de datos, la cual combina múltiples disciplinas de conocimiento que son base de estudio de la maestría en ingeniería para la industria con especialidad en ciencias de la computación.

En la actualidad no se cuenta con pronósticos del comportamiento de importaciones de tubería de acero proveniente de China, por lo cual hacer uso de la ciencia de datos puede ayudar a proveer información adecuada para detectar tendencias de importación.

Se plantea la propuesta de un modelo de pronóstico del comportamiento de importaciones de tubería de acero proveniente de China.

El conocer la proyección de importaciones de tubería de acero proveniente de china puede ayudar a las empresas que se dedican al mercado de comercialización de tubería, porque les permitirá contar con información confiable para la toma de decisiones empresariales, dando solución a problemas e interrogantes presentes en la sociedad por medio de la ciencia de datos con la finalidad de brindar solución a dichos problemas de una manera rápida y automatizada.



## 5. OBJETIVOS

### General

Diseñar un modelo que pueda predecir el comportamiento de importaciones de tubería de acero proveniente de China hacia Guatemala.

### Específicos

- Identificar los métodos factibles para el análisis predictivo del comportamiento de importaciones de tubería de acero proveniente de China hacia Guatemala.
- Identificar que variables económicas influyen principalmente el comportamiento de importaciones de tubería de acero proveniente de China hacia Guatemala.
- Determinar cómo contribuye el uso de machine learning al análisis predictivo del comportamiento de importaciones de tubería de acero proveniente de China hacia Guatemala.



## **6. NECESIDADES POR CUBRIR Y ESQUEMA DE LA SOLUCIÓN**

Propuesta de modelo para predictivo del comportamiento de importaciones de tubería de acero proveniente de China para analizar ventajas competitivas.

En la actualidad no se cuenta con pronósticos del comportamiento de importaciones de tubería de acero proveniente de China, el desconocimiento de métodos de predicción y variables económicas dificulta el realizar pronósticos de montos de importaciones.

La carencia de un modelo de predicción (forecasting) para el comportamiento de importación de tubería de acero proveniente de China limita el panorama general del estado del mercado de tubería de acero, lo que dificulta la toma de decisiones. El uso de la data science ayudará a proveer información adecuada para detectar tendencias de importación.

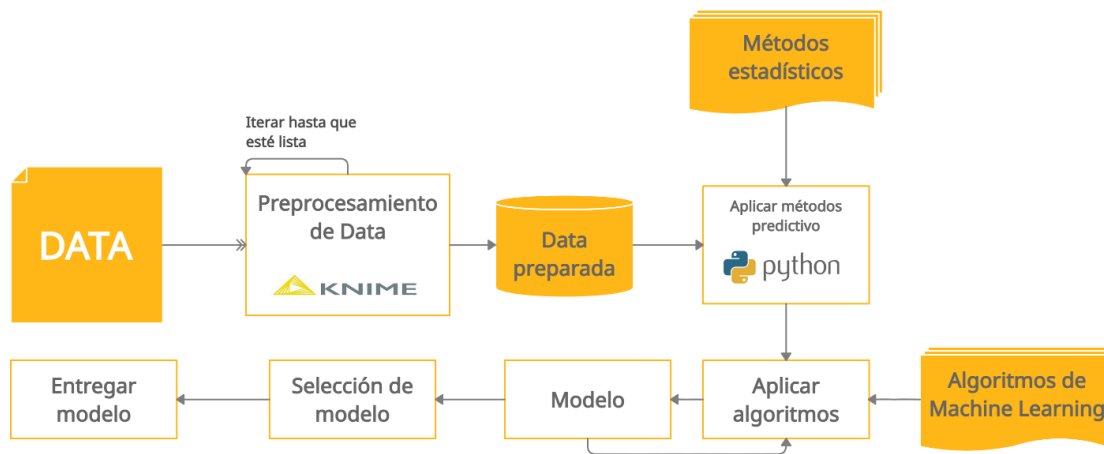
La investigación propuesta pretende crear un modelo de forecasting para el comportamiento de importaciones de tubería de acero proveniente de China.

El diseño del modelo utilizará datos históricos de la SAT sobre importaciones, también se recolectará información de variables económicas del banco de Guatemala y del banco mundial, así como información pública relacionada al acero. Se analizará, reformateará y limpiará la información recolectada con la herramienta KNIME, para su posterior análisis en Python, en donde se hará uso de series de tiempo como ARIMA para analizar datos

históricos de montos de importaciones de tubería de acero proveniente de china, con el fin de conocer si al modelo generado se le puede mejorar su desempeño por medio de técnicas de machine learning como gradient boosting.

Estos estudios se llevarán a cabo en un entorno de desarrollo local, haciendo uso de sets de entrenamiento, test y validación.

Figura 1. Diagrama de arquitectura



Fuente: elaboración propia, utilizando draw.io

El modelo propuesto brindará la posibilidad de realizar análisis de ventajas competitivas a las empresas que se encuentren en el ramo de comercio de tubería de acero.

## **7. MARCO TEÓRICO**

### **7.1. Variables económicas**

El crecimiento económico es un factor importante para el desarrollo de un país, estos conceptos están relacionados a las variables económicas ya que si una se ve afectada las otras también. Las variables económicas recolectan información agregada de una región en específico las cuales nos permiten analizar el comportamiento económico.

#### **7.1.1. Producto interno bruto**

El producto interno bruto (PIB) contabiliza todo el producto económico generado dentro de las fronteras de la nación, este es una medida bruta ya que no hace ningún ajuste por la depreciación de los activos (Coyle, 2014). Este está medido durante un periodo definido de tiempo que generalmente es de un año.

#### **7.1.2. Tipo de cambio**

El precio al que una moneda cambia por otra se le llama tipo de cambio, las fluctuaciones del tipo de cambio ocurren porque las fluctuaciones de la demanda y oferta de la moneda no son independientes (Parking, 2007).

#### **7.1.3. Inflación**

La inflación es un incremento a nivel general en los precios de productos durante un periodo de tiempo (Comley, 2013).



#### **7.1.4. Balanza de pagos**

Ningún país es capaz de producir todos los satisfactorios de su población dentro de sus fronteras, estos pueden ser por circunstancias como clima, nivel de mar, fuentes de recursos naturales por mencionar algunas (Toro, 2003). Esto hace necesario realizar comercio con otras naciones con el fin de deshacerse de productos que produce en excedente o adquirir nuevos productos ya sea que no los produzca o los produzca en cantidades limitadas. Estos intercambios de compra y venta se registran en lo que conocemos como balanza de pagos la cual es utilizado para análisis económico.

### **7.2. Estadística**

Es una ciencia derivada de la matemática encargada de recoger, analizar, interpretar y representar los datos (Cruz, 2016).

#### **7.2.1. Estadística Descriptiva**

Busca describir los datos de manera reducida y ordenada haciendo uso de distintos métodos (Cruz, 2016).

#### **7.2.2. Estadística Inferencial**

El objetivo de esta es llegar a conclusiones de un conjunto de datos grande, haciendo uso de observaciones a una pequeña porción de este (Cruz, 2016).

### **7.2.3. P-Value**

Es la probabilidad de tener un valor, del estadístico calculado, más extremo que el observado (Millones, Barreno y Vásquez, 2017). Este sirve para rechazar o aceptar una hipótesis nula.

## **7.3. Pronósticos**

Los pronósticos están basados en el principio de que hechos y ocurrencias actuales vienen en cierta forma condicionados por hechos y ocurrencias pasados (Pando, 1981). Esto quiere decir que para poder estimar el futuro se debe de conocer el pasado.

### **7.3.1. Métodos de pronóstico**

Existen distintos métodos para pronostico estandarizados, estos difieren el uno del otro en términos de precisión del pronóstico, nivel de satisfacción cuantitativa y en la base lógica de la cual se deriva el pronóstico (Ballou, 2004).

### **7.3.2. Métodos cualitativos**

Los métodos cualitativos utilizan el juicio, la intuición, las encuestas o técnicas comparativas para generar estimados cuantitativos acerca del futuro. Debido a que este método carece de naturaleza científica se dificulta la estandarización y validación de su precisión (Ballou, 2004). No obstante el uso de estos métodos se torna necesario cuando se desea pronosticar el éxito de productos nuevos en múltiples ámbitos.

### **7.3.3. Métodos de proyección histórica**

Se utilizan cuando se cuenta con información histórica en una cantidad razonable y las variaciones de tendencia y estacionalidad de las series de tiempo son estables y bien definidos (Ballou, 2004). Estas se basan en la proposición de que el futuro se puede deducir del pasado.

### **7.3.4. Métodos causales**

Estos se basan en la premisa de que el nivel de la variable a pronosticar se puede derivar del nivel de las otras variables relacionadas.

## **7.4. Series de tiempo**

Una serie de tiempo es una serie de observaciones de un evento o proceso en intervalos iguales de tiempo. Es llamada serie de tiempo, porque las observaciones son usualmente tomadas con respecto al tiempo (Ky y Vu, 2007).

### **7.4.1. Prueba Dickey Fuller**

El objetivo básico de esta prueba es examinar la hipótesis nula que:

$$\theta = 1 \text{ en } y_t = \theta y_{t-1} + u_t$$

Contra la alternativa de un lado  $\theta < 1$ . Por lo tanto las hipótesis de interés son:

H<sub>0</sub>: La serie contiene raíz unitaria.

H<sub>1</sub>: La serie es estacionaria.

La prueba de Dickey Fuller se pueden llevar a cabo permitiendo una intersección o una intersección y una tendencia determinística (Brooks, 2008).

## **7.5. Análisis predictivo con series de tiempo**

El objetivo de del análisis de series de tiempo para realizar pronóstico es estudiar las técnicas para analizar las inferencias generadas por las series de tiempo, para esto se debe seleccionar un método adecuado el cual permitirá estimar parámetros, revisar la bondad de ajuste de los datos y la posibilidad de encontrar un modelo ajustado (Brockwell, 2010).

### **7.5.1. Método ARMA**

El modelo autorregresivo de media móvil (ARMA por sus siglas en inglés) es un modelo de forecasting en el cual los métodos de análisis de autorregresión (AR) y el de media móvil (MA) son aplicados a datos de series de tiempo que se comportan bien.

En este método se asume que la serie de tiempo es estacionaria esto quiere decir que cuando fluctúa lo hace de manera uniforme en un tiempo en particular (Gordon, 2021).

Este modelo de series de tiempo trabaja utilizando mecanismos de extrapolación, en la descripción de las series de tiempo., esto basado en la ley cambiante de la serie de tiempo. Este método se utiliza principalmente en la investigación de mercado, es un método adecuado para una gran variedad de problemas.

El método ARMA tiene una precisión y rendimiento superior que utilizar los métodos AR y MA por separado, pero puede ser complicada su implementación.

El método ARMA se basa en la siguiente ecuación:

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \Theta_i \varepsilon_{t-i}$$

En donde:

- $\phi$  = Los parámetros autorregresivo del modelo
- $\theta$  = Los parámetros de la media móvil
- $c$  = una constante
- $\Sigma$  = sumatoria
- $\varepsilon$  = error

### 7.5.2. Método ARIMA

El método autorregresivo integrado de media móvil (ARIMA por sus siglas en inglés) es un modelo estadístico utilizado para forecasting en series de tiempo. El cual se utiliza comprender el pasado de la información o poder predecir el futuro de una serie de tiempo.

El método ARIMA realiza predicciones sobre una serie de tiempo basada en sus valores en el pasado. Puede ser utilizada para para cualquier serie de números no estacional media vez estas muestren un patrón y no es una serie de eventos aleatorios.

ARIMA es la unión de tres modelos AR que es el modelo autorregresivo, I es la diferencia entre observaciones no estacionales y por último MA que es la media móvil este utiliza los valores de los tres modelos.

El método ARIMA está definido por:

AR: Auto Regressive

$$Y_t = \beta_1 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p}$$

Donde:

- $Y_t$  = es una observacion en el tiempo t que depende de  $Y_{t-1}$ ,  $Y_{t-2}$  ...,  $Y_{t-p}$
- $P$ = Son las observaciones previas

I: Integrated

$$\begin{aligned} Z_t &= Y_{t+1} - Y_t && \dots d = 1 \\ Q_t &= Z_{t+1} - Z_t && \dots d = 2 \end{aligned}$$

Donde:

- Se puede operarar las funciones de diferenciacion multiples niveles dependiendo del parametro d que fijemos para el entrenamiento del modelo.

MA: Moving Average

$$Y_t = \beta_2 + \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} + \dots + \omega_q \varepsilon_{t-q} + \varepsilon_t$$

Donde:

- $\varepsilon$  = representa el error residual de las funciones de agregación
- $q$ = es un hiper parámetro que es igual a  $p$ , pero en la ventana de tiempo  $q$ .

### 7.5.3. Método SARIMA

El método autorregresivo integrado de media móvil estacional (SARIMA por sus siglas en inglés) es una extensión del método ARIMA el cual incorpora el componente estacional ya que el método ARIMA no lo soporta.

Este método introduce el orden de estacionalidad autorregresivo  $P$ , el orden de diferencia estacional  $D$ , el orden de la media móvil estacional  $Q$  y el número de pasos en un periodo estacional  $m$  al método ARIMA.

#### **7.5.4. Método de suavizado exponencial**

El método de suavizado exponencial es un método de forecasting en el cual las combinaciones de las observaciones pasadas tienen un peso, en el cual las observaciones recientes tienen más peso que las observaciones más antiguas. El nombre de suavizado exponencial refleja que los pesos decrecen exponencialmente mientras la observación se hace más vieja (Hyndman, 2008).

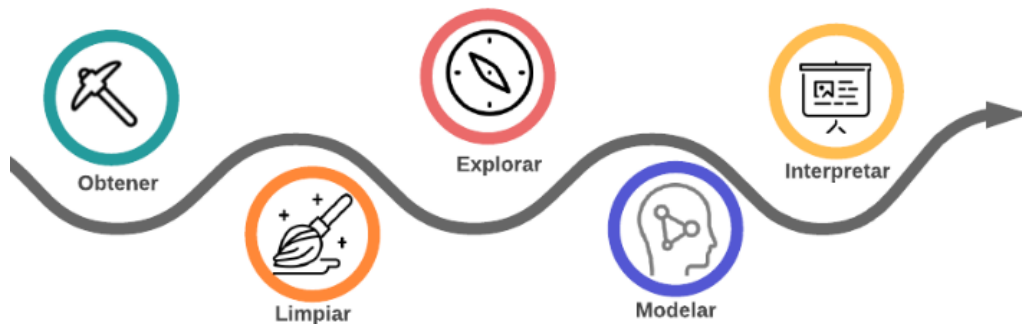
#### **7.6. Ciencia de datos**

La ciencia de datos se basa en principios y técnicas requeridas para realizar actividades de descubrimiento, las cuales son definidas típicamente en términos de secuencias de pasos, llamados flujo de trabajo (workflow o pipeline) (Braschler, 2019).

##### **7.6.1. Flujo de proceso de ciencia de datos**

Según Manson H. el flujo de proceso en ciencia de datos se puede reducir a 5 pasos los cuales son: obtener la información (obtaining data), limpiar la información (scrubbing data), explorar la información (exploring data), modelar la información (modeling data) e interpretar la información (interpreting data) estos pasos forman el acrónimo del modelo (OSEM, 2010).

Figura 2. Imagen flujo de proceso de ciencia de datos



Fuente: elaboración propia, basado en Mason y Wiggins, (2010). *Descripciones de flujo de proceso de ciencia de datos.*

#### 7.6.1.1. Obtener

El primer paso es obtener la información, para esto puede ser necesario:

- Descargar información (privada o pública)
- Consultar bases de datos o API
- Extraer desde archivos pudiendo ser digitales o físicos
- Generarla

#### 7.6.1.2. Limpiar la información

Es común que la información recolectada contenga errores, inconsistencias, datos vacíos, datos no relevantes o caracteres extraños. Esto hace necesario realizar un proceso de limpieza antes de realizar cualquier actividad con la información la cual puede llevar procesos como:

- Filtrar líneas
- Extraer columnas



- Reemplazar valores
- Extraer valores
- Manipular valores faltantes y duplicados
- Transformar datos de un formato a otro

#### **7.6.1.3. Explorar la información**

Una vez se cuenta con la información limpia se procede a explorarla. El fin de este paso es entender la información con la que se cuenta por lo que se puede hacer uso de:

- Ver la información
- Realizar análisis estadístico de la información
- Crear visualizaciones para comprender el comportamiento

#### **7.6.1.4. Modelar la información**

Para poder explicar la información o poder realizar predicciones, es necesaria la realización de un modelo estadístico. Dicho modelo puede ser realizado con distintas técnicas como:

- Clusters
- Clasificación
- Regresiones
- Reducciones dimensionales

#### **7.6.1.5. Interpretar la información**

Este es el paso final en el cual se le da un significado a la información generada por el modelo, este paso involucra:

- Crear conclusiones de los datos
- Evaluar el significado de los resultados
- Comunicar los resultados

### **7.7. Inteligencia artificial**

Según Nilson (1998) la inteligencia artificial (AI por sus siglas en inglés) está interesada en el comportamiento inteligente de los artefactos. El comportamiento inteligente, a su vez, involucra percepción, razonamiento, aprendizaje, comunicación, y actuar en ambientes complejos.

La AI busca realizar actividades lo más parecido a los seres humanos o si es posible mejor así mismo también busca comprender el comportamiento de los humanos, máquinas y los animales. Esto le da a la AI componentes, tanto de ciencia, como de ingeniería.

### **7.8. Machine Learning**

Según Bonaccorso (2017) la principal meta de machine learning es estudiar, diseñar y mejorar modelos matemáticos que pueden ser entrenados con información relacionada, con el fin de inferir el futuro y tomar decisiones sin conocer completamente todos los elementos que influyen.

En otras palabras ML busca tomar información recolectada por medio de un agente y realizar acciones en base a la meta que se busca, adopta un método de aprendizaje estadístico y trata de determinar la distribución de probabilidad y llegar a un resultado con la menor cantidad de errores posibles por medio de los datos procesados con el modelo.

### **7.8.1. Supervised learning**

El aprendizaje supervisado (supervised learning) hace uso de un supervisor, el cual su principal tarea es proveer al agente con una medición precisa de su error. Esto quiere decir que se usan datos debidamente etiquetados con los que se alimenta el modelo, ajusta debidamente su peso hasta que el modelo se haya ajustado correctamente, esto con un proceso de validación cruzada.

### **7.8.2. Unsupervised learning**

El aprendizaje no supervisado (unsupervised learning) a diferencia del aprendizaje supervisado usa algoritmos de machine learning para analizar y agrupar sets de datos no etiquetados. Dichos algoritmos encuentran los patrones existentes o agrupamientos de datos sin necesidad de intervención humana para clasificarlos.

### **7.8.3. Reinforcement learning**

El aprendizaje reforzado (reinforcement learning) se basa en retroalimentación, aunque no haya supervisores la toma de su entorno. Sin embargo esta retroalimentación es mas de tipo cualitativa y no ayuda al agente a determinar una medición más precisa de los errores (Bonaccorso, 2017).

Este método aprende por medio de prueba y error para llegar a la solución del problema suministrado. Para llegar a dicho cometido se basa en un sistema de recompensas en la que si desempeña bien una acción se premia y si desempeña mal se penaliza.

#### **7.8.4. Métodos de pronóstico de machine learning**

Machine learning nos brinda múltiples métodos para inferencia o pronóstico de información esto lo puede lograr con distintos algoritmos extrapolando reglas generales y comprender su estructura con una precisión alta (Bonaccorso, 2017).

##### **7.8.4.1. XGBoost**

XGBoost es uno de los algoritmos dominantes en machine learning aplicado, este es una implementación de árboles de decisiones de aumento de gradiente (gradient boosting) diseñado con enfoque en desempeño y velocidad (Brownlee, 2021).

Este método se utiliza ya que tiene una alta velocidad de ejecución comparada a otros métodos de gradient boosting y un alto desempeño en modelos de predicción con set de datos tabulares y estructurados por medio de regresiones.

#### **7.9. Herramientas**

A continuación se describen las herramientas que se utilizarán para llevar a cabo la realización del estudio.

##### **7.9.1. Python**

Python es un lenguaje de programación interpretado multiplataforma, intuitivo, definido por funciones con tipos de datos compuestos el cual es fácil y rápido de aprender. (Python, 2021).

### **7.9.2. NumPy**

NumPy es un proyecto open source creado en 2005, el cual permite realizar cálculos numéricos en Python (Numpy, 2021).

### **7.9.3. Pandas**

Pandas es una librería open source, bajo licencia BSD, la cual provee estructuras de datos de alto desempeño fáciles de usar y herramientas de análisis de datos para Python (Pandas, 2021).

### **7.9.4. Statsmodels**

Statsmodels es un módulo de Python el cual provee las herramientas necesarias para poder realizar estimaciones en diferentes modelos estadísticos, realizar pruebas estadísticas y exploración de datos. (Seabold y Perktold, 2010).

### **7.9.5. Scikit-learn**

Scikit-learn es un proyecto creado en el año 2007 como parte del proyecto de código de Google Summer. Es considerada la librería de machine learning en Python más útil y robusta (Pedragosa, 2011).

### **7.9.6. KNIME**

KNIME es un software de código abierto para análisis de datos, reportería y una plataforma de integración. El atractivo de KNIME es que integra varios componentes de machine learning y minería de datos sumado a su concepto de

diseño por medio de bloques lo hace una herramienta muy potente y fácil de usar (KNIME, 2021).

### **7.9.7. Jupyter Notebook**

Es un proyecto sin ánimos de lucro y de código abierto nacido en 2014 como una herramienta interactiva para ciencia de datos y computo científico (Jupyter, 2021).

### **7.9.8. Visual Studio Code**

Es un editor de código gratuito el cual permite iniciar a programar rápidamente soportando múltiples lenguajes de programación, sin necesidad de cambiar de editor (Visual studio code,2021).



## 8. PROPUESTA DE ÍNDICE DE CONTENIDOS

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES

LISTA DE SÍMBOLOS

GLOSARIO

RESUMEN

PLANTAMIENTO DEL PROBLEMA Y FORMULACIÓN DE PREGUNTAS  
ORIENTADORAS

OBJETIVOS

RESUMEN DE MARCO METODOLÓGICO

INTRODUCCIÓN

1. ANTECEDENTES
2. JUSTIFICACIÓN
3. ALCANCES
4. MARCO TEÓRICO
  - 4.1. Variables económicas
    - 4.1.1. Producto interno bruto
    - 4.1.2. Tipo de cambio
    - 4.1.3. Inflación
    - 4.1.4. Balanza de pagos
  - 4.2. Estadística
    - 4.2.1. Estadística Descriptiva



- 4.2.2. Estadística Inferencial
- 4.2.3. P-Value
- 4.3. Pronósticos
  - 4.3.1. Métodos de pronóstico
    - 4.3.1.1. Métodos cualitativos
    - 4.3.1.2. Métodos de proyección histórica
    - 4.3.1.3. Métodos causales
- 4.4. Series de tiempo
  - 4.4.1. Prueba Dickey Fuller
- 4.5. Análisis predictivo con series de tiempo
  - 4.5.1. Método ARMA
  - 4.5.2. Método ARIMA
  - 4.5.3. Método SARIMA
  - 4.5.4. Método de suavizado exponencial
- 4.6. Ciencia de datos
  - 4.6.1. Flujo de proceso de Data Science
    - 4.6.1.1. Obtener la información
    - 4.6.1.2. Limpiar la información
    - 4.6.1.3. Explorar la información
    - 4.6.1.4. Modelar la información
    - 4.6.1.5. Interpretar la información
- 4.7. Inteligencia Artificial
- 4.8. Machine Learning
  - 4.8.1. Supervised learning
  - 4.8.2. Unsupervised learning
  - 4.8.3. Reinforcement learning
  - 4.8.4. Métodos de pronóstico de machine learning
    - 4.8.4.1. XGBoost
- 4.9. Herramientas

- 4.9.1. Python
- 4.9.2. NumPy
- 4.9.3. Pandas
- 4.9.4. Statmodels
- 4.9.5. Scikit-learn
- 4.9.6. KNIME
- 4.9.7. Jupyter Notebook
- 4.9.8. Visual Studio Code

## 5. DESARROLLO DE MODELO

- 5.1. Diseño del modelo de pronóstico
  - 5.1.1. Diagrama de bloques de arquitectura
  - 5.1.2. Diagrama de flujo
  - 5.1.3. Diagrama de componentes
  - 5.1.4. Diagrama de arquitectura
  - 5.1.5. Diagrama de actividades
- 5.2. Preparación de la información
  - 5.2.1. Obtener información
    - 5.2.1.1. Recolección de datos de importaciones
    - 5.2.1.2. Recolección de variables económicas
    - 5.2.1.3. Recolección de datos internacionales
    - 5.2.1.4. Transformar los datos a formato CSV
  - 5.2.2. Configuración de ambientes de desarrollo
    - 5.2.2.1. Instalar herramientas necesarias
  - 5.2.3. Limpiar información
    - 5.2.3.1. Filtrar líneas
    - 5.2.3.2. Extraer columnas
    - 5.2.3.3. Reemplazar valores
    - 5.2.3.4. Extraer valores

- 5.2.3.5. Manipular valores faltantes y duplicados
    - 5.2.3.6. Transformación de datos
    - 5.2.3.7. Unir datos
    - 5.2.3.8. Agregar dimensión de tiempo
  - 5.3. Análisis de información
    - 5.3.1. Explorar información
    - 5.3.2. Revisar datos
    - 5.3.3. Serializar la información
    - 5.3.4. Pruebas de Dickey Fuller
    - 5.3.5. Pruebas de Ljung Box
    - 5.3.6. Resumen estadístico
  - 5.4. Modelar la información con series de tiempo
    - 5.4.1. Realizar modelo de regresión
    - 5.4.2. Refinar modelo
    - 5.4.3. Realizar pruebas al modelo
    - 5.4.4. Validar modelo
  - 5.5. Modelar la información con machine learning
    - 5.5.1. Etiquetar información
    - 5.5.2. Realizar modelo de ML
    - 5.5.3. Revisar Errores
    - 5.5.4. Validación de modelo
- 6. INTERPRETACIÓN DE RESULTADOS
  - 6.1. Creación de conclusiones
  - 6.2. Evaluar el significado de resultados
  - 6.3. Comunicar los resultados
- 7. DISCUSIÓN DE RESULTADOS
  - 7.1. Selección de modelo de pronóstico adecuado

CONCLUSIONES  
RECOMENDACIONES  
REFERENCIAS  
APÉNDICES  
ANEXOS



## **9. METODOLOGÍA**

### **9.1. Características del estudio**

El enfoque del estudio propuesto es de carácter cuantitativo, ya que se analizarán datos de importaciones por medio de métodos estadísticos y matemáticos, con el fin de predecir el comportamiento que tendrán estos datos en el futuro.

El alcance es diseño de tendencia, dado que se desea interpretar datos, con el fin de descubrir el comportamiento que tendrán las importaciones en el futuro, utilizando múltiples técnicas de análisis.

El diseño adoptado será no experimental, ya que la información se analizará con el fin de realizar la propuesta de modelo para pronóstico del comportamiento de importaciones de tubería de acero proveniente de china para analizar ventajas competitivas; además será de tipo longitudinal pues se estudiará la información en un intervalo de tiempo predefinido, dado que se analizarán variables económicas con la finalidad de determinar el impacto que tienen sobre el comportamiento de importaciones de tubería.

### **9.2. Unidades de análisis**

La población en estudio serán las importaciones de tubería de acero proveniente de China comprendida en los años 2013 a 2019, que serán estudiadas en su totalidad.

### 9.3. Variables

Las variables en estudio se describen a continuación:

Tabla I. **Tipos de variables**

<b>Variable</b>	<b>Definición teórica</b>	<b>Definición operativa</b>	<b>Indicador</b>
Importación de tubería anual	Monto de tubería importada proveniente de China	Monto de importaciones por año iniciando en 2013 y terminando en 2019	Monto de tubería importada proveniente de china en los años 2013 - 2019
VARIABLES ECONÓMICAS	VARIABLES QUE INFLUYEN EN COMPORTAMIENTO DE LAS IMPORTACIONES DE TUBERÍA DE ACERO.	Recolección de datos históricos de variables económicas por año iniciando en 2013 y terminando en 2019.	PIB, tipo de cambio, inflación, balanza de pagos de Guatemala en el periodo 2013-2019.

Continuación de la tabla I.

Métodos de análisis predictivo	Métodos de análisis estadístico que permitirán predecir el comportamiento de patrones.	Aplicabilidad de acuerdo con el comportamiento de la muestra.	Aplicable No Aplicable
Métodos de machine learning	Métodos de machine learning utilizados para predecir patrones en el comportamiento de los mercados económicos.	Facilidad de implementación y aplicabilidad según los datos.	Aplicable No Aplicable

Fuente: elaboración propia.

#### **9.4. Fases del estudio**

A continuación, se detallan las fases que se llevarán a cabo para desarrollar el trabajo de investigación.

##### **9.4.1. Fase 1. Revisión de literatura**

Contiene la exposición y análisis de enfoques teóricos y teorías conceptuales utilizadas para fundamentar la investigación, haciendo uso las referencias bibliográficas, artículos científicos y tesis.

El presente trabajo de investigación busca la creación de un modelo predictivo, utilizando métodos estadísticos y machine learning, por lo que se hace necesario adquirir conocimiento por medio análisis documental respecto a los siguientes temas:



- Limpieza y transformación de datos
- Métodos estadísticos para análisis predictivo con series de tiempo
- Flujo de datos para modelo de predictivo
- Funcionamiento de machine learning
- Análisis predictivo con métodos de machine learning
- Herramientas para implementación de machine learning y series de tiempo

#### **9.4.2. Fase 2. Diseño del modelo de pronóstico**

Se realiza el análisis de la solución y se representa por medio de diagramas para comprender claramente lo planteado.

#### **9.4.3. Fase 3. Gestión o recolección de la información**

La información se recolectará de información pública recolectada de la SAT sobre importaciones de tubería proveniente de China comprendida en los años 2013 a 2019.

#### **9.4.4. Fase 4. Configuración de ambientes**

Se instalarán todas las herramientas necesarias para poder realizar el desarrollo del modelo.

#### **9.4.5. Fase 5. Análisis de información**

La información recolectada será analizada por medio de diferentes técnicas y herramientas como las siguientes:

- Limpieza y transformación de datos por medio de KNIME: se limpiará la información recolectada y será transformada a la forma deseada con el fin de que sea más fácil su análisis.
- Análisis estadístico por medio de NumPy y Pandas en Python: se realizará un análisis estadístico de los datos transformados con el fin de detectar patrones en el comportamiento de estos.
- Se utilizarán gráficas para visualizar el comportamiento haciendo uso de las librerías Matplotlib y Plotly para Python.

#### **9.4.6. Fase 6. Modelar la información con series de tiempo**

Se realizará la propuesta de un modelo por medio de métodos de series de tiempo.

- Utilización de métodos predictivos por medio de *statmodels* en Python: Se hará uso de librerías como *statmodels* para realizar análisis predictivo por medio de series de tiempo.

#### **9.4.7. Fase 7. Modelar la información con machine learning**

Se realizará el modelo de series de tiempo y se analizaran optimizaciones por métodos de machine learning.

- Utilización de métodos de machine learning con scikit-learn en Python: se utilizará scikit-learn para implementar métodos de machine learning para *forecasting*.

#### **9.4.8. Fase 8. Interpretación de información**

Elaboración de conclusiones, evaluación y comunicación de los resultados obtenidos.

#### **9.4.9. Fase 9. Selección de modelo de pronóstico adecuado**

Se analizará si el uso de machine learning como un método de pronóstico, representa una ventaja para la eficiente predicción del comportamiento de las importaciones de tubería de acero.

## **10. TÉCNICAS DE ANÁLISIS DE INFORMACIÓN**

Para analizar la información en este estudio se utilizarán métodos de estadística descriptiva e inferencial, esto con el fin de verificar la validez de los datos obtenidos.

### **10.1. Preparación de la información**

Una vez se hayan recolectado los datos de las fuentes necesarias según la sección 9.4.3, se procederá a preparar la información para su posterior análisis, para esto se realizarán los siguientes procesos:

- Convertir los datos recolectados a formato CSV si no estuviesen ya en este.
- Importar los archivos CSV a un flujo de trabajo en KNIME.
- Tratar datos nulos, remover columnas innecesarias y corregir formato de datos.
- Agregar columna para la dimensional de tiempo, la cual podrá estar dada en meses, semanas o días.
- Se exportarán los data sets procesados para su posterior análisis en Python.

### **10.2. Análisis de información**

Para realizar el análisis de la información se utilizará estadística descriptiva e inferencial según la sección 9.4.5, utilizando técnicas como:

- Prueba Dickey Fuller para detectar la estacionalidad en los datos esto se hará haciendo uso de la librería statsmodels de Python.
- Prueba Ljung Box para verificar la correlación de los datos esto se hará haciendo uso de la librería statsmodels de Python.
- Resumen estadístico para resumir los datos analizados con el fin de generar conclusiones esto se hará haciendo uso de la librería Pandas para Python.
- Series de tiempo para generar un modelo el cual pueda predecir el comportamiento de la tubería.
- Transformaciones de series para estabilizar la serie si esta no fuese estacionaria por medio de uno de los siguientes métodos:
  - Transformaciones de estabilización de varianza: por medio de logaritmo natural o raíz cuadrada.
  - Transformaciones de estabilización del nivel: por medio de diferenciación simple o diferenciación estacional.
- Técnicas de potenciación del gradiente las cuales se utilizarán para el análisis de la regresión esto con el fin de generar un modelo de tipo predictivo. Se hará por medio de la librería scikit-learn para Python.
- Se utilizarán gráficas para analizar el comportamiento de los datos a lo largo del análisis estas se realizarán con las librerías Matplotlib y Plotly para Python.

### **10.3. Presentación de los resultados**

























En esta etapa del análisis, se elaborarán gráficas para visualizar el comportamiento del modelo creado, así como tablas en las cuales se mostrará la precisión alcanzada.

# 11. CRONOGRAMA

Tabla II. Cronograma de actividades

Id	Modo de tarea	Nombre de tarea	Duración	Comienzo	Fin	Predecesoras
1		<b>Trabajo de investigación</b>	<b>37.2 sem.</b>	<b>lun 17-01-22</b>	<b>lun 03-10-22</b>	
2		<b>Revisión de avances</b>	<b>20.2 sem.</b>	<b>vie 21-01-22</b>	<b>vie 10-06-22</b>	
3		Revisión de avances 1	0.2 sem.	vie 21-01-22	vie 21-01-22	
4		Revisión de avances 2	0.2 sem.	vie 28-01-22	vie 28-01-22	
5		Revisión de avances 3	0.2 sem.	vie 04-02-22	vie 04-02-22	
6		Revisión de avances 4	0.2 sem.	vie 11-02-22	vie 11-02-22	
7		Revisión de avances 5	0.2 sem.	vie 18-02-22	vie 18-02-22	
8		Revisión de avances 6	0.2 sem.	vie 25-02-22	vie 25-02-22	
9		Revisión de avances 7	0.2 sem.	vie 04-03-22	vie 04-03-22	
10		Revisión de avances 8	0.2 sem.	vie 11-03-22	vie 11-03-22	
11		Revisión de avances 9	0.2 sem.	vie 18-03-22	vie 18-03-22	
12		Revisión de avances 10	0.2 sem.	vie 25-03-22	vie 25-03-22	
13		Revisión de avances 11	0.2 sem.	vie 01-04-22	vie 01-04-22	
14		Revisión de avances 12	0.2 sem.	vie 08-04-22	vie 08-04-22	
15		Revisión de avances 13	0.2 sem.	vie 15-04-22	vie 15-04-22	
16		Revisión de avances 14	0.2 sem.	vie 22-04-22	vie 22-04-22	
17		Revisión de avances 15	0.2 sem.	vie 29-04-22	vie 29-04-22	
18		Revisión de avances 16	0.2 sem.	vie 06-05-22	vie 06-05-22	
19		Revisión de avances 17	0.2 sem.	vie 13-05-22	vie 13-05-22	
20		Revisión de avances 18	0.2 sem.	vie 20-05-22	vie 20-05-22	
21		Revisión de avances 19	0.2 sem.	vie 27-05-22	vie 27-05-22	
22		Revisión de avances 20	0.2 sem.	vie 03-06-22	vie 03-06-22	
23		Revisión de avances 21	0.2 sem.	vie 10-06-22	vie 10-06-22	
24		<b>Diagramas de solución</b>	<b>2 sem.</b>	<b>lun 17-01-22</b>	<b>vie 28-01-22</b>	
25		Diagrama de flujo	0.5 sem.	lun 17-01-22	mié 19-01-22	
26		Diagrama de componentes	0.5 sem.	mié 19-01-22	vie 21-01-22	25
27		Diagrama de arquitectura	0.5 sem.	lun 24-01-22	mié 26-01-22	26
28		Diagrama de actividades	0.5 sem.	mié 26-01-22	vie 28-01-22	27
29		<b>Preparación de la información</b>	<b>12.2 sem.</b>	<b>lun 31-01-22</b>	<b>lun 25-04-22</b>	<b>28</b>
30		<b>Obtener información</b>	<b>3.2 sem.</b>	<b>lun 31-01-22</b>	<b>lun 21-02-22</b>	
31		Recolección de datos de importaciones	1 sem.	lun 31-01-22	vie 04-02-22	28
32		Recolección de variables económicas	1 sem.	lun 07-02-22	vie 11-02-22	31
33		Recolección de datos internacionales	1 sem.	lun 14-02-22	vie 18-02-22	32
34		Transformar los datos a formato CSV	0.2 sem.	lun 21-02-22	lun 21-02-22	33
35		<b>Configurar ambiente de desarrollo</b>	<b>1 sem.</b>	<b>mar 22-02-22</b>	<b>lun 28-02-22</b>	<b>30</b>
36		Instalar herramientas necesarias	1 sem.	mar 22-02-22	lun 28-02-22	34
37		<b>Limpiar la información</b>	<b>8 sem.</b>	<b>mar 01-03-22</b>	<b>lun 25-04-22</b>	<b>35</b>
38		Filtrar líneas	1 sem.	mar 01-03-22	lun 07-03-22	36
39		Extraer columnas	1 sem.	mar 08-03-22	lun 14-03-22	38
40		Reemplazar valores	1 sem.	mar 15-03-22	lun 21-03-22	39
41		Extraer valores	1 sem.	mar 22-03-22	lun 28-03-22	40
42		Manipular valores faltantes y duplicados	1 sem.	mar 29-03-22	lun 04-04-22	41
43		Transformar formato de datos	1 sem.	mar 05-04-22	lun 11-04-22	42
44		Unir datos	1 sem.	mar 12-04-22	lun 18-04-22	43
45		Agregar dimensión de tiempo	1 sem.	mar 19-04-22	lun 25-04-22	44
46		<b>Análisis de información</b>	<b>23 sem.</b>	<b>mar 26-04-22</b>	<b>lun 03-10-22</b>	<b>37</b>
47		<b>Explorar información</b>	<b>2.3 sem.</b>	<b>mar 26-04-22</b>	<b>mié 11-05-22</b>	
48		Revisar datos	0.5 sem.	mar 26-04-22	jue 28-04-22	45

Continuación de la tabla II.

Id		Modo de tarea	Nombre de tarea	Duración	Comienzo	Fin	Predecesoras
49			Serializar la información	1 sem	jue 28-04-22	jue 05-05-22	48
50			Pruebas de Dickey Fuller	0.2 sem.	jue 05-05-22	vie 06-05-22	49
51			Pruebas de Ljung Box	0.3 sem.	vie 06-05-22	lun 09-05-22	50
52			Resumen estadístico	0.3 sem.	mar 10-05-22	mié 11-05-22	51
53			<b>Modelar la información con series de</b>	<b>6.2 sem.</b>	<b>mié 11-05-22</b>	<b>jue 23-06-22</b>	<b>47</b>
54			Realizar modelo de regresión	4 sem.	mié 11-05-22	mié 08-06-22	51
55			Refinar modelo	1 sem	mié 08-06-22	mié 15-06-22	54
56			Realizar pruebas al modelo	1 sem	mié 15-06-22	mié 22-06-22	55
57			Validar modelo	0.2 sem.	mié 22-06-22	jue 23-06-22	56
58			<b>Revisión parcial</b>	<b>1 sem</b>	<b>jue 23-06-22</b>	<b>jue 30-06-22</b>	<b>53</b>
59			Revisión de avances	1 sem	jue 23-06-22	jue 30-06-22	57
60			<b>Modelar la información con machine</b>	<b>9 sem.</b>	<b>jue 30-06-22</b>	<b>jue 01-09-22</b>	<b>58</b>
61			Etiquetar información	1 sem	jue 30-06-22	jue 07-07-22	59
62			Realizar modelo de ML	8 sem.	jue 07-07-22	jue 01-09-22	61
63			Revisar errores	8 sem.	jue 07-07-22	jue 01-09-22	61
64			Validación de modelo	8 sem.	jue 07-07-22	jue 01-09-22	61
65			<b>Interpretación de resultados</b>	<b>1.5 sem.</b>	<b>jue 01-09-22</b>	<b>lun 12-09-22</b>	<b>60</b>
66			Creación de conclusiones	0.5 sem.	jue 01-09-22	lun 05-09-22	62,63,64
67			Evaluar el significado de resultados	0.5 sem.	mar 06-09-22	jue 08-09-22	66
68			Comunicar los resultados	0.5 sem.	jue 08-09-22	lun 12-09-22	67
69			Creación de informe final	2 sem.	mar 13-09-22	lun 26-09-22	68
70			Revisión Final	1 sem	mar 27-09-22	lun 03-10-22	69
71			Aprobación	0 sem.	lun 03-10-22	lun 03-10-22	70

Fuente: elaboración propia.

## **12. FACTIBILIDAD DEL ESTUDIO**

Para la realización e implementación de la propuesta de estudio se cuenta con los siguientes recursos humanos, tecnológicos y económicos.

### **12.1. Recurso humano necesario**

- 1 estudiante de la maestra de ingeniería para la industria con especialidad en ciencias de la computación.
- 1 maestro encargado que supervisará y guiará el presente estudio.

### **12.2. Acceso a la información**

Para la elaboración de este estudio se hará uso de información disponible públicamente en internet, tanto en fuentes nacionales, como internacionales.

### **12.3. Recursos tecnológicos**

- Características de Hardware y software para desarrollo
  - Sistema operativo Windows 10
  - Procesador AMD Ryzen 7 de 16 núcleos
  - 64Gb de Ram
  - 2Tb de almacenamiento SSD
  - Tarjeta de video nvidia RTX 3070 8Gb de Vram
- Infraestructura de red
  - Red doméstica de 100MB



- Software necesario
  - KNIME edición gratuita
  - Python
  - Jupyter Notebook
  - Visual Studio Code

#### 12.4. Tiempo necesario para el desarrollo de la solución

- El estudio está planificado para llevarse a cabo de enero 2022 a noviembre 2022.
- Se hará uso de 2h en horario no laborable.

Las librerías y software para utilizar serán open source o versiones gratuitas.

#### 12.5. Recursos financieros

Los recursos financieros se describen a continuación:

Tabla III. Recursos financieros

Actividad	Costo unitario	Costo proyecto	Observaciones
Costo por recursos tecnológicos	Q800	Q7200	Costos asociados al uso de internet
Costo por software	Q0	Q0	Se utilizará software open source o versiones gratuitas

Continuación de la tabla III.

Costos de energía eléctrica	Q200	Q1800	Costo por consumo de energía eléctrica del equipo de computo
Costos de desarrollo	Q0	Q0	El costo de desarrollo será asumido por mi persona
<b>Total proyecto</b>	<b>Q9000</b>		
<b>Costo de inversión</b>	<b>Q9000</b>		

Fuente: elaboración propia

Se concluye en esta fase que el presupuesto es acorde para el desarrollo del estudio.



## REFERENCIAS

1. Ballou, R. (2004). *Logística: Administración de la cadena de suministro*. Estado de México, México: Pearson Educación.
2. Batarseh, F.; Gopinath, M.; Nalluru, G. y Beckman, J. (octubre, 2019). *Application of machine learning in forecasting international trade trends*. [Mensaje en un blog]. Recuperado de [https://www.researchgate.net/publication/336361589\\_Application\\_of\\_Machine\\_Learning\\_in\\_Forecasting\\_International\\_Trade\\_Trends](https://www.researchgate.net/publication/336361589_Application_of_Machine_Learning_in_Forecasting_International_Trade_Trends)
3. Bonaccorso, G. (2017). *Machine Learning Algorithms*. Birmingham, Reino Unido: Packt Publishing Ltd.
4. Braschler, M. (2019). *Applied Data Science: Lessons Learned for the Data-Driven Business*. Winterthur, Suiza: Springer.
5. Brockwell, P. y Davis, R. (2006). *Introduction to Time Series and Forecasting*. New York, USA: Springer Science y Business Media.
6. Brooks, C. (2008). *Introductory Econometrics for Finance*. New York, USA: Cambridge University Press.
7. Brownlee, J. (2021). *XGBoost With Python: Gradient Boosted Trees with XGBoost and scikit-learn*. USA: Machine Learning Mastery.

8. Calof, J. y Viviers, W. (junio, 2020). Big data analytics and international market selection: An exploratory study. *Journal of Intelligence Studies in Business*, 10(2).1-14. Recuperado de <https://journal.hh.se/index.php/JISIB/article/view/581>
9. Carta, S.; Medda, A.; Pili, A.; Reforgiato, D. y Saia, R. (septiembre 2019). Forecasting e-commerce products prices by combining an autoregressive integrated moving average (ARIMA) model and google trends data. *Future Internet*, 11(1), 1-19. Recuperado de <https://www.mdpi.com/1999-5903/11/1/5>
10. Cerrillo, A. (octubre, 2018). Datos masivos y datos abiertos para una gobernanza inteligente. *El profesional de la información (EPI)*, 27(5), 1-8. Recuperado de <http://profesionaldelainformacion.com/contenidos/2018/sep/16.html>
11. Comley, P. (2013). *Inflation Tax: The Plan to Deal with the Debts*. Reino Unido: Pete Comley.
12. Coyle, D. (2017). *El producto interno bruto: Una historia breve pero entrañable*. Ciudad de México, México: Fondo de Cultura Económica.
13. Cruz, E. (2016). *Estadística básica: Introducción a la estadística con R*. Bogotá, Colombia: Ediciones de la U.
14. Darapaneni, N.; Reddy, D.; Paduri, R.; Acharya, P. y Nithin, S. (octubre, 2020). *Forecasting of COVID-19 in India using ARIMA model*. [Mensaje en un blog]. Recuperado de <https://ieeexplore.ieee.org/abstract/document/9298045>

15. Gopinath, M.; Batarseh, F. A.; Beckman, J.; Kulkarni, A. y Jeong, S. (enero, 2021). International agricultural trade forecasting using machine learning. *Data & Policy*. 3, 1-20. Recuperado de <https://www.cambridge.org/core/journals/data-and-policy/article/international-agricultural-trade-forecasting-using-machine-learning/0AB17DA0D2BD6E72274232656BA29B91>
16. Gordon, J. (julio, 2021). *Autoregressive Moving Average (ARMA)*. Explained. Recuperado de [https://thebusinessprofessor.com/en\\_US/research-analysis-decision-science/autoregressive-moving-average-arma-definition](https://thebusinessprofessor.com/en_US/research-analysis-decision-science/autoregressive-moving-average-arma-definition)
17. Hyndman, R.; Koehler, A.; Ord, K. y Snyder, R. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Washington DC, USA: Springer Science y Business Media.
18. Jupyter. (Mayo 2021). *JupyterLab: A Next-Generation Notebook Interface*. [Mensaje en un blog]. Recuperado de <https://jupyter.org/>
19. Karingula, S.; Ramanan, N.; Tahsambi, R.; Amjadi, M.; Jung, D; Si, R. y Coelho, N. (2021). Boosted Embeddings. *Time Series Forecasting*, 1-13. Recuperado de <https://arxiv.org/abs/2104.04781>
20. KNIME. (2021). *End to End Data Science*. KNIME. Recuperado de [www.knime.com](http://www.knime.com)
21. Ky, M. (2007). *The ARIMA and VARIMA Time Series: Their Modelings, Analyses and Applications*. Canada: AuLac Technologies Inc.

22. Mason, H. y Wiggins, C. (octubre 2010). A Taxonomy of Data Science. [Mensaje en un blog]. Recuperado de <http://www.dataists.com/2010/09/a-taxonomy-of-data-science/>
23. Millones, R.; Barreno, E.; Vásquez, F. y Castillo, C. (2017). *Estadística aplicada a la ingeniería y los negocios*. Lima, Perú: Fondo editorial Universidad de Lima.
24. Nikou, M.; Mansourfar, G. y Bagherzadeh, J. (diciembre, 2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, 26(4), 164-174. Recuperado de <https://onlinelibrary.wiley.com/doi/abs/10.1002/isaf.1459>
25. Nilsson, N. (1998). *Artificial Intelligence: A New Synthesis*. USA: Elsevier
26. NumPy. (2021). *The fundamental package for scientific computing with Python*. NumPy. Recuperado de <https://numpy.org/>
27. Ohyver, M.; Pudjihastuti, H. (2018). Arima model for forecasting the price of medium quality rice to anticipate price fluctuations. *Procedia Computer Science*, 135, 707-711. Recuperado de <https://www.sciencedirect.com/science/article/pii/S1877050918315072>
28. Pandas. (2021). *Documentation*. Pandas. Recuperado de <https://pandas.pydata.org/>

29. Pando, J. (1981). *Fundamentos de Los Pronósticos de Mercado*. Costa Rica: Bib. Orton IICA / CATIE.
30. Parking, D.; Esquivel, G. y Muñoz, M. (2007). *Macroeconomía: versión para Latinoamérica*. Estado de México, México: Pearson Educación.
31. Pedregosa, et al., (2011). *Scikit-learn: Machine Learning in Python*. *JMLR* 12, 2825-2830. [Mensaje en un blog]. Recuperado de <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
32. Perktold, J.; Seabold, S. (enero, 2010). Statsmodels. *Econometric and Statistical Modeling with Python*, 1-6. Recuperado de [https://www.researchgate.net/publication/264891066\\_Statsmodels\\_Econometric\\_and\\_Statistical\\_Modeling\\_with\\_Python](https://www.researchgate.net/publication/264891066_Statsmodels_Econometric_and_Statistical_Modeling_with_Python)
33. Python. (2021). *Compound Data Types*. Python. Recuperado de <https://www.python.org/>
34. Siami-Namini, S. y Namin, A. S. (marzo, 2018). Forecasting economics and financial time series. *ARIMA vs. LSTM*, 1-19. Recuperado de <https://arxiv.org/abs/1803.06386>
35. Toro, J. (2003). *La balanza de pagos de México: su importancia en el análisis económico*. San Luis Potosí, México: UASLP.
36. Visual studio code. (2021) *Code editing*. Visual Studio. Recuperado de <https://code.visualstudio.com/>



