

# **La Corrupción y su Impacto en el Crecimiento Económico Sostenible: Análisis de Datos y Big Data para la Identificación de Factores y Generación de un Modelo de Machine Learning**

Danilo Esteban Rodríguez, Jorge Humberto Amórtegui, Sandra Milena Trujillo

Especialización en Big Data, Fundación Universitaria Compensar

Propuesta de Consultoría

Yeimmy Katherine Serrano León

27 de noviembre del 2023



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/).



**La Corrupción y su Impacto en el Crecimiento Económico  
Sostenible: Análisis de Datos y Big Data para la Identificación  
de Factores y Generación de un Modelo de Machine Learning**

**Danilo Esteban Rodríguez, Jorge Humberto Amórtegui,  
Sandra Milena Trujillo**

Propuesta de consultoría presentada como requisito parcial para optar al título de:  
**Especialista en Big Data**

Fundación Universitaria Unipanamericana – Compensar  
Facultad de Ingeniería, Especialización en Big Data  
Bogotá, Colombia

2023



## Resumen

La lucha contra la corrupción debido a su impacto en los índices de desarrollo es esencial, con el Análisis de Datos y Big Data es posible identificar factores y generar un modelo de Machine Learning que aborda la relación entre corrupción y crecimiento económico sostenible. Para abordar esta problemática, se deben intervenir múltiples aspectos clave, como la obtención, el procesamiento y el análisis de los datos.

El objetivo principal del estudio es identificar los factores que influyen en esta relación y generar un modelo predictivo que pueda brindar asesoramiento estratégico a organizaciones interesadas en mitigar los efectos de la corrupción en el desarrollo económico.

La investigación se basa en un enfoque multidisciplinario y utiliza varias técnicas y herramientas, como la extracción y limpieza de datos, análisis estadísticos, minería de datos y modelos de regresión. Se recopilan datos de fuentes confiables y verificables, tales como informes gubernamentales y bases de datos internacionales.

Los resultados muestran una fuerte correlación entre indicadores de gobernanza y el Producto Interno Bruto per cápita, lo que sugiere que mejoras en áreas como el control de la corrupción, la efectividad gubernamental, el estado de derecho y la voz ciudadana pueden tener un impacto positivo en el desarrollo económico.

Por lo tanto, es fundamental abordar la corrupción y mejorar la gobernanza para promover un entorno propicio para el crecimiento económico y la prosperidad. Se proporcionan recomendaciones informadas, así como visualizaciones para facilitar la comprensión de los resultados.

**Palabras clave: (Machine Learning, PCA, corrupción, PIB per cápita).**



## Abstract

The fight against corruption and its impact on development rates is essential; with Data Analysis and Big Data we can identify factors and generate a Machine Learning model that addresses the relationship between corruption and sustainable economic growth. To address this problem, multiple key factors must be intervened.

The main objective of the study is to identify the factors that influence this relationship and generate a predictive model that can provide strategic advice to cooperative organizations in mitigating the effects of corruption on economic development.

The research is based on a multidisciplinary approach and uses various techniques and tools, such as data extraction and cleaning, statistical analysis, data mining and regression models. Data is collected from reliable sources, such as government reports and international databases.

The results show a strong evaluation between governance indicators and the Gross Domestic Product per capita, suggesting that improvements in areas such as corruption control, government effectiveness, the rule of law and citizen voice can have a positive impact on economic development.

Addressing corruption and improving governance is critical to promoting an enabling environment for economic growth and prosperity. Informed recommendations are provided, as well as visualizations to facilitate understanding of the results.

**Keywords: (Machine Learning, PCA, corruption, GDP per cápita).**



# Contenido

RESUMEN .....	V
ABSTRACT .....	VII
INTRODUCCIÓN .....	11
ANTECEDENTES Y JUSTIFICACIÓN .....	11
FORMULACIÓN DEL PROBLEMA .....	12
OBJETIVO GENERAL .....	12
<i>Objetivos Específicos</i> .....	13
ALCANCES Y LIMITACIONES .....	13
1.1 CORRUPCIÓN: DEFINICIÓN Y DIMENSIONES .....	14
1.2 CRECIMIENTO ECONÓMICO SOSTENIBLE .....	15
1.3 RELACIÓN ENTRE CORRUPCIÓN Y CRECIMIENTO ECONÓMICO SOSTENIBLE .....	15
1.4 FACTORES DETERMINANTES DE LA RELACIÓN .....	15
2.DISEÑO METODOLÓGICO .....	17
2.1 TIPO DE INVESTIGACIÓN: .....	17
2.2 MUESTRA O DATOS UTILIZADOS: .....	17
2.3 RECOPIACIÓN DE DATOS:.....	17
2.4 ANÁLISIS DE DATOS Y MINERÍA DE DATOS:.....	18
2.5 MODELOS PREDICTIVOS: .....	18
2.6 IDENTIFICACIÓN DE FACTORES SIGNIFICATIVOS:.....	18
3.RESULTADOS .....	19
3.1 MATRIZ DE CORRELACIÓN .....	20
3.2 COMPROBACIÓN DE NORMALIDAD .....	20
3.3 CORRELACIONES NO PARAMÉTRICAS.....	21
3.4 MODELOS DE REGRESIÓN .....	21
3.5 ANÁLISIS DE COMPONENTES PRINCIPALES.....	23
4.DISCUSIÓN .....	24
5.CONCLUSIONES Y RECOMENDACIONES.....	25
5.1 CONCLUSIONES.....	25
5.2 RECOMENDACIONES.....	26
BIBLIOGRAFIA .....	27



# Introducción

La corrupción es un problema persistente y sistémico que afecta a numerosos países en todo el mundo. Sus efectos negativos en el crecimiento económico sostenible y el desarrollo social son ampliamente reconocidos. Por esta razón, se ha vuelto fundamental abordar en profundidad este fenómeno y encontrar soluciones efectivas para combatirlo. En esta investigación se pretende mostrar la relación entre la corrupción y el crecimiento económico sostenible, utilizando el análisis de datos y el Big Data como herramientas fundamentales en este proceso. Nuestro objetivo es identificar los factores clave que influyen en esta relación y generar un modelo predictivo que sirva para realizar asesorías estratégicas a organizaciones interesadas, como repositorio público para empresas que quieran orientar su impacto social. A través de un enfoque multidisciplinario y el uso de técnicas como el ETL, la minería de datos, análisis de correlación, de componentes principales y de regresión, buscamos proporcionar una visión enriquecedora sobre este tema y contribuir al debate académico y político en la lucha contra la corrupción.

## Antecedentes y Justificación

Los métodos de Deep Learning se han aplicado cada vez más en la predicción de series de tiempo económicas. Podemos ver como la aplicación de redes neuronales recurrentes para predecir el comportamiento de las variables macroeconómicas de las economías de América Latina. Los modelos de deep learning, como el Perceptrón Multi Capa, Redes Neuronales Convolucionales (CNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) y SeqToSeq se aplicaron para cada país y frecuencia de las series (Anual y trimestral). Los países seleccionados para el análisis fueron Argentina, Brasil, Chile, Colombia, México y Perú. Los datos anuales abarcaron desde 1960 hasta 2020, mientras que los datos trimestrales variaron según el país, desde 1993 hasta 2021. (Alegre, V; Lozano, J; 2022)

Otro estudio encontrado nos muestra las predicciones de las tasas de crecimiento en los PIB's de los países de la Comunidad Europea, en el cual, con la aplicación de modelos lineales bayesianos, son una buena alternativa para la predicción del PIB, ya que logran resultados similares a los modelos tradicionales con una menor complejidad (Basulto, J. y Ortega, F.; 2003).

También encontramos la evaluación del impacto del VIII Plan de la Nación en el crecimiento del PIB per cápita en Venezuela durante el período 1970-1994. Para evaluar los resultados potenciales, los autores emplean un método de Contrafactual Sintético, considerando escenarios hipotéticos donde las medidas del plan se implementaron sin la interferencia de eventos como el Caracazo y los golpes de Estado (Colmenares, C; 2020).

El presente trabajo busca favorecer a la comprensión de la corrupción, a la identificación de los factores y con ello que las entidades que correspondan puedan actuar para mejorar los índices de desarrollo de los países. Para ello, se empleará una metodología que combina el análisis estadístico de datos públicos, minería de datos y elaboración de un modelo de Machine Learning. Los resultados de este estudio nos permiten identificar los factores que inciden en los indicadores de desarrollo sostenible y como se combina la corrupción con los mismos.

## **Formulación del Problema**

¿Cómo puede modelarse la relación entre la corrupción y los indicadores de desarrollo económico sostenible en los países de la CEPAL utilizando técnicas de Machine Learning identificando los factores más influyentes en esta compleja interacción? Esta propuesta busca diagnosticar y comprender de manera empírica las conexiones entre la corrupción y el desarrollo económico sostenible, con el fin de identificar vías para mejorar la situación en la región y avanzar hacia un estado más favorable.

## **Objetivo General**

Modelar la relación entre la corrupción y el crecimiento económico de los países de la CEPAL. Identificando factores influyentes, por medio de procesos de extracción, análisis y modelado

de datos, para generar un insumo que permita identificar insights y oportunidades relevantes de intervención para mitigar los efectos de la corrupción en las metas del objetivo.

## **Objetivos Específicos**

- Recopilar, extraer y limpiar (ETL) datos relacionados con indicadores de corrupción y metas del desarrollo económico en países de la CEPAL. Incluyendo datos provenientes de fuentes confiables como organismos de control internacionales.
- Utilizar técnicas de análisis de datos y minería de datos, para explorar patrones, correlaciones y relaciones causales entre los factores de corrupción y los índices de desarrollo. Emplear modelos predictivos para identificar variables clave que impactan la relación para una comprensión más profunda de las dinámicas subyacentes.
- Generar un insumo que permita identificar oportunidades relevantes de intervención y así mitigar los efectos de la corrupción para llegar al cumplimiento de los objetivos de desarrollo sostenible.

## **Alcances y Limitaciones**

- Recopilación y Análisis de Datos: El proyecto incluirá la recopilación de datos relevantes de países miembros de la CEPAL donde se analizarán con el fin de identificar patrones y correlaciones.
- Minería de Datos y Modelado Predictivo: Se utilizarán para identificar relaciones significativas entre la corrupción y los índices de desarrollo, centrándose en el crecimiento económico sostenible. Esto incluirá el uso de algoritmos de aprendizaje automático y modelos de regresión.
- Identificación de Factores de Influencia: Busca identificar factores específicos que influyen en la relación entre la corrupción y el crecimiento económico sostenible, haciendo uso de análisis de relaciones entre las variables, análisis de interacción y análisis de componentes principales esto podría incluir variables como el gasto público, la inversión extranjera, la educación, la salud, entre otros.

- Recomendaciones Informadas: Con los resultados se generarán recomendaciones informadas y estrategias para abordar los factores que contribuyen a la corrupción en países miembros de la CEPAL. (4)
- Disponibilidad de Datos: La calidad y disponibilidad de datos pueden variar entre los países de la CEPAL, lo que podría limitar la comparabilidad y la generalización de los resultados.
- Complejidad de los Factores: El proyecto puede no abarcar todos los factores relevantes en la medición de variables independiente y dependientes, ya que son multifacéticos y complejos.
- Factores Externos: Los eventos inesperados pueden influir en la relación de la corrupción y el desarrollo, como crisis económicas globales o eventos políticos significativos.
- Implementación de Recomendaciones: Se pueden generar recomendaciones, pero esto no garantiza la implementación de estas.
- Cambio a Largo Plazo: La mitigación sobre efectos de la corrupción en el desarrollo pueda que sus resultados no se materialicen de inmediato.
- 

## **1. Marco teórico**

### **1.1 Corrupción: Definición y Dimensiones**

La corrupción se define como el uso indebido del poder público para beneficio personal o privado. Esta práctica puede manifestarse en diversas formas, como sobornos, nepotismo, malversación de fondos públicos, tráfico de influencias y falta de transparencia en las instituciones gubernamentales (Transparency International, 2022).

## 1.2 Crecimiento Económico Sostenible

El crecimiento económico sostenible es un concepto que abarca el desarrollo económico a largo plazo sin comprometer los recursos y la calidad de vida de las generaciones futuras. Este tipo de crecimiento se caracteriza por la sostenibilidad ambiental, la equidad social y el desarrollo humano (Naciones Unidas, 2018).

## 1.3 Relación entre Corrupción y Crecimiento Económico Sostenible

Diversas teorías y estudios han explorado la relación entre la corrupción y el crecimiento económico sostenible:

- Teoría de la Economía de la Corrupción: Sostiene que la corrupción disuade la inversión extranjera, socava la eficiencia económica y debilita la competitividad de las empresas, lo que a su vez afecta negativamente el crecimiento económico.
- Teoría de la Maldición de los Recursos: Sugiere que los países ricos en recursos naturales son más propensos a la corrupción, lo que limita la diversificación económica y obstaculiza el desarrollo sostenible.
- Teoría de la Efectividad Gubernamental: Destaca la importancia de instituciones gubernamentales efectivas y no corruptas para promover el crecimiento sostenible.
- Enfoque de Desarrollo Humano: Se centra en la relación entre la corrupción y la calidad de vida de la población, sosteniendo que la corrupción afecta negativamente la prestación de servicios públicos esenciales (Naciones Unidas, 2018).

## 1.4 Factores Determinantes de la Relación

Identificar los factores que explican la relación entre corrupción y crecimiento económico sostenible es fundamental. Estos factores pueden incluir ser: (1) gravedad de la corrupción, (2) efectividad y transparencia de las instituciones gubernamentales, (3) Flujo de inversión extranjera, (4) recursos naturales, (5) existencia y aplicación de políticas contra la corrupción y (6) el nivel educativo de la mano con la conciencia ciudadana (Díaz, 2018).



## 2. Diseño Metodológico

La investigación se estructuró en varias etapas clave que abarcan desde la recopilación de datos hasta la generación de recomendaciones. Estas etapas se detallan a continuación:

### 2.1 Tipo de Investigación:

Diseño de la Investigación: Este estudio se enmarca en una investigación de tipo descriptiva correlacional. Se buscó describir la relación entre la corrupción y el crecimiento económico sostenible, así como identificar posibles correlaciones y patrones causales.

### 2.2 Muestra o Datos Utilizados:

Fuentes de Datos: Incluyeron informes gubernamentales, bases de datos internacionales almacenadas en los datasets del catálogo de datos del Banco Mundial.

### 2.3 Recopilación de Datos:

- a. Identificación de Fuentes de Datos: Se identificaron y seleccionaron fuentes de datos relacionadas con indicadores de desarrollo económico sostenible, desarrollo global y gobernanza (Banco Mundial).
- b. Proceso de Extracción: Los datos se extrajeron directamente de la página Web del Banco Mundial por medio de una API. Se tuvieron en cuenta las consideraciones éticas y legales pertinentes durante este proceso.
- c. Limpieza de Datos (ETL): En primer lugar, se realiza la limpieza de los DataFrames generados, eliminando la información con valores perdidos superior al 26%. Seguidamente, se realiza una imputación de los valores faltantes calculando la mediana por columna en subconjuntos de indicadores. Esto garantiza la calidad y coherencia de los datos utilizados en el análisis subsiguiente.

## 2.4 Análisis de Datos y Minería de Datos:

Técnicas de Análisis: Se utilizan técnicas de análisis de datos y minería de datos para evaluar las relaciones entre las variables, extraer las más relevantes para el análisis detallado y asegurar que las variables resultantes cuentan con registro para la misma ventana de tiempo.

Análisis de correlación específico: Se lleva a cabo una prueba de normalidad Anderson-Darling para evaluar la distribución de las variables numéricas y determinar los análisis de correlación a realizar. Se aplican análisis de correlación de Pearson y Spearman según la distribución de cada variable para comprobar la asociación entre las mismas y se extraen las variables con correlaciones más altas respecto a la variable dependiente.

## 2.5 Modelos predictivos:

Se realizaron entrenamientos de modelos de regresión lineal para prever la variable objetivo 'GDE\_SL.GDP.PCAP.EM.KD' (PIB per cápita). En el primer escenario, se realiza la división de muestras sin considerar la variable categórica país, utilizando 'CRP\_CC.EST' (corrupción estatal), 'CRP\_GE.EST' (efectividad del gobierno), 'CRP\_RL.EST' (cumplimiento del gobierno), 'CRP\_VA.EST' (voz y rendición de cuentas) como variables independientes. En los otros escenarios se generan modelos adicionales, uno con el conjunto completo de datos globales y otro con un conjunto reducido, evaluando así la influencia de esta variable en las predicciones. Se utiliza la técnica de regresión lineal y se evalúa el rendimiento del modelo en conjuntos de entrenamiento y prueba.

## 2.6 Identificación de Factores Significativos:

Análisis de interacción: En este análisis, con las variables que guardan mayor relación con la variable dependiente, se utilizaron visualizaciones de interacción para identificar la relación entre las variables.

Factores Clave: Se identificaron indicadores específicos que demostraron tener un impacto significativo en los índices de desarrollo afectados por la corrupción como: Control de la Corrupción (CC.EST), efectividad del Gobierno (GE.EST), estado de Derecho - Cumplimiento de la ley (RL.EST) y voz y Rendición de Cuentas (VA.EST) variables independientes y el indicador dependiente, el Producto Interno Bruto (PIB) per cápita ajustado por paridad de poder adquisitivo (SL.GDP.PCAP.EM.KD).

### 3.Resultados

En este análisis, hemos utilizado una serie de indicadores del Banco Mundial que se centran en la gobernanza y el desempeño de los gobiernos. Una vez realizados la captura de los datos, la transformación de las variables y el cargue de la base de datos, se procede a realizar el análisis exploratorio, los análisis descriptivos, análisis correlacionales y análisis predictivos para poder medir la relación que existe entre la corrupción y el desarrollo económico e identificar las variables influyentes, en este sentido las variables objeto de análisis son:

**Control de la Corrupción (CC.EST):** Control de la Corrupción evalúa la percepción de hasta qué punto el poder público se ejerce para beneficio privado, abarcando formas tanto menores como mayores de corrupción.

**Efectividad del Gobierno (GE.EST):** Efectividad del Gobierno mide la calidad de los servicios públicos, la independencia del servicio civil y la credibilidad del compromiso del gobierno con sus políticas.

**Estado de Derecho - Cumplimiento de la ley (RL.EST):** Estado de Derecho refleja la confianza y el cumplimiento de las reglas de la sociedad, incluyendo la aplicación de contratos, derechos de propiedad, la actuación policial y judicial, así como la probabilidad de crimen y violencia.

**Voz y Rendición de Cuentas (VA.EST):** Voz y Rendición de Cuentas evalúa la participación ciudadana en la selección del gobierno y la libertad de expresión, asociación y medios de comunicación.

Al realizar un modelo de regresión lineal con estos indicadores, la primera parte del análisis muestra que hay unas relaciones significativas entre las variables independientes y el indicador dependiente, el **Producto Interno Bruto (PIB) per cápita ajustado por paridad de poder adquisitivo (SL.GDP.PCAP.EM.KD)**. Si bien el modelo de regresión no identifica un comportamiento específico para los datos de Colombia, la matriz de correlación y las correlaciones específicas sugieren que las mismas variables guardan relación de ocurrencia.

Sin embargo, es importante destacar que el modelo no presenta un alto nivel predictivo para los datos específicos de Colombia, a pesar de las correlaciones específicas. Esto puede deberse a factores únicos en el contexto colombiano que no están plenamente capturados por los indicadores utilizados.

A pesar de esta limitación, las fuertes correlaciones entre los indicadores de gobernanza y el PIB per cápita sugieren que mejoras en el control de la corrupción, la efectividad gubernamental, el estado de derecho y la voz ciudadana podrían tener implicaciones positivas en el desarrollo económico de un país. Es crucial abordar estos aspectos para fomentar un entorno propicio para el crecimiento económico sostenible y la prosperidad.

### 3.1 Matriz de correlación

A partir de la matriz de correlación generada, se identificaron las variables dependientes que guardan mayor relación con las variables independientes, tanto para el subconjunto de datos completos como para los datos de Colombia.

Subconjunto	Variable	CC.EST	GE.EST	PV.EST	RL.EST	RQ.EST	VA.EST
Completo	SL.GDP.PCAP.EM.KD	0,80	0,89		0,86	0,79	0,81
Colombia	SL.GDP.PCAP.EM.KD			0,86		0,76	0,92

Tabla 1. Indicadores de desarrollo global con correlaciones mayores a 0,7 respecto a los indicadores de corrupción.

### 3.2 Comprobación de normalidad

Para las variables extraídas se realiza una prueba de normalidad, comprobando que todas las variables del conjunto global de datos no se comportan de manera normal. Debido a que posteriormente el modelo de regresión no obtuvo valor predictivo para el subconjunto de datos de Colombia, no se profundizaron las relaciones específicas en este segmento de datos.

Subconjunto	Variable	Estadístico	Valor Crítico	Conclusión
Completo	CRP_CC.EST	22,31	0,78	No Normal
Completo	CRP_GE.EST	15,09	0,78	No Normal
Completo	CRP_PV.EST	1,57	0,78	No Normal
Completo	CRP_RL.EST	20,92	0,78	No Normal
Completo	CRP_RQ.EST	5,74	0,78	No Normal
Completo	CRP_VA.EST	10,59	0,78	No Normal
Completo	GDE_SL.GDP.PCAP.EM.KD	18,25	0,78	No Normal

Tabla 2. Comprobación de normalidad de variables a nivel global.

### 3.3 Correlaciones no paramétricas

Ya que las variables para el subconjunto global de datos se comportan de manera no normal, se aplicaron correlaciones de Pearson entre las variables independientes y la variable para comprobar la relación establecida por la matriz de correlación, validando nuevamente una evidencia a favor de la relación entre los indicadores.

Subconjunto	Tipo de análisis	Variable	Correlación	Valor p
Completo	Pearson	CRP_CC.EST	0,76	4,54E-85
Completo	Pearson	CRP_GE.EST	0,86	1E-133
Completo	Pearson	CRP_PV.EST	0,69	2,83E-65
Completo	Pearson	CRP_RL.EST	0,83	1,9E-117
Completo	Pearson	CRP_RL.EST	0,56	0,004766
Completo	Pearson	CRP_RQ.EST	0,77	4,77E-87
Completo	Pearson	CRP_VA.EST	0,84	1,2E-121

Tabla 3. Análisis de correlación de variables independientes con la variable dependiente a nivel global.

### 3.4 Modelos de regresión

Se encontró que los modelos con mayor valor predictivo requieren el conjunto de datos completo e incluir el país como variable categórica. Adicionalmente, la capacidad predictiva de los modelos más predictivos se aplica con datos entrenados desde el año 1996 hasta el 2002, pudiendo predecir la variable dependiente con las variables independientes de 1997 en adelante. Adicionalmente, en la utilización del método random forest demuestra que hay mayor precisión en el nivel predictivo cuando se incluye el país como variable categórica, pero sigue siendo inferior esta capacidad respecto a los modelos con el método de regresión lineal. La Tabla 3 muestra un consolidado de los modelos realizados para este análisis, indicando para cada uno el MSE -que indica mejor ajuste del modelo en cuanto más cercano esté a cero- y el R2 -que indica mejor ajuste del modelo en cuanto más cercano esté a 1-. Así mismo, se presentan gráficos que representan el contraste de la variable dependiente original en comparación con la predicción de los modelos.

Subconjunto	Variable		MSE	R2	MSE	R2
	categórica	Método	Entrenamiento	Entrenamiento	Prueba	Prueba
Completo	Ninguna	Regresión lineal	0,01	0,84	0,01	0,84
Completo	País	Regresión lineal	4,9E-29	1	7,53E-28	1
Entrenado desde 1996 hasta 2002	País	Regresión lineal	8,63E-30	1	1,17E-28	1
Predicción desde 2003 en adelante	País	Regresión lineal	8,63E-30	1	7,53E-28	1
Completo	Ninguna	Random forest	0,005	0,91	0,006	0,006
Completo	País	Random forest	0,002	0,96	0,008	0,86

Tabla 3. Modelos de regresión entrenados y probados para predecir la variable objetivo.

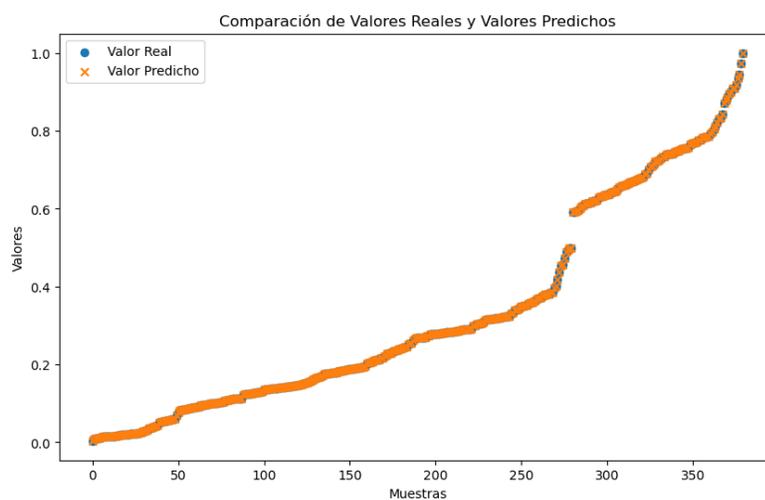


Figura 1. Comparación de valores reales y valores predichos por el modelo de regresión con el conjunto de prueba, con el país como variable categórica.

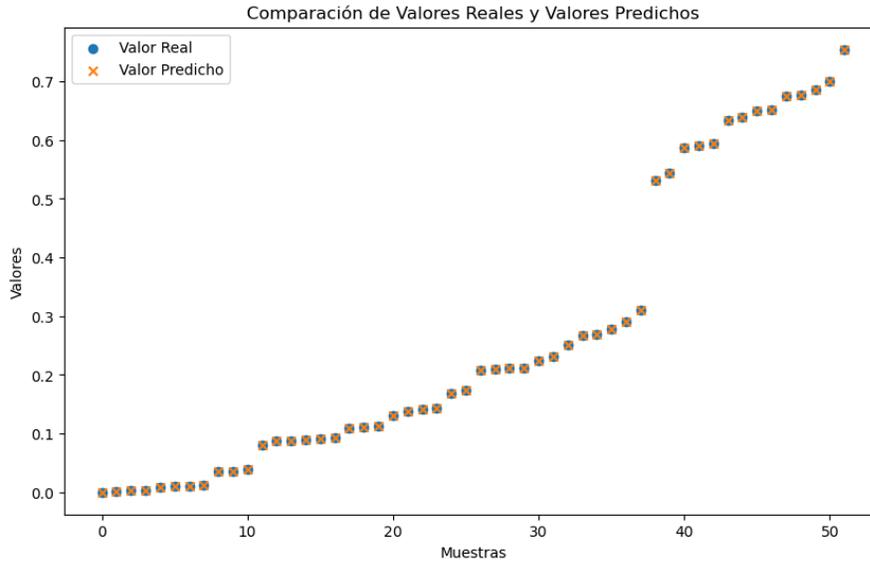


Figura 2. Comparación de valores reales y valores predichos por el modelo de regresión con el conjunto de prueba desde 1996 hasta 2002 , con el país como variable categórica.

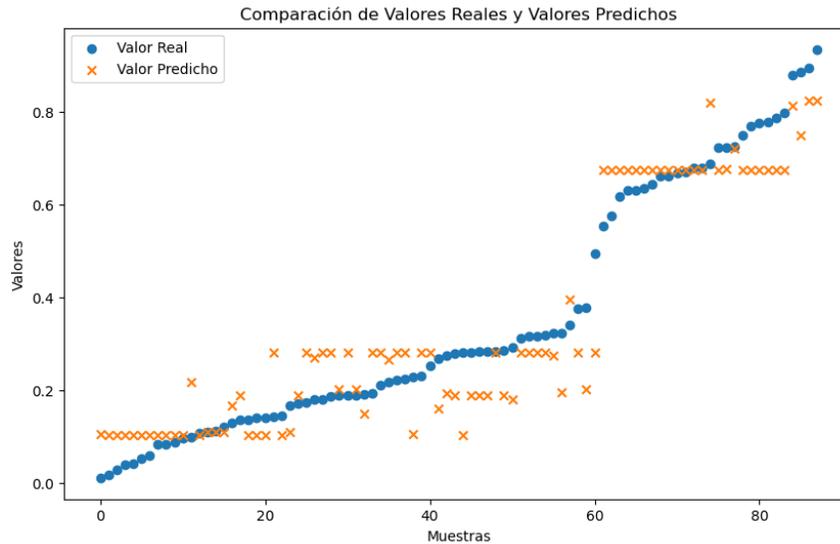


Figura 3. Comparación de valores reales y valores predichos por el modelo random forest con el conjunto de prueba, con el país como variable categórica.

### 3.5 Análisis de componentes principales

Finalmente, el análisis de componentes principales permitió identificar que los indicadores de corrupción agrupados en 2 componentes explican el 95% de la varianza, los que sugiere que este número de componentes puede capturar la mayoría de la información de las

variables de corrupción. De acuerdo a los pesos de los componentes, el primero tiene pesos relativamente altos para todas las variables y el segundo parece capturar variaciones específicas relacionadas con 'CRP\_GE.EST' y 'CRP\_VA.EST'.

Componentes	CRP_CC.EST	CRP_GE.EST	CRP_RL.EST	CRP_VA.EST
1	0,50	0,48	0,54	0,45
2	0,04	0,56	0,13	-0,81

tabla 4. pesos de los componentes identificados en el análisis de componentes principales.

## 4. Discusión

En la propuesta se explora la relación entre la corrupción y el crecimiento económico sostenible mediante técnicas avanzadas de análisis de datos y minería de datos. Los resultados revelan una correlación significativa entre los indicadores de gobernanza, como el control de la corrupción, la efectividad gubernamental, el estado de derecho y la voz ciudadana, y el Producto Interno Bruto (PIB) per cápita. Este hallazgo sugiere que la mejora de estos aspectos de la gobernanza puede incidir positivamente en el desarrollo económico de un país. No obstante, se reconoce que el modelo de regresión lineal empleado presenta limitaciones y podría no capturar exhaustivamente los factores específicos que afectan la relación entre la corrupción y el crecimiento económico. A pesar de estas restricciones, la investigación proporciona recomendaciones fundamentadas y visualizaciones para facilitar la interpretación de los resultados y fomentar estrategias efectivas en la lucha contra la corrupción y la promoción del crecimiento económico sostenible.

## 5. Conclusiones y recomendaciones

### 5.1 Conclusiones

La corrupción tiene un impacto negativo en el crecimiento económico sostenible: El estudio reconoce que la corrupción junto con otros factores afecta negativamente el crecimiento económico sostenible y en específico el PIB per cápita. Se destaca la importancia de abordar este problema para promover un entorno propicio para el crecimiento económico y la prosperidad.

El análisis resalta la importancia de variables como el control de la corrupción, la efectividad gubernamental, el estado de derecho y la voz y rendición de cuentas como aspectos clave en la lucha contra la corrupción. Mejorar la gobernanza en estos aspectos puede tener implicaciones positivas en el desarrollo económico de un país.

Se plantea la utilización del modelo propuesto para evaluar que acciones se deben tomar según el caso particular del país o entidad que esté pensando en aumentar su PIB per cápita. Estas herramientas pueden proporcionar información y conocimientos importantes para desarrollar estrategias efectivas de intervención y mitigación de los efectos de la corrupción.

Adicional al modelo de regresión lineal, aplicamos un análisis de Componentes Principales (PCA) que reveló que con 2 componentes se explica más del 95% de la varianza en los indicadores de corrupción. La identificación de dos clusters sugiere patrones específicos en los datos, en donde habría un primer componente que describe una tendencia general de todas las variables independientes y un segundo componente destacando la importancia de variables como 'CRP\_GE.EST' y 'CRP\_VA.EST', en la formación de tendencias específicas. La contribución relativa de ciertos aspectos de la gobernanza en estos componentes refuerza la importancia de estos indicadores en la relación entre gobernanza y desarrollo económico.

Se reconocieron las limitaciones del estudio, incluyendo posibles sesgos, datos faltantes o restricciones de tiempo que pudieron influir en la profundización de los factores de corrupción que generan variaciones específicas en la variable dependiente.

Se dejan a continuación los enlaces del repositorio en GitHub y la aplicación que se realizó como evidencia del trabajo: <https://worldbank-corruption-insights-5jlyewujypilprgdb9eyeg.streamlit.app/> - <https://github.com/derodriguez/WorldBank-Corruption-Insights>

## 5.2 Recomendaciones

- a. Basadas en Hallazgos: Las fuertes correlaciones entre los indicadores de gobernanza y el PIB per cápita sugieren que mejoras en el control de la corrupción, la efectividad gubernamental, el estado de derecho y la voz ciudadana podrían tener implicaciones positivas en el desarrollo económico de los países de la CEPAL.
- b. Visualizaciones: Se crearon visualizaciones pertinentes para representar los hallazgos de manera efectiva y facilitar la comprensión de los resultados.

## Bibliografía

Alegre Ibáñez, V. A., & Lozano Aparicio, J. M. (2022). Aplicación de métodos de Deep Learning en series de tiempo para el pronóstico de la situación macroeconómica en América Latina. *Interfases*, 15(015), 102-130. <https://doi.org/10.26439/interfases2022.n015.5817>

Basulto Santos, J. y Ortega Irizo, F.J. (2003). Predicciones de las tasas de crecimiento relativas de los productos interiores brutos reales de un conjunto de países de la Comunidad Europea. Aplicación de los modelos lineales bayesianos. En V Reunión de Economía Mundial Sevilla: Universidad de Sevilla, Secretariado de Recursos Audiovisuales y Nuevas Tecnologías

Colmenares, C (2020). ¿La sociedad tuvo la culpa? Contrafactual sintético para modelar las consecuencias del VIII plan de la nación en el PIB per cápita entre 1970 y 1994. Universidad católica Andrés Bello facultad de ciencias económicas y sociales escuela de economía. <http://biblioteca2.ucab.edu.ve/anexos/biblioteca/marc/texto/AAU4538.pdf>

Díaz, J. M. (2018). Causas y efectos de la corrupción en las sociedades democráticas. *Revista de la Escuela Jacobea de Posgrado*, 14, 1-26.

El Índice de Percepción de la Corrupción 2021 revela que los niveles de corrupción se han estancado en los últimos diez años, en medio de un entorno de abusos a los derechos humanos y deterioro de la democracia (2022). Transparency.org; #creator. <https://www.transparency.org/es/press/2021-corruption-perceptions-index-press-release>

Naciones Unidas (2018), La Agenda 2030 y los Objetivos de Desarrollo Sostenible: una oportunidad para América Latina y el Caribe (LC/G.2681-P/Rev.3), Santiago.

Paredes, R., Reyes, N., & Campos, S. (2020). Aplicación de métodos de Deep Learning en series de tiempo para el pronóstico de la situación macroeconómica en América Latina. *Revista chilena de ingeniería*, 28(1), 135-144.

Verónica, R. D., & Gustavo, A. B. (2016). Predicciones de las tasas de crecimiento relativas de los productos interiores brutos reales de un conjunto de países de la Comunidad Europea. Aplicación de los modelos lineales bayesianos. *Estudios de Economía Aplicada*, 34(2), 437-460.