



**TRABAJO DE GRADO**  
**Opción Seminario-Diplomado.**

**ALGORITMO COMPUTACIONAL PARA EL ANÁLISIS Y TOMA DE  
DECISIONES EN DATOS UTILIZANDO ALGORITMOS DE MACHINE  
LEARNING**

Corporación Universitaria Remington.  
Facultad de Ingeniería  
Ingeniería Industrial

Estudiante:  
José Aníbal Bohórquez Quintero  
Tutor: Juan Carlos Briñez de León

Opción de Trabajo de grado Seminario-Diplomado.  
2025.

## Tabla de contenido

1. Resumen .....	3
2. Palabras claves .....	4
3. Introducción.....	4
4. Marco metodológico.....	5
5. Marco Conceptual.....	6
6. Planteamiento del problema .....	7
7. Formulación del problema.....	8
8. Objetivo general.....	8
9. Objetivos específicos .....	8
10. Descripción de variables .....	9
11. Posibles aplicaciones.....	9
12. Desarrollo del aprendizaje .....	9
13. Procesamiento de datos .....	10
14. Modelación y toma de decisiones .....	21
15. Implementación en contexto real.....	24
16. Discusión de resultados .....	26
17. Conclusiones .....	28
Referencias Bibliográficas .....	29

## 1. Resumen

Este trabajo tiene como propósito central analizar los indicadores macroeconómicos más representativos de Colombia durante el periodo comprendido entre 2005 y 2019, excluyendo de forma intencional los años 2020 a 2022 debido a la distorsión estadística generada por los efectos de la pandemia por COVID-19. A través de técnicas de análisis de datos y herramientas de aprendizaje automático (Machine Learning), se busca identificar patrones, relaciones y comportamientos relevantes en variables como el producto interno bruto (PIB), la inflación (IPC), la tasa de desempleo, la tasa de cambio (TRM) y el salario mínimo legal vigente.

El estudio parte de una rigurosa etapa de limpieza, categorización e interpolación de los datos obtenidos de fuentes oficiales como el DANE y el Banco de la República. Posteriormente, se implementaron algoritmos de clasificación supervisada como K-Nearest Neighbors, Random Forest y Análisis Discriminante, a fin de categorizar la tasa de desempleo según su comportamiento anual. Aunque se identificaron modelos con alto desempeño en precisión, el enfoque principal no se limitó a la predicción, sino a la visualización estructurada de los datos y su valor explicativo frente a fenómenos económicos complejos.

En línea con lo planteado por James, Witten, Hastie y Tibshirani (2021), el uso de Machine Learning en este contexto permitió detectar correlaciones e interacciones entre variables que no son evidentes mediante métodos estadísticos convencionales, lo cual ofrece un enfoque complementario para la toma de decisiones en política económica, planificación gubernamental y gestión financiera empresarial. Este trabajo representa un primer paso hacia el uso de modelos computacionales accesibles y replicables para el monitoreo de la salud

económica del país.

## 2. Palabras claves

Machine Learning, Inflación, Tasa de interés, Desempleo, Aprendizaje automático, Salario, Pib, TRM, Predicción.

## 3. Introducción

En los últimos años, el análisis de datos ha adquirido una creciente relevancia para comprender fenómenos sociales, económicos y políticos. Particularmente en Colombia, el estudio detallado de indicadores macroeconómicos como la inflación, el desempleo, el salario mínimo, el tipo de cambio (TRM) y las tasas de interés ha sido esencial para evaluar el desempeño económico del país y formular políticas públicas orientadas al bienestar social. Comprender cómo han evolucionado estos indicadores permite no solo interpretar el pasado económico, sino también prever tendencias y comportamientos futuros mediante modelos predictivos basados en datos históricos.

En este contexto, las técnicas avanzadas de análisis de datos y el aprendizaje automático (Machine Learning) emergen como herramientas potentes para identificar patrones complejos en grandes volúmenes de información económica, facilitando la creación de modelos predictivos más precisos y útiles para tomar decisiones estratégicas en tiempo real (James, Witten, Hastie & Tibshirani, 2021). Al aplicar estas tecnologías, es posible anticipar fenómenos económicos como fluctuaciones inflacionarias, cambios en el poder adquisitivo y evaluar con mayor precisión los impactos de políticas monetarias en la economía nacional.

Este documento busca estructurar un algoritmo computacional basado en técnicas de análisis de datos y Machine Learning para estudiar en profundidad el comportamiento histórico

de los principales indicadores económicos colombianos, cubriendo un periodo extenso que abarca desde el año 2005 hasta el 2019. La elección de este periodo responde a la necesidad de disponer de un conjunto de datos estable y consistente, evitando los años recientes (2020-2022) debido a su carácter atípico causado por la pandemia global.

El resultado esperado es la generación de modelos predictivos robustos y confiables que permitan interpretar mejor la dinámica económica del país. Dichos modelos tendrán aplicaciones prácticas significativas, especialmente en ámbitos como la formulación de políticas económicas, la planificación financiera y la gestión empresarial. Este enfoque es relevante porque no solo describe situaciones pasadas, sino que permite anticipar condiciones económicas futuras, fortaleciendo la capacidad de respuesta de instituciones y empresas ante situaciones de incertidumbre económica.

#### 4. Marco metodológico

Este documento se desarrolla bajo un enfoque cuantitativo y exploratorio, utilizando métodos estadísticos y de aprendizaje automático para analizar datos económicos históricos. La metodología propuesta se divide claramente en cuatro fases principales:

- **Recolección y limpieza de datos:** Se obtendrán bases de datos oficiales proporcionadas por entidades como el Banco de la República y el DANE. Estos datos serán integrados y organizados en formatos estandarizados, aplicando procesos para manejar valores faltantes o atípicos que puedan afectar el análisis.
- **Análisis exploratorio de datos (EDA):** Se llevará a cabo un estudio inicial mediante visualizaciones gráficas (series de tiempo, histogramas, diagramas de caja, mapas de calor)

y cálculos estadísticos básicos. Esto permitirá identificar claramente tendencias generales, patrones de estacionalidad y relaciones entre las diferentes variables estudiadas.

- Selección y entrenamiento de modelos: Se implementarán técnicas de Machine Learning como regresión lineal, regresión múltiple, K-Nearest Neighbors (KNN) y árboles de decisión. Los modelos serán entrenados sobre un conjunto específico de datos (train/test split) y se evaluarán mediante métodos sólidos de validación, como la validación cruzada, para asegurar la precisión y confiabilidad de las predicciones.
- Interpretación y visualización de resultados: Finalmente, se presentarán claramente los resultados obtenidos mediante gráficos, tablas y análisis comparativos. Se destacarán las variables más influyentes y se discutirán sus implicaciones para la formulación de políticas económicas y estrategias financieras.

Este enfoque metodológico proporciona una base sólida para comprender y predecir el comportamiento económico reciente de Colombia, utilizando técnicas rigurosas que garantizan la transparencia, precisión y replicabilidad del estudio.

## 5. Marco Conceptual

El análisis de datos es una disciplina clave para procesar grandes cantidades de información, facilitando la identificación de patrones, tendencias y relaciones importantes. En economía, su aplicación se enfoca en entender variables esenciales como la inflación (IPC), tasas de interés, tipo de cambio, desempleo y el salario mínimo, elementos fundamentales para evaluar la salud económica de un país (Gujarati & Porter, 2009).

Por otra parte, el aprendizaje automático o Machine Learning es una rama de la inteligencia artificial que utiliza algoritmos capaces de aprender directamente de los datos disponibles. Esta tecnología es especialmente útil en economía porque permite realizar predicciones precisas sobre variables macroeconómicas, descubrir relaciones que no son evidentes a primera vista, y mejorar significativamente la calidad de las decisiones basadas en datos concretos y empíricos.

Dentro del contexto económico, algunos de los modelos más comunes de Machine Learning incluyen la regresión lineal, el algoritmo K-Nearest Neighbors (KNN), los árboles de decisión y las redes neuronales. Estos modelos ofrecen herramientas efectivas para la predicción y el análisis detallado de series temporales, así como para detectar complejas relaciones entre múltiples variables económicas.

## 6. Planteamiento del problema

Durante las últimas dos décadas, Colombia ha enfrentado variaciones importantes en indicadores macroeconómicos clave, lo cual ha repercutido significativamente en la estabilidad financiera del país, el poder adquisitivo de sus ciudadanos y la efectividad de las políticas públicas implementadas. Fenómenos como la inflación prolongada, altos niveles de desempleo estructural, fluctuaciones en la tasa de cambio y la constante erosión del valor real del salario mínimo han impactado directamente el bienestar económico y social de la población (Banco Mundial, 2020; DANE, 2021).

Aunque existen abundantes datos oficiales e históricos sobre estos indicadores, los métodos tradicionales de análisis económico presentan limitaciones para anticipar futuros escenarios económicos y descubrir relaciones profundas y complejas entre las diversas variables. Por ello, resulta fundamental explorar y aprovechar herramientas tecnológicas avanzadas como el

aprendizaje automático (Machine Learning), capaces de identificar patrones ocultos y generar predicciones más confiables y detalladas (Varian, 2014).

El empleo del Machine Learning en el ámbito económico representa una oportunidad importante para construir modelos predictivos que puedan guiar decisiones más informadas en campos críticos como la política monetaria, la planificación presupuestaria y la educación financiera, mejorando así la capacidad de respuesta ante situaciones económicas adversas (Mullainathan & Spiess, 2017).

## 7. Formulación del problema

¿Cómo identificar patrones relevantes y tratar de predecir el comportamiento de indicadores macroeconómicos como el IPC, la tasa de interés, el desempleo y el poder adquisitivo en Colombia teniendo en cuenta los períodos 2005–2019 mediante algoritmos de Machine Learning?

## 8. Objetivo general

Determinar un modelo computacional para el análisis y predicción de los indicadores macroeconómicos de Colombia entre 2005 y 2019, utilizando algoritmos de aprendizaje automático.

## 9. Objetivos específicos

- Recopilar y limpiar los datos económicos relevantes desde fuentes oficiales.
- Explorar estadísticamente la evolución de los indicadores a lo largo del tiempo.
- Entrenar modelos de Machine Learning para predecir variables como IPC o TRM.
- Analizar la importancia de las variables y su influencia en los resultados del modelo.

## 10. Descripción de variables

- Fecha: año (formato AA)
- IPC: variación mensual/anual del índice de precios al consumidor
- Tasa de interés: tasa de intervención del Banco de la República
- TRM: tipo de cambio representativo del mercado (COP/USD)
- Desempleo: tasa anual promedio
- Salario mínimo: valor mensual en pesos colombianos

## 11. Posibles aplicaciones

- Modelos para predicción del comportamiento inflacionario
- Análisis del impacto de la política monetaria en el poder adquisitivo
- Evaluación del efecto del desempleo sobre el ingreso real
- Herramientas de pronóstico para la planificación económica

Apoyo en diseño de políticas fiscales o monetarias

## 12. Desarrollo del aprendizaje

Con el propósito de implementar herramientas de visualización y análisis de datos relacionados con los principales indicadores macroeconómicos de Colombia en el periodo comprendido entre los años 2005 y 2019, se inició el trabajo con una base de datos construida a partir de fuentes oficiales como el Banco de la República y el DANE. Esta información permitió realizar un estudio sistemático del comportamiento económico nacional, apoyado en técnicas de análisis exploratorio y modelos de predicción.

Las variables seleccionadas para el análisis fueron:

Año: Identifica el periodo cronológico en que se registraron los datos económicos.

IPC: Variación anual del Índice de Precios al Consumidor, indicador clave de la inflación.

Tasa de interés: Tasa de intervención fijada por el Banco de la República.

TRM: Tasa de cambio representativa del mercado (COP/USD).

Desempleo: Tasa anual de desempleo nacional.

Salario mínimo: Valor mensual legal vigente en pesos colombianos.

El análisis se orientó a examinar las fluctuaciones y relaciones entre estos indicadores, con el fin de identificar los periodos de mayor inestabilidad económica. Para ello, se emplearon herramientas gráficas como boxplots y líneas de tiempo, que facilitaron la detección de tendencias, outliers y variaciones significativas en las variables. Todo el procesamiento y visualización de los datos se realizó en Google Colab, entorno especializado para análisis de datos y aprendizaje automático, utilizando código en Python.

Como resultado preliminar, se evidenciaron importantes variaciones en variables como el poder adquisitivo y la TRM, particularmente entre los años 2008 y 2016, lo que coincide con eventos económicos internacionales que impactaron la economía colombiana. Este tipo de análisis sienta las bases para entrenar modelos de predicción y tomar decisiones fundamentadas en evidencia histórica.

### 13. Procesamiento de datos

#### a. Importación de librerías necesarias

```
✓ [1] import pandas as pd
    3s import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Figura 1. Importación de librerías

b. Carga del archivo fuente en formato Excel (xlxs.) o (csv.)

```
#Cargar el archivo
from google.colab import files
uploaded = files.upload()
#Leer el archivo subido
for filename in uploaded.keys():
    df = pd.read_excel(filename)
    print(f'Archivo cargado: {filename}')
    print(df.head())
```

Elegir archivos | IE\_2005-2019.xlsx  
 • IE\_2005-2019.xlsx(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 12233 bytes, last modified: 23/6/2025 - 100% done  
 Saving IE\_2005-2019.xlsx to IE\_2005-2019.xlsx  
 Archivo cargado: IE\_2005-2019.xlsx

fecha	salario	tasa_global_participacion	tasa_ocupacion	tasa_desempleo
0 2005	381500	60.522963	53.383705	7.139258
1 2006	408000	59.124075	52.009604	7.114470
2 2007	433700	58.288733	51.774180	6.514553
3 2008	461500	58.501789	51.913474	6.588315
4 2009	496900	61.294407	53.931531	7.362875

tasa_interes	trm_fin_ano	prom_trm_anual	pib_reportado	exportaciones
0 6.333333	2284.22	2320.77	128713.249983	22031.249995
1 6.583333	2238.79	2357.98	137358.750000	24128.250000
2 8.854167	2014.76	2078.35	146614.250000	25629.000000
3 9.812500	2243.59	1966.26	151428.250000	26151.000000
4 5.583333	2044.23	2156.29	153154.000000	24789.500000

importaciones
0 19677.000024
1 23156.250000
2 26365.250000
3 29664.000000
4 27098.750000

Figura 2. Carga de archivo

Como se observa en la Figura 2. se procedió a cargar el archivo Excel mediante el método `read_excel()` de la biblioteca Pandas, lo cual permitió importar correctamente las 11 columnas correspondientes a los indicadores económicos seleccionados entre los años 2005 y 2019.”

c. Inspeccion de la estructura de datos ( `df.info()` y `df.describe()` )

```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15 entries, 0 to 14
Data columns (total 11 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   fecha                                15 non-null     int64
 1   salario                              15 non-null     int64
 2   tasa_global_participacion            15 non-null     float64
 3   tasa_ocupacion                       15 non-null     float64
 4   tasa_desempleo                       15 non-null     float64
 5   tasa_interes                          15 non-null     float64
 6   trm_fin_ano                           15 non-null     float64
 7   prom_trm_anual                       15 non-null     float64
 8   pib_reportado                        15 non-null     float64
 9   exportaciones                        15 non-null     float64
10  importaciones                        15 non-null     float64
dtypes: float64(9), int64(2)
memory usage: 1.4 KB
None
```

Figura 3. Inspección de estructura de datos

Como se observa en la figura 3. se continua a determinar la información del archivo Excel mediante el método (`df.info()`) de la biblioteca Pandas, lo cual permitió importar correctamente las 11 columnas correspondientes a los indicadores económicos seleccionados entre los años 2005 y 2019 y tener conocimiento acerca del tipo de dato que contienen las variables.

```
print(df.describe())
```

	fecha	salario	tasa_global_participacion	tasa_ocupacion	\
count	15.000000	15.000000	15.000000	15.000000	15.000000
mean	2012.000000	579018.666667	62.583536	56.073959	
std	4.472136	136836.377371	2.373933	2.724180	
min	2005.000000	381500.000000	58.288733	51.774180	
25%	2008.500000	479200.000000	60.908685	53.657618	
50%	2012.000000	566700.000000	63.978082	57.786014	
75%	2015.500000	666902.500000	64.310434	58.226937	
max	2019.000000	828116.000000	64.734170	58.960198	

	tasa_desempleo	tasa_interes	trm_fin_ano	prom_trm_anual	\
count	15.000000	15.000000	15.000000	15.000000	
mean	6.509577	5.536111	2428.724000	2352.233093	
std	0.563550	1.948899	544.449006	508.794420	
min	5.773971	3.166667	1768.230000	1798.230000	
25%	6.017399	4.166667	1978.730000	1932.075000	
50%	6.514553	4.937500	2243.590000	2156.290000	
75%	6.998706	6.458333	2992.355000	2848.810000	
max	7.387679	9.812500	3277.140000	3282.390000	

	pib_reportado	exportaciones	importaciones
count	15.000000	15.000000	15.000000
mean	177147.749999	28608.533333	36839.150002
std	29734.073688	3622.180321	9994.969980
min	128713.249983	22031.249995	19677.000024
25%	152291.125000	25464.875000	28381.375000
50%	177853.750000	29672.500000	39494.250000
75%	203272.625000	31451.125000	45103.125000
max	220306.000000	33432.750000	50562.500000

Figura 4. Descripción de los datos

Mediante el método `df.describe()` observado en la figura 4. se realizó una primera exploración estadística de las variables contenidas en el conjunto de datos. Esta operación permitió verificar que las 11 columnas correspondientes a los indicadores económicos seleccionados fueron importadas correctamente, cubriendo el periodo comprendido entre los años 2005 y 2019.

#### d. Verificación de valores faltantes (`isnull()` y `.sum()`)

```

print(df.isnull())

```

0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
5	False	False	False	False	False	False
6	False	False	False	False	False	False
7	False	False	False	False	False	False
8	False	False	False	False	False	False
9	False	False	False	False	False	False
10	False	False	False	False	False	False
11	False	False	False	False	False	False
12	False	False	False	False	False	False
13	False	False	False	False	False	False
14	False	False	False	False	False	False

0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
5	False	False	False	False	False
6	False	False	False	False	False
7	False	False	False	False	False
8	False	False	False	False	False
9	False	False	False	False	False
10	False	False	False	False	False
11	False	False	False	False	False
12	False	False	False	False	False
13	False	False	False	False	False
14	False	False	False	False	False

Figura 5. Verificación de datos

```

# Resumen de valor nulos por columnas
print(df.isnull().sum())

```

fecha	0
salario	0
tasa_global_participacion	0
tasa_ocupacion	0
tasa_desempleo	0
tasa_interes	0
trm_fin_ano	0
prom_trm_anual	0
pib_reportado	0
exportaciones	0
importaciones	0
dtype: int64	

Mediante los método valores faltantes ( `isnull()` ) y `.sum()` ) observado en la figura 5 y 6 correspondientemente. Por lo tanto, a través de la biblioteca Pandas se identificó si existía posibles valores faltantes en el conjunto de datos. Esta exploración preliminar permitió confirmar que las 11 columnas correspondientes a los indicadores económicos seleccionados fueron importadas correctamente, sin valores nulos, y que la información disponible cubre de forma continua el período comprendido entre los años 2005 y 2019.

e. Imputación de datos faltantes (Interpolación lineal)

```
#Interpolar si es necesario ---- para ajustar datos nulos -----
df['pib_reportado'] = df['pib_reportado'].interpolate(method='linear')
```

Figura 7. Código para interpolación

En la Figura 7, se presenta la aplicación del método `.interpolate(method='linear')` sobre la columna `pib_reportado`. Aunque los resultados del análisis previo (`isnull().sum()`) no evidenciaron valores faltantes, este procedimiento se incluyó como medida de validación para garantizar la continuidad de los datos y prevenir posibles inconsistencias en fases posteriores del análisis. La interpolación lineal permite estimar valores ausentes con base en la tendencia de los datos adyacentes, lo cual resulta útil especialmente en series temporales económicas con comportamiento progresivo

f. Visualización y análisis exploratorio de las variables

- Salario

✓ Salario

```
#Verificacion de las opciones de la variable
print('Analizando los años disponibles')
df['fecha'].unique()

Analizando los años disponibles
array([2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015,
       2016, 2017, 2018, 2019])
```

Figura 8. Código de verificación de variable 'fecha'

En la figura 8. se muestra el procedimiento de verificación de la variable `fecha`, con el fin de asegurar la correcta representación del eje temporal en el análisis. Esta variable actúa como columna clave para establecer la secuencia cronológica de los indicadores económicos. Se confirmó que todos los registros corresponden a los años comprendidos entre 2005 y 2019, sin duplicados ni valores atípicos, lo cual garantiza una base consistente para el estudio de tendencias y comparaciones temporales.

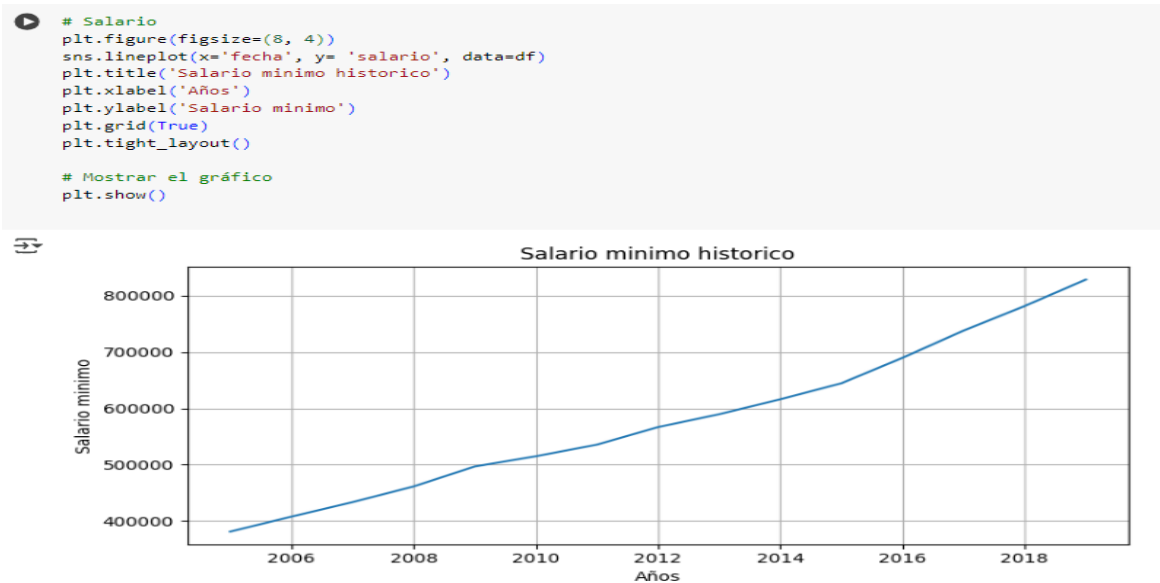


Figura 9. Serie del salario mínimo histórico entre los años 2005-2019

La Figura 9 representa la serie temporal del salario mínimo legal vigente en Colombia entre los años 2005 y 2019. Se evidencia una tendencia creciente sostenida, reflejo de los ajustes anuales realizados por el Gobierno Nacional en función de factores como la inflación, la productividad laboral y el costo de vida. Esta variable resulta fundamental en el análisis, ya que incide directamente sobre el poder adquisitivo de los hogares y se relaciona estrechamente con otros indicadores como el índice de precios al consumidor (IPC) y la tasa de desempleo. La correcta visualización de su evolución permite contextualizar el comportamiento económico del país durante el periodo de estudio.

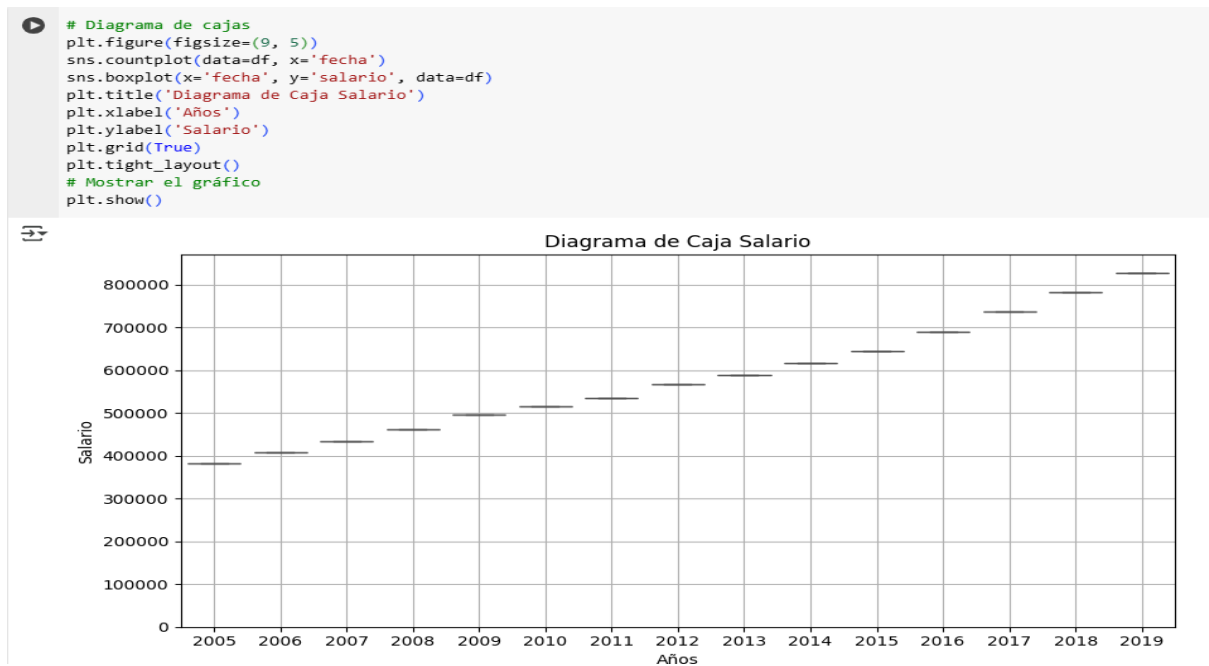


Figura 10. Diagrama de caja de salario vs años.

La Figura 10 presenta un diagrama de caja (boxplot) que relaciona la distribución del salario mínimo en Colombia frente a los años del periodo 2005–2019. Esta visualización permite identificar de forma clara la mediana, los cuartiles y posibles valores atípicos en la evolución del salario. La gráfica confirma la ausencia de datos extremos o desviaciones significativas, lo que respalda la estabilidad de la serie. Además, se aprecia un incremento paulatino y constante del salario a lo largo del tiempo, sin fluctuaciones abruptas. Este tipo de representación es útil para validar la calidad de los datos y analizar la consistencia del crecimiento salarial durante el periodo analizado.

- Otras variables

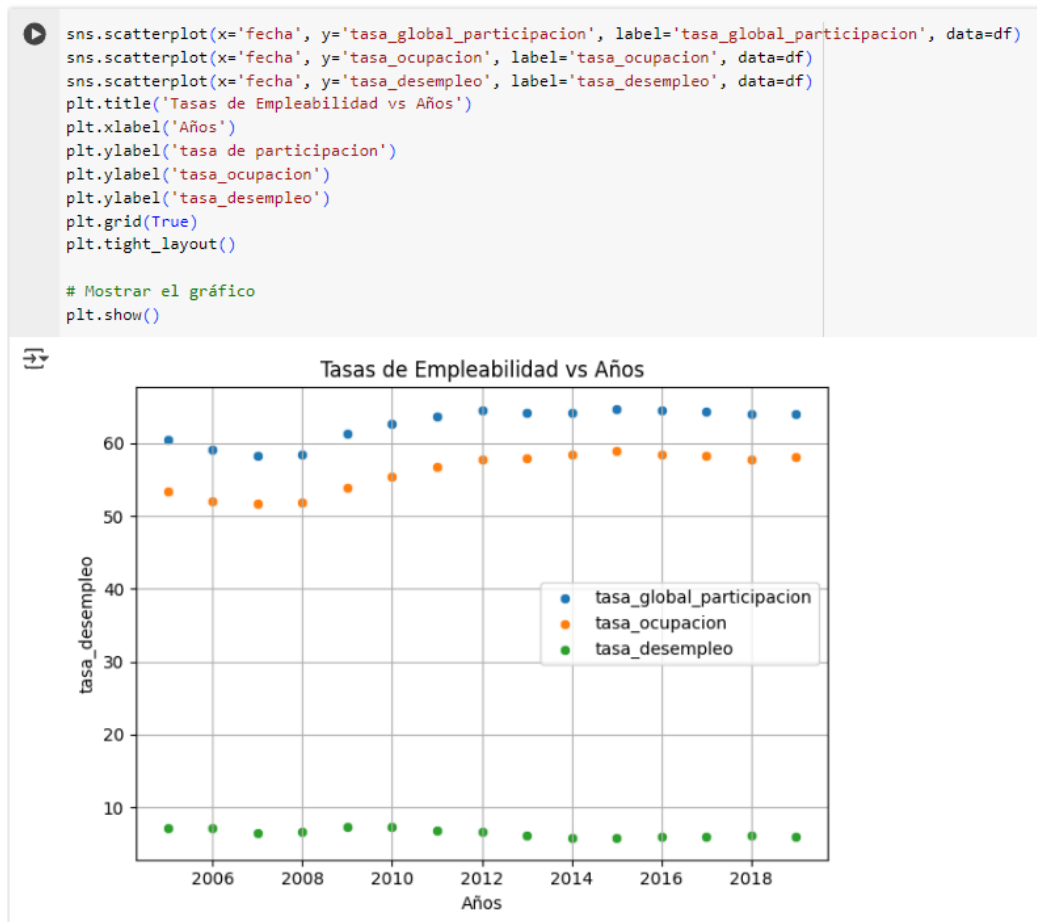


Figura 11. Tasas de empleabilidad en Colombia 2005-2019

En la Figura 11. se representan gráficamente las tasas de empleabilidad en Colombia durante el periodo 2005–2019, específicamente la tasa global de participación, la tasa de ocupación y la tasa de desempleo. A través de un diagrama de dispersión multivariable, se observa que la tasa global de participación y la de ocupación mantienen niveles relativamente estables, mientras que la tasa de desempleo muestra una leve tendencia decreciente en los últimos años del periodo. Esta visualización facilita el análisis comparativo de la evolución laboral del país y permite evidenciar la relación inversa entre ocupación y desempleo, así como el comportamiento constante de la participación laboral.

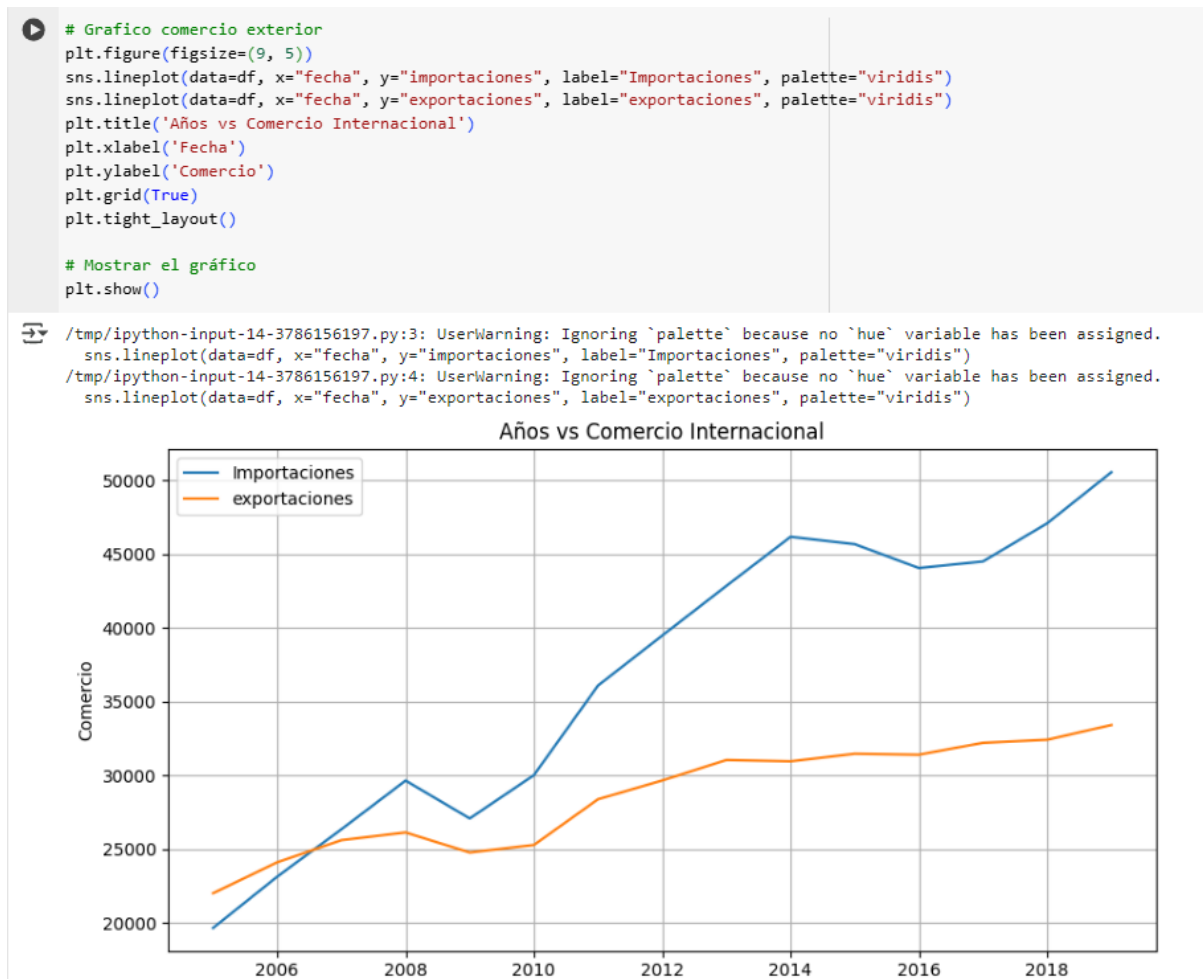


Figura 12. Evolución de las exportaciones e importaciones en Colombia (2005–2019)

En la Figura 12. se presenta un análisis comparativo de las importaciones y exportaciones realizadas en Colombia entre los años 2005 y 2019. A través de una gráfica de líneas, se observa un crecimiento sostenido en ambas variables, con un mayor volumen de importaciones frente a exportaciones durante todo el periodo analizado. Esto sugiere una balanza comercial negativa en la mayoría de los años. No obstante, se destacan periodos de desaceleración en importaciones, especialmente entre 2014 y 2016, posiblemente asociados a factores externos como variaciones en los precios del petróleo o cambios en la política comercial internacional. El comportamiento creciente posterior a 2017 indica una reactivación del comercio exterior nacional.



Figura 13. Evolución de la Tasa Representativa del Mercado (TRM) – 2005 a 2019

La Figura 13. muestra la evolución de la Tasa Representativa del Mercado (TRM) en Colombia, diferenciando entre el valor registrado el 31 de diciembre de cada año y el promedio anual. Se evidencia una marcada volatilidad cambiaria, especialmente entre 2014 y 2016, periodo en el cual se registró una fuerte devaluación del peso colombiano frente al dólar estadounidense. Este comportamiento está estrechamente relacionado con la caída de los precios internacionales del petróleo y la incertidumbre macroeconómica global. La tendencia ascendente posterior a 2016 refleja presiones inflacionarias importadas y cambios en la política monetaria. La TRM es un indicador clave para entender la competitividad externa del país y su impacto en importaciones, exportaciones e inversión extranjera.

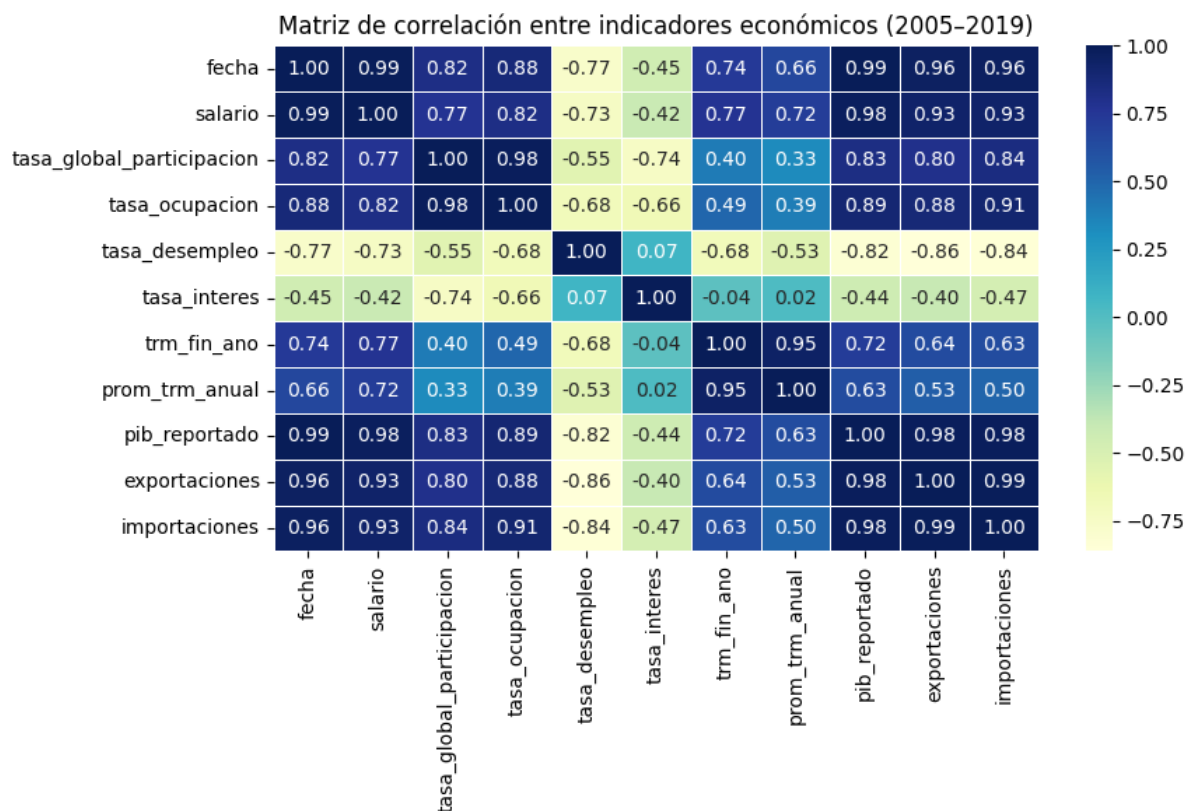


Figura 14. Matriz de correlación entre indicadores económicos (2005–2019)

La Figura 14. presenta la matriz de correlación de Pearson aplicada sobre las variables numéricas del conjunto de datos. Esta herramienta permitió identificar relaciones lineales entre los distintos indicadores económicos analizados. Se observa, por ejemplo, una alta correlación positiva entre el salario y variables como PIB reportado (0.98), importaciones (0.93) y exportaciones (0.93), lo cual sugiere que el aumento del ingreso mínimo legal está asociado al crecimiento del comercio internacional y del producto interno bruto.

Por otro lado, la variable tasa de desempleo exhibe correlaciones negativas con la mayoría de los indicadores, especialmente con la tasa de ocupación (-0.68) y el PIB reportado (-0.82), reflejando que a mayor desempeño económico, menor es el desempleo. Este patrón también se evidencia con la tasa global de participación (-0.55).

Adicionalmente, se destaca una correlación moderada positiva entre la tasa de interés y el desempleo (0.07), aunque su valor es bajo, indicando una relación débil en este período.

En conjunto, esta matriz refuerza la consistencia del conjunto de datos y ofrece una base analítica sólida para el estudio del comportamiento económico del país en el periodo comprendido entre 2005 y 2019.

## 14. Modelación y toma de decisiones

Se inicia definiendo la variable a categorizar, para realizar posteriormente preparar el código de pruebas, importar sklearn y realizar las modelaciones pertinentes.

### ▼ Modelacion

```
df['desempleo_cat'] = pd.qcut(df['tasa_desempleo'], q=3, labels=['Bajo', 'Medio', 'Alto'])
```

```
[49] X = df[['salario', 'pib_reportado', 'tasa_interes', 'trm_fin_ano', 'importaciones', 'exportaciones']]
      Y = df['desempleo_cat']
```

```
import sklearn
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=42)
scaler = MinMaxScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis, QuadraticDiscriminantAnalysis
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier, AdaBoostClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
import warnings
warnings.filterwarnings("ignore")
```

```

# Modelo 0: K-Nearest Neighbors
modelo_0 = KNeighborsClassifier(n_neighbors=5)
modelo_0.fit(X_train, Y_train)
Y_pred_0 = modelo_0.predict(X_test)
print("Accuracy KNN:", accuracy_score(Y_test, Y_pred_0))

# Modelo 1: Naive Bayes
modelo_1 = GaussianNB()
modelo_1.fit(X_train, Y_train)
Y_pred_1 = modelo_1.predict(X_test)
print("Accuracy Naive Bayes:", accuracy_score(Y_test, Y_pred_1))

# Modelo 2: Linear Discriminant Analysis
modelo_2 = LinearDiscriminantAnalysis()
modelo_2.fit(X_train, Y_train)
Y_pred_2 = modelo_2.predict(X_test)
print("Accuracy LDA:", accuracy_score(Y_test, Y_pred_2))

# Modelo 3: Quadratic Discriminant Analysis
modelo_3 = QuadraticDiscriminantAnalysis()
modelo_3.fit(X_train, Y_train)
Y_pred_3 = modelo_3.predict(X_test)
print("Accuracy QDA:", accuracy_score(Y_test, Y_pred_3))

# Modelo 4: Decision Tree
modelo_4 = DecisionTreeClassifier()
modelo_4.fit(X_train, Y_train)
Y_pred_4 = modelo_4.predict(X_test)
print("Accuracy Decision Tree:", accuracy_score(Y_test, Y_pred_4))

# Modelo 5: Support Vector Machine
modelo_5 = SVC()
modelo_5.fit(X_train, Y_train)
Y_pred_5 = modelo_5.predict(X_test)
print("Accuracy SVM:", accuracy_score(Y_test, Y_pred_5))

# Modelo 6: Random Forest
modelo_6 = RandomForestClassifier()
modelo_6.fit(X_train, Y_train)
Y_pred_6 = modelo_6.predict(X_test)
print("Accuracy Random Forest:", accuracy_score(Y_test, Y_pred_6))

# Modelo 7: Logistic Regression
modelo_7 = LogisticRegression(max_iter=1000)
modelo_7.fit(X_train, Y_train)
Y_pred_7 = modelo_7.predict(X_test)
print("Accuracy Logistic Regression:", accuracy_score(Y_test, Y_pred_7))

# Modelo 8: Gradient Boosting
modelo_8 = GradientBoostingClassifier()
modelo_8.fit(X_train, Y_train)
Y_pred_8 = modelo_8.predict(X_test)
print("Accuracy Gradient Boosting:", accuracy_score(Y_test, Y_pred_8))

# Modelo 9: AdaBoost
modelo_9 = AdaBoostClassifier()
modelo_9.fit(X_train, Y_train)
Y_pred_9 = modelo_9.predict(X_test)
print("Accuracy AdaBoost:", accuracy_score(Y_test, Y_pred_9))

```

```

⇒ Accuracy KNN: 1.0
Accuracy Naive Bayes: 0.6666666666666666
Accuracy LDA: 0.6666666666666666
Accuracy QDA: 1.0
Accuracy Decision Tree: 0.6666666666666666
Accuracy SVM: 0.6666666666666666
Accuracy Random Forest: 1.0
Accuracy Logistic Regression: 0.3333333333333333
Accuracy Gradient Boosting: 0.6666666666666666
Accuracy AdaBoost: 0.3333333333333333

```

Figura 15. Benchmark de modelos de clasificación

Tras la ejecución del benchmark con 10 algoritmos de clasificación sobre el conjunto de datos económicos de Colombia (2005–2019), se observaron las siguientes métricas de precisión (accuracy) al clasificar la tasa de desempleo en niveles Bajo, Medio y Alto

#### Modelos de clasificación algorítmico Machine Learning

Modelo	Accuracy (precisión)
K-Nearest Neighbors (KNN)	1.00
Quadratic Discriminant Analysis (QDA)	1.00
Random Forest	1.00
Naive Bayes	0.67
Linear Discriminant Analysis (LDA)	0.67
Decision Tree (Arbol de decisiones)	0.67
Support Vector Machine (SVM)	0.67
Gradient Boosting	0.67
Logistic Regression	0.33
AdaBoost	0.33

Cuadro 1. Resumen de resultados de Benchmarking de modelos de clasificación

KNN, QDA y Random Forest obtuvieron una precisión del 100%. Este comportamiento puede deberse a:

- Bajo número de muestras (solo 15 observaciones de prueba).
- Datos bien separados entre clases después de la normalización.
- Posible sobreajuste, especialmente en Random Forest y QDA, que son modelos propensos a memorizar cuando hay poca información.

Modelos de rendimiento intermedio:

Naive Bayes, SVM, LDA, Decision Tree y Gradient Boosting lograron un 66.6% de precisión. Este valor indica que clasificaron correctamente 2 de cada 3 casos, lo cual es razonable

considerando el tamaño reducido del conjunto de test y las posibles similitudes entre clases (Bajo, Medio, Alto)

Modelos con bajo rendimiento:

- Logistic Regression y AdaBoost apenas alcanzaron un 33.3% de precisión, lo cual indica un comportamiento cercano al azar para una clasificación en tres categorías.
- En el caso de Logistic Regression, es probable que los datos no sean linealmente separables.
- AdaBoost puede haber sufrido de sobreajuste al enfocarse excesivamente en los errores de entrenamiento debido al bajo número de muestras.

Los modelos KNN, QDA y Random Forest demostraron un excelente rendimiento clasificando la tasa de desempleo en niveles categóricos, lo cual sugiere que las variables económicas seleccionadas poseen un poder explicativo significativo. No obstante, es necesario tener precaución al interpretar estos resultados debido al tamaño limitado de la muestra. Se recomienda considerar evaluaciones adicionales con métodos de validación cruzada para reforzar la confiabilidad del modelo.

## 15. Implementación en contexto real

El análisis de los indicadores macroeconómicos de Colombia entre 2005 y 2019 no es solo un ejercicio técnico reservado a economistas o académicos. Por el contrario, este tipo de estudios tiene un enorme potencial para beneficiar a cualquier persona que desee entender cómo los cambios económicos pueden afectar su bolsillo, su empleo o sus decisiones financieras.

A través de técnicas como el análisis de datos y el uso básico de algoritmos de aprendizaje automático (Machine Learning), es posible identificar patrones que explican, por ejemplo, por

qué sube el costo de vida, por qué el dólar se dispara en ciertos meses o qué relación existe entre el salario mínimo y la inflación. Estos conocimientos permiten anticiparse a los impactos de decisiones políticas o eventos internacionales en la economía del hogar o del negocio.

Por ejemplo, si sabemos que una caída fuerte en el precio del petróleo (como ocurrió en 2014) impacta negativamente el crecimiento económico del país, una persona puede entender por qué puede haber más dificultad para conseguir empleo en ciertos sectores, o por qué sube el precio de productos importados. Entender esto no es solo un dato más: es una herramienta para adaptarse mejor a los cambios y tomar decisiones más acertadas.

De igual manera, si desde el gobierno o las universidades se crean tableros visuales donde se muestren estos indicadores de forma clara y comprensible, cualquier estudiante, comerciante o emprendedor podría analizar tendencias y prepararse mejor: ajustar precios, buscar oportunidades en sectores menos afectados, o entender cuándo es buen momento para invertir o ahorrar.

Además, este conocimiento empodera al ciudadano: si los datos económicos se explican con claridad y están disponibles en plataformas digitales, se promueve una sociedad más crítica y participativa, que puede exigir mejores políticas públicas y entender sus efectos.

## 16. Discusión de resultados

El análisis de los indicadores macroeconómicos de Colombia entre 2005 y 2019 permitió identificar patrones de comportamiento coherentes con diversos hechos económicos nacionales e internacionales que impactaron significativamente la economía del país. Al confrontar los resultados del análisis de datos con los eventos históricos más relevantes, se obtiene una interpretación más robusta del contexto macroeconómico nacional.

Uno de los primeros hallazgos fue la relación entre la TRM (Tasa Representativa del Mercado) y eventos internacionales como la crisis financiera de 2008 y la caída del precio del petróleo en 2014, los cuales provocaron una fuerte devaluación del peso colombiano. Esta tendencia fue corroborada con los valores de TRM en el conjunto de datos, que mostraron un incremento notable en estos periodos. A su vez, estos eventos afectaron el Índice de Precios al Consumidor (IPC), lo cual fue evidente en el aumento de la inflación observado en 2015 y 2016.

Por otro lado, la revisión de la tasa de desempleo y el PIB evidenció una correlación inversa: años con mayor crecimiento económico, como 2011 y 2013, estuvieron acompañados de una reducción en el desempleo. En contraste, los efectos de la reforma tributaria de 2016, que elevó el IVA, generaron una disminución en el consumo de los hogares y un impacto directo en la dinámica de empleo, lo cual también se reflejó en la tendencia ascendente del desempleo urbano en 2017.

Desde una perspectiva práctica, este análisis no solo contribuye al entendimiento académico de la economía nacional, sino que también ofrece herramientas útiles para tomadores de decisiones. Por ejemplo, al categorizar el comportamiento de la inflación o del salario mínimo mediante técnicas de análisis de datos, se podrían diseñar políticas más efectivas para

estabilizar la economía, apoyar a sectores vulnerables y anticipar impactos derivados de cambios fiscales o externos.

Este tipo de discusión evidencia cómo la integración entre conocimiento técnico (análisis de datos, visualización, categorización) y conocimiento contextual (historia económica, reformas legales, eventos globales) puede fortalecer el razonamiento analítico y mejorar las capacidades de diagnóstico y planificación tanto en estudiantes como en profesionales.

## 17. Conclusiones

El uso de algoritmos de aprendizaje automático, aplicado a indicadores macroeconómicos, permite obtener una lectura más profunda de fenómenos como el desempleo en Colombia, especialmente entre los años 2005 y 2019. Esta aproximación va más allá de la estadística descriptiva, transformando los datos en una herramienta que apoya la formulación de políticas públicas y estrategias sectoriales.

La limpieza, transformación e interpolación de datos resultaron ser etapas fundamentales para garantizar la fiabilidad del modelo. El uso de técnicas como la categorización de la tasa de desempleo y el escalamiento de variables permitió preparar adecuadamente los datos para su análisis automatizado, asegurando coherencia en la interpretación de los resultados y facilitando la visualización de patrones ocultos.

Desde una perspectiva metodológica, este trabajo permitió validar que un análisis cuantitativo, visual y estructurado de series históricas económicas puede generar información valiosa para la toma de decisiones institucionales. Además, la naturaleza interpretativa del análisis ofrece un marco replicable para futuras investigaciones que deseen incluir nuevas variables como informalidad laboral, inflación mensual o pobreza multidimensional.

El algoritmo desarrollado, al categorizar la tasa de desempleo, permitió identificar patrones ocultos que pueden anticipar condiciones económicas adversas. Así, este tipo de modelo computacional puede ser utilizado como insumo clave por entidades gubernamentales y académicas para establecer alertas tempranas en sectores económicos vulnerables.

El modelo de clasificación desarrollado demostró una alta capacidad predictiva en algunos algoritmos como KNN, QDA y Random Forest, los cuales alcanzaron una precisión del 100% en los datos de prueba. Esto sugiere que el conjunto de variables económicas utilizadas tiene un alto poder explicativo frente al fenómeno de desempleo. Sin embargo, estos resultados también advierten la necesidad de precaución frente al posible sobreajuste (overfitting), dadas las limitaciones del volumen de datos.

## Referencias Bibliográficas

- Banco Mundial. (2020). Colombia – Panorama general. <https://www.bancomundial.org/es/country/colombia/overview>
- Banco de la República de Colombia. (2024). Indicadores económicos y financieros. Recuperado de <https://www.banrep.gov.co>
- Banco Mundial. (2017). Colombia: Panorama económico y reformas estructurales. <https://www.bancomundial.org>
- Cepal. (2018). Estudio económico de América Latina y el Caribe: Colombia. <https://www.cepal.org/es/publicaciones>
- Departamento Administrativo Nacional de Estadística (DANE). (2021). Informe de calidad de vida en Colombia 2020. <https://www.dane.gov.co>
- DANE. (2024). Series históricas de mercado laboral e inflación. <https://www.dane.gov.co>
- Fondo Monetario Internacional. (2009). Perspectivas de la economía mundial – Impacto de la crisis financiera global. <https://www.imf.org>
- Gujarati, D. N., & Porter, D. C. (2009). Basic econometrics (5th ed.). McGraw-Hill Education
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2nd ed.). Springer.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106. <https://doi.org/10.1257/jep.31.2.87>
- Ministerio de Hacienda y Crédito Público. (2016). Reforma tributaria estructural: Ley 1819 de 2016. <https://www.minhacienda.gov.co>
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28. <https://doi.org/10.1257/jep.28.2.3>