

**¿SON MÁS CORRUPTOS LOS PAÍSES MENOS ABIERTOS  
A LOS MERCADOS INTERNACIONALES?**

APLICACIÓN DE UN MODELO PREDICTIVO DE  
CLASIFICACIÓN BASADO EN REDES NEURONALES

IS THERE MORE CORRUPTION IN COUNTRIES  
LESS OPENED TO INTERNATIONAL MARKETS?  
APPLICATION OF A PREDICTIVE CLASSIFICATION  
MODEL BASED ON NEURAL NETWORKS

Cristian Picón Viana\*

---

\* Economista de la Universidad del Atlántico. Especialista en Estadística Aplicada y Magister (c) en Economía de la Universidad del Norte. Docente catedrático de la Universidad del Norte y de la Universidad del Atlántico. [crispicon@gmail.com](mailto:crispicon@gmail.com)

## RESUMEN

Los planteamientos más comunes a favor de la liberalización del comercio y las finanzas internacionales afirman que políticas aperturistas tendrán un impacto positivo en la reducción de la corrupción. No obstante la complejidad del estudio de estas relaciones, la mayoría de investigaciones al respecto se limitan a estudios correlacionales o determinísticos. En este estudio se aplicó un modelo predictivo de clasificación basado en redes neuronales denominado Perceptrón Multicapa (MLP por sus siglas en inglés) que cumple con un conjunto de cualidades estadísticas deseables, con el fin de estimar las características o “síntomas” que presenta un país catalogado como más o menos corrupto. De las variables utilizadas, los niveles de desarrollo humano (IDH) y los niveles de apertura económica son las características comunes que comparten los países con niveles semejantes de corrupción, permitiendo clasificarlos correctamente. Encontrándose evidencia de que entre menor sea el nivel de apertura económica de un país, mayor será la posibilidad de ser clasificado en un nivel de mayor corrupción

**PALABRAS CLAVE:** Apertura económica, comercio internacional, finanzas internacionales, corrupción, redes neuronales, perceptrón multicapa (MLP)

**CLASIFICACIÓN JEL:** C45, D73, H10, K49, F19, F39.

## ABSTRACT

The most common approaches in favor of liberalizing international trade and international finances suggest that opening policies will have a positive impact in reducing corruption. Despite of the complexity of studying these relationships, most research in this field is limited to correlational or deterministic studies. In this paper we applied a predictive model of classification based on neural networks called Multilayer Perceptron (MLP) that meets a set of desired statistics qualities, with the purpose of estimating the characteristics or “symptoms” present in countries categorized as more or less corrupt. Of the variables used, the levels of human development (HDI) and levels of economic openness are the common characteristics shared by countries with similar levels of corruption. We found evidence that shows us that if the economic openness level of a country is lower, the chance of being classified at a higher level of corruption increases.

**KEYWORDS:** Economic Integration, international trade, international finances, corruption, neural networks, Multilayer Perceptron (MLP)

**JEL CODES:** C45, D73, H10, K49, F19, F39.

## INTRODUCCIÓN

La discusión sobre la interacción entre las políticas de liberación comercial y financiera y el nivel de desarrollo institucional de los países que las adoptan ha cobrado relevancia gracias a los procesos aperturistas generalizados en el ámbito mundial desde la década del ochenta y las recientes crisis mundiales contagiadas y aceleradas por los niveles de interdependencia económica. Por lo tanto, el nivel de corrupción de un país, como componente importante de su institucionalidad, se transforma en objeto de interés, sobre todo en lo referente a su relación con objetivos de política comercial.

En este contexto, los planteamientos más comunes a favor de la liberalización del comercio y las finanzas internacionales afirman que las políticas aperturistas tendrán un impacto positivo en la reducción de la corrupción al disminuir el poder del Estado, fomentar la descentralización y la competencia, y generar incentivos para desarrollar de una manera más eficiente los procesos productivos (Friedman, 1962); desde un enfoque neo-institucional (Lambsdorff, 2007).

No obstante, la corrupción no es un fenómeno lineal ni simplemente el resultado de las decisiones de individuos ante incentivos perversos, sino un fenómeno complejo con muchos matices e incluso anclado a reglas de juego rígidas y resistentes a la generación de incentivos, como indican Boehm y Lambsdorf (2009) en su introducción. La corrupción es entonces el resultado de las instituciones legales, económicas, culturales y políticas de un país, por lo que puede ser una respuesta a reglas positivas o perjudiciales (Svensson, 2005).

Lo anterior muestra, como lo expresan Berg y Krueger (2002), lo complicado de investigar sobre la interacción entre

los niveles de apertura económica y factores institucionales como la corrupción y, por ende, la dificultad de proponer una relación de causalidad utilizando datos agregados sin correr el riesgo de realizar análisis mecanicistas y faltos de contenido real. Por ejemplo, Dreher, Kotsogiannis y McCorrison (2007) exponen los problemas de encontrar relaciones causales y proponen una metodología diferente en la medición de índices de corrupción.

Sin embargo, la mayoría de investigaciones al respecto utilizan herramientas estadísticas y econométricas que no tienen en cuenta lo anterior, ya que se limitan a estudios correlacionales o determinísticos (regresión) planteando la exogeneidad y/o linealidad de algunas variables que en el mundo real no lo son o por lo menos no hay evidencia de que lo sean; así mismo, analizan la relevancia de estas variables en términos parciales o *ceteris paribus* cuando la multicolinealidad es lo común.

Por tanto, lo ideal es analizar la relación entre el nivel de apertura a los mercados internacionales y los niveles de corrupción con herramientas robustas para el cumplimiento de supuestos y que no presupongan relaciones determinísticas o causales mecanicistas.

Este trabajo pretende incursionar en esta forma de entender el problema al analizar la relación entre los niveles de corrupción y un grupo de factores socio-económicos para los países estudiados, planteando las siguientes preguntas: primero, ¿el nivel de apertura económica de los países es una de las variables socio-económicas que permiten clasificarlos según su nivel de corrupción? Segundo, ¿cuál es la importancia relativa del nivel de apertura económica con respecto a las demás variables en la clasificación de los países según su nivel de corrupción? Y tercero, ¿cuál es el signo de esta relación?

Para contestarlas, se propone la utilización de un modelo predictivo de clasificación, basado en redes neuronales y denominado Perceptrón Multicapa (MLP, por sus siglas en inglés), que cumple con las características deseadas.

Los países estudiados se clasificaron, específicamente, por cuartiles, de acuerdo con sus niveles de corrupción y según su posición en el CPI publicado por Transparencia Internacional. En función de los valores de los campos de entrada se determinaron las variables, que al interactuar, permiten predecir en qué rango se encontrará un país dadas sus características. Por lo tanto, no se analiza la relación determinista entre variables sino las características o “síntomas” que presenta un país catalogado como más o menos corrupto.

Así pues, basados en el estudio de un conjunto de 155 países con datos de corte transversal para el año 2010, se encontró que:

1. De las variables utilizadas, los niveles de desarrollo humano (IDH) y los niveles de apertura económica son las características comunes que comparten los países con niveles semejantes de corrupción, lo que permite que sean clasificados correctamente.
2. El nivel de apertura económica a los mercados internacionales es la segunda característica en importancia para clasificar correctamente a los países según sus niveles de corrupción.
3. Se encontró evidencia de que entre menor sea el nivel de apertura económica de un país, mayor será la posibilidad de ser clasificado en un nivel de mayor corrupción.

Por otro lado, el documento está constituido de la siguiente manera: inicialmente se muestra un conjunto de investigaciones que se han realizado sobre la interrelación entre la corrupción y otras variables, especialmente los niveles de inserción en los mercados internacionales, así como una breve introducción al funcionamiento de una red neuronal MLP; una tercera y cuarta parte, en las que se explican de manera concisa las variables utilizadas y se expone la metodología desarrollada, un quinto componente, en el que se muestran los resultados y, por último, se plantean las conclusiones y propuestas pertinentes.

## 2. MARCO TEÓRICO Y ANTECEDENTES

### 2.1. Corrupción y apertura

No existe una definición concreta de corrupción, pero la más generalizada y la utilizada en este trabajo es la de Klitgaard (1988), que hace hincapié en las actividades de los funcionarios públicos que no obedecen a sus deberes debido a la búsqueda de beneficios personales. Véase también (Rose-Ackerman, 1999) y una aproximación detallada al fenómeno de la corrupción en Svensson (2005).

No obstante la complejidad del fenómeno, se han realizado intentos de plantear relaciones causales entre variables económicas e institucionales y los niveles de corrupción internacional. Tanzi (1998) discute aspectos relacionados con las causas, consecuencias y el alcance de la corrupción, y las posibles acciones correctivas. Hace hincapié en los costos de la corrupción en términos de crecimiento económico y en

el hecho de que la lucha contra ese fenómeno no puede ser independiente de la reforma del Estado.

En un estudio más técnico Treisman (2000) analiza varios índices de corrupción percibida y encuentra argumentos que indican que los países con tradiciones protestantes, los que han sido colonias británicas, las economías con mayores niveles de desarrollo y, aunque no tan concluyente, con mayores importaciones fueron menos “corruptos”.

Sandholtz y Koetzle (2000) estudian la variación en el nivel percibido de corrupción, definido como el abuso del cargo público para beneficio privado en una muestra de cincuenta países. Proponen un conjunto de hipótesis que explican la variación en los niveles de corrupción en términos de la estructura político-económica nacional, las normas democráticas, la integración en la economía internacional, y la afiliación religiosa protestante. Encuentran, así mismo, que los niveles de corrupción son más altos cuanto menor sea el nivel de ingreso promedio, cuanto mayor sea el grado de control estatal de la economía, más débiles sean las normas y las instituciones democráticas y menor sea el grado de integración en la economía mundial.

Interesados en estudiar y evaluar la literatura empírica sobre las fuentes de corrupción y hacer uso de la mejora en la disponibilidad de datos, Pellegrini y Gerlagh (2007) utilizaron un modelo econométrico con datos de diferentes países para poner a prueba hipótesis bien establecidas y más recientes sobre corrupción en forma conjunta. Encontraron, nuevamente, que haber sido en el pasado una colonia británica es un buen predictor de la corrupción y apoyan las teorías culturales sobre las causas de la corrupción, sugiriendo que una exposición de mediano o largo plazo a la democracia sin

interrupción se asocia con niveles de corrupción más bajos, mientras que la inestabilidad política hace que se incremente la corrupción.

Centrado en el debate sobre la pertinencia de los índices basados en la percepción como indicadores de la corrupción real de un país, Dreher, Kotsogiannis y McCorrison (2007) diseñan un modelo de ecuaciones estructurales que trata a la corrupción como una variable latente que está directamente relacionada con sus causas y efectos, para obtener un índice de corrupción. En el proceso, realizan un análisis de las diferentes causas económicas e institucionales de la corrupción dentro de las cuales incluyen al nivel de apertura o interrelación con los mercados internacionales.

Análogamente, algunos estudios tratan específicamente de la interrelación entre los niveles de integración con los mercados internacionales y la corrupción de los países. La principal hipótesis plantea que una medida de la presencia de competencia en los mercados es la integración de un país en la economía mundial. Si la competencia reduce la corrupción, una mayor apertura al comercio internacional y a la inversión debe estar acompañada de una menor corrupción.

Ades y Di Tella (1999) demuestran que la apertura, definida como la razón entre las importaciones y el PIB, se asocia negativamente con la corrupción, y encuentran resultados robustos cuando incluyen otras variables explicativas. Los autores concluyen que la competencia económica internacional reduce la corrupción. Un hallazgo similar es reportado por Gerring y Thacker (2005). Sin embargo, Treisman (2000) no encontró evidencia significativa utilizando el índice de transparencia internacional (CPI).

Otra medida posible del grado de competencia en un país puede ser el número de años que ha estado abierto al comercio internacional. Algunos estudios demuestran que el impacto de esta variable sobre la corrupción es significativamente negativo. Sin embargo Tavares (2007) afirma que el efecto inmediato de la liberalización depende de cuestiones institucionales y de la rapidez con la que se ejecuten las políticas aperturistas.

Más recientemente se ha presentado un giro copernicano y se ha desplazado el interés hacia el análisis de la influencia de la corrupción sobre las políticas comerciales. Por ejemplo, Bandyopadhyay y Roy (2007) controlan estadísticamente la endogeneidad de la corrupción con respecto a la política comercial y encuentran que la corrupción y la falta de cumplimiento de los contratos tienen un fuerte impacto en el aumento de la protección comercial y los efectos negativos de la apertura comercial.

Así mismo, Dutt y Traca (2007) analizan el papel de la corrupción como una barrera a las importaciones, poniendo de relieve la interacción con los aranceles nominales. Demuestran que el efecto de la corrupción no es lineal y que debe tenerse en cuenta el nivel de los aranceles para prever su influencia.

También, Jong y Bogmans (2010) investigan los efectos de la corrupción en el comercio internacional y comparan los resultados para diferentes países. Distinguen los efectos de la corrupción en una economía exportadora de los que se presentan en una economía importadora. Concluyen que la corrupción en general dificulta el comercio internacional, mientras que el pago de sobornos en las aduanas aumenta las

importaciones. Este efecto es más fuerte en los países importadores con aduanas ineficientes.

Dado que muchos de los resultados de investigaciones previas son ambiguos y solo incluyen la apertura en el mercado de bienes y servicios, sumado al hecho de que existen diferentes maneras de plantear la relación entre el nivel de apertura a los mercados internacionales y la corrupción, podemos afirmar, como lo hace Lambsdorff (2007, pág. 14), que nuestro conocimiento sobre las tendencias en los niveles de corrupción es todavía limitado y, sin duda, los resultados pueden requerir una mayor validación.

## **2.1 Red Neuronal Perceptrón Multicapa (MLP)**

Una Red Neuronal Artificial, RNA, desde ahora es básicamente un modelo simplificado del modo en que los sistemas nerviosos procesan información. Funciona en forma simultánea con un número de unidades simples de procesamiento interconectadas que emulan a las neuronas (llamadas también nodos), están organizadas en niveles denominados capas. Cada nodo está conectado con otros mediante enlaces de comunicación, cada uno de los cuales tiene asociado un peso. En los pesos se encuentra el “conocimiento” que tiene la RNA acerca de un determinado problema. Algunas de las redes neuronales son herramientas útiles en muchas aplicaciones de predicción en minería de datos debido a su potencia, flexibilidad y facilidad de uso.

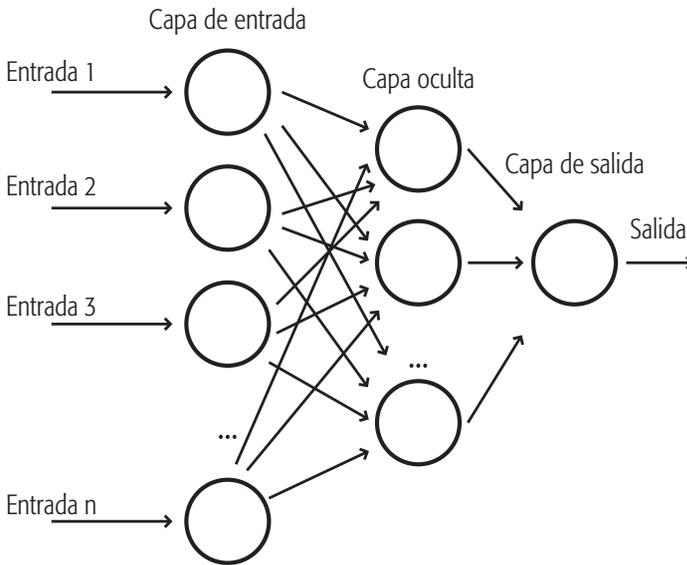
Una de las RNA más ampliamente utilizadas en el análisis de clasificación es el Perceptrón Multicapa (MLP por sus siglas en inglés). Rumelhart, Hinton y Williams (1986) formalizaron un método para que una red de este tipo aprendiera la asociación que existe entre un conjunto de patrones de en-

trada y sus salidas correspondientes. Este método se conoce como *backpropagation error* (propagación del error hacia atrás).

Un MLP está compuesto por una capa de entrada, una de salida y una o más capas ocultas; aunque se ha demostrado que para la mayoría de problemas es suficiente con una sola capa oculta Funahashi, 1989. En este tipo de modelos las conexiones entre nodos siempre van desde las neuronas de una determinada capa hacia las neuronas de la siguiente; no hay conexiones laterales ni hacia atrás. Por tanto, la información siempre se transmite desde la capa de entrada hacia la capa de salida. En el gráfico 1 se puede observar la arquitectura de un MLP.

Complementando, generalmente se considera  $W_{ij}$  como el peso de conexión entre la neurona de entrada  $i$  y la neurona oculta  $j$ , y  $V_{jk}$  como el peso de conexión entre la neurona oculta  $j$  y la neurona de salida  $k$ , estos pesos se observan como flechas que unen a los nodos en el gráfico 1.

Según Palmer, Montaña y Jiménez (2001), en el algoritmo *backpropagation* podemos considerar una etapa donde se presenta ante la red un patrón de entrada y este se transmite a través de las sucesivas capas de neuronas hasta obtener una salida y, por otro lado, una etapa de entrenamiento o aprendizaje en la que se modifican los pesos de la red de manera que coincida la salida deseada con la salida obtenida por la red ante la presentación de un determinado patrón de entrada.



**Fuente:** elaboración del autor.

**Gráfico 1**  
Arquitectura de una red MLP

El patrón de entrada  $p$  está expresado como un vector  $pX$ :  $x_{p1}, \dots, x_{p1}, \dots, x_{pN}$ , este se transmite a través de los pesos  $W_{ji}$  desde la capa de entrada hacia la capa oculta, la entrada neta que recibe una neurona oculta  $j$ ,  $net_{pj}$ , es:

$$net_{pj} = \sum_{(i=j)}^N w_{ji} x_{pi} + \theta_j$$

donde  $\theta$  es el umbral de la neurona que se supone como un peso asociado a una neurona ficticia con valor de salida igual a 1.

Las neuronas de esta capa intermedia transforman las señales recibidas mediante la aplicación de una función de activación (sigmoidea generalmente) y se obtiene así un valor de salida:

$$b_{pj} = f(\text{net}_{pj})$$

donde  $b_{pj}$  es el valor de salida de la neurona  $j$ .

Este valor se transfiere a través de los pesos  $V_{kj}$  hacia la capa de salida:

$$\text{net}_{pk} = \sum_{j=1}^L v_{kj} b_{pj} + \theta_k$$

En la capa de salida se aplica la misma operación que en la capa anterior, las neuronas de esta última capa proporcionan la salida,  $y_{pk}$ , de la red:

$$y_{pk} = f(\text{net}_{pk})$$

Luego, empieza una etapa de aprendizaje o entrenamiento de la RNA, el objetivo de este es minimizar el error entre la salida obtenida por la red y la salida deseada. Se dice que el aprendizaje en las redes mediante el algoritmo *backpropagation* y similares es de tipo supervisado, debido a que el usuario (o supervisor) determina la salida deseada ante la presentación de un determinado patrón de entrada (Palmer, Montaña,

La función de error que se intenta minimizar para cada patrón  $p$ ,  $E_p$ , está definida por:

$$E_p = \frac{1}{2} \sum_{k=1}^M (d_{pk} - y_{pk})^2$$

donde  $d_{pk}$  es la salida esperada para la neurona de salida  $k$  ante la presentación del patrón  $p$ . A partir de la anterior expresión se obtiene una ecuación general de error mediante:

$$E = \sum_{p=1}^p E_p$$

Como  $E_p$  es función de todos los pesos de la red, el gradiente de  $E_p$  es un vector igual a la derivada parcial de  $E_p$  respecto a cada uno de los pesos. El gradiente toma la dirección que determina el incremento más rápido en el error, mientras que la dirección opuesta determina el decremento más rápido en el error. Por tanto, el error puede reducirse ajustando cada peso en la dirección<sup>1</sup>:

$$- \sum_{p=1}^p \frac{\delta E_p}{\delta w_{ji}}$$

---

<sup>1</sup>Un análisis más concienzudo de este algoritmo puede estudiarse en Rumelhart, Hinton y Williams (1986); un estudio generalizado sobre las RNA y sus aplicaciones en ciencias sociales se puede encontrar en Palmer, Montaña y Jiménez (2001) y una visión más práctica para el uso de software en Pères (2008).

Para la modificación de los pesos de la red, estos se actualizan después de haber presentado todos los patrones de entrenamiento; este proceso se denomina aprendizaje por lotes o modo *batchy*, es el modo usual de proceder. Además, se han desarrollado diferentes variantes del algoritmo original *backpropagation* para realizar la etapa de aprendizaje, las cuales tienen por objeto acelerar el proceso de aprendizaje.

### 3. LAS VARIABLES

El nivel de apertura económica no es una variable tan fácil de medir como se podría pensar, por lo tanto, se deben tener en cuenta tres cuestiones principales:

**i) Lo que se quiere medir:** tener claro si lo que se pretende medir es el nivel de interacción con los mercados internacionales per se, las posibilidades reales (libertad) de los agentes de un país para comerciar con el extranjero o una combinación de ambas. En el primer caso, la variable más utilizada es la razón entre componentes de la Balanza Comercial y el PIB u otras similares. En el segundo caso generalmente se utilizan índices que mezclan indicadores sobre políticas aperturistas o proteccionistas (Sachs & Warner, 1995). Muy pocos estudios logran unir los dos criterios.

**ii) Los mercados:** saber qué mercados se incluirán en la medición del nivel de apertura. En la mayoría de estudios de este tipo se toma únicamente el mercado de bienes y servicios.

**iii) Lo temporal:** si se quiere medir el nivel de apertura económica de un país en un momento determinado del tiempo y su relación con otras características se utilizan datos de corte transversal; cuando el interés del estudio se centra en

su influencia histórica sobre otras variables, generalmente se utiliza el tiempo que se ha estado abierto a los mercados internacionales o datos en panel y promedios. Asimismo, cuando interesa la influencia de los cambios en la política aperturista se suele utilizar la tasa de variación de algún indicador con respecto al tiempo.

Teniendo en cuenta lo anterior, en este trabajo se entiende por apertura económica un conjunto de políticas de carácter económico y legal que incentivan el ejercicio del comercio y las finanzas internacionales, la liberación del mercado de capitales y el intercambio de tecnologías sin restricciones por parte del Estado. Para cuantificarla se genera la variable “GTRADEFINANCFREED”, calculada a partir de la media geométrica<sup>2</sup> de los índices de libertad comercial y libertad financiera publicados por The Heritage Foundation y The Wall Street Journal. Este consiste en un índice que toma valores de 0 a 100, donde los niveles de mayor apertura son los más cercanos a 100; y se calcula para 155 países correspondientes al año 2010.

La libertad de comercio es, más específicamente, una medida compuesta por la ausencia de barreras arancelarias y no arancelarias que afectan las importaciones y exportaciones de bienes y servicios. Los puntajes del índice de libertad financiera constan de aspectos como el alcance de la regulación gubernamental de los servicios financieros, el grado de intervención estatal en las entidades financieras, el grado de desarrollo de los mercados financieros internacionales y de capitales, y la apertura a la competencia extranjera. Los primeros forman parte

---

<sup>2</sup> Se puede demostrar fácilmente que la media geométrica es más eficiente que la media aritmética para promediar tasas o porcentajes.

de un índice más general de libertad económica que tiene en cuenta otros aspectos económicos e institucionales.

Por otra parte, esta investigación tiene como objetivo

cuarta corresponde al cuartil de los países con menor percepción de corrupción y la clase 1 al cuartil con mayor nivel de percepción de corrupción para el año 2010. Hay que aclarar que como los cuartiles fueron creados teniendo en cuenta el ranking y no los puntajes, puede haber cuartiles con diferente número de países.

Dados los objetivos propuestos, es necesario controlar la influencia de otras variables socio-económicas, para esto se toma el Índice de Desarrollo Humano (IDH) del 2010 para los países estudiados como proxy de su nivel de pobreza y calidad de vida. Este índice es elaborado por el Programa de las Naciones Unidas para el Desarrollo (PNUD) y contiene los indicadores de esperanza de vida al nacer, tasa de alfabetización de adultos y la tasa bruta combinada de matrícula en educación primaria, secundaria y superior, así como los años de duración de la educación obligatoria. También incluye el PIB per cápita (PPA) en dólares internacionales. El IDH toma valores entre 0 y 100, donde 100 es el mayor puntaje posible en desarrollo humano.

Otra variable incluida debido a su importancia teórica es el tamaño del Estado<sup>3</sup>. Se supone que entre mayor sea el tamaño de este con respecto al total de la economía, mayores niveles de corrupción deben observarse. Para este trabajo la variable se denomina “GOVSPENDING” y corresponde al índice de gastos del Gobierno, publicado por The Heritage Foundation y The Wall Street Journal en el año 2010.

Este índice se basa en el nivel de gasto público, incluido el consumo del Gobierno y las transferencias como porcentaje

---

<sup>3</sup> Puede encontrarse una lista de estudios realizados sobre la interacción entre corrupción y tamaño del Estado en Lambsdorff, 2007, pág. 4.

del PIB. No trata de identificar un nivel óptimo de gasto público, ya que el ideal puede variar de país a país, dependiendo de factores culturales, geográficos y de desarrollo. Según la ficha metodológica, la escala de puntuación de este índice no es lineal, lo que significa que los países que se acercan a cero en su participación de gasto público con respecto al PIB son ligeramente penalizados, mientras que los niveles de gasto público que exceden el 30% del PIB reciben resultados mucho peores en forma cuadrática en este índice, de modo que solo los gobiernos realmente grandes reciben puntuaciones muy bajas.

Por último, se incluye una variable de control sin interés teórico y sin esperar relación alguna: el nivel o tasa de urbanización de cada país “URBANIZACION” en el año 2010, tomando valores entre 0 y 100. El cuadro 1 (ver pág 64) muestra un resumen de las variables utilizadas:

**Tabla 1**  
Variables utilizadas

VARIABLE	DESCRIPCIÓN	FUENTE
GTRADEFINANCFREED	Media geométrica de los índices de libertad comercial y libertad financiera publicados por The Heritage Foundation y The Wall Street Journal 2010.	Cálculo del autor.
CATEGIPC	Variable categórica correspondiente a cuartiles de acuerdo con los niveles de corrupción según la posición (ranking) de cada país en el CPI, publicado por Transparencia Internacional 2010.	Cálculo del autor a partir del CPI publicado por Transparencia Internacional.
IDH	Índice de Desarrollo Humano 2010.	Programa de las Naciones Unidas para el Desarrollo (PNUD).
GOVSPENDING	Índice de gastos del Gobierno 2010.	Publicado por The Heritage Foundation y The Wall Street Journal en el año 2010.
URBANIZACIÓN	Nivel o tasa de urbanización de cada país para el año 2010.	Base de datos del Banco Mundial.

**Fuente:** elaborado por el autor.

#### 4. METODOLOGÍA

En este trabajo se analiza una muestra de 155 países para los cuales existen datos correspondientes al año 2010 para todas las variables propuestas.

Se pretende generar un modelo predictivo que pronostique el grupo al que pertenece cada país según su nivel de corrupción y a partir del CPI de Transparencia Internacional y de determinadas características socio-económicas que delimiten su perfil. Este tipo de modelos se denominan predictivos de clasificación, los cuales permiten agrupar los elementos de interés en dos o más categorías diferentes representadas por una variable no métrica; el perfil se genera con base en los va-

lores de un conjunto de variables independientes combinadas funcionalmente.

Por lo tanto, los modelos de clasificación se diferencian de los modelos de regresión con variables dependientes categóricas en que los primeros predicen la probabilidad de pertenencia de un elemento a una clase específica, dado un conjunto de características que comparte con los demás miembros de esa categoría; mientras que los segundos predicen la probabilidad de pertenecer a una clase para valores dados de las variables independientes, lo que presupone algún nivel de causalidad a priori.

No obstante su interés para este tipo de estudios, los modelos predictivos de clasificación clásicos parten de supuestos muy rigurosos sobre la distribución de los datos y la naturaleza de las variables independientes que limitan de sobremaneira su aplicabilidad en problemas económicos; y exigen, entre otras propiedades, la normalidad multivariante de los datos, homogeneidad de matrices de varianza-covarianza, linealidad y ausencia de multicolinealidad. Además, son muy sensibles al tratamiento de valores perdidos o extremos<sup>4</sup>.

En contraste, las Redes Neuronales MLP poseen una serie de características deseables, tales como la habilidad para procesar datos con ruido o incompletos, la alta tolerancia a fallos, lo que le permite a la red operar satisfactoriamente con neuronas o conexiones dañadas, y la capacidad de funcionar como modelos no paramétricos y robustos a supuestos. Asimismo, su amplia aplicabilidad se debe a que es capaz de actuar como un aproximador universal de funciones (Funa-

---

<sup>4</sup> Un acercamiento matemático del análisis discriminante y otros métodos clasificatorios puede encontrarse en Peña, 2002.

hashi, 1989); más concretamente, una red MPL con al menos una capa oculta con suficientes unidades no lineales puede aprender cualquier tipo de función o relación continua entre un grupo de variables de entrada y de salida.

Estas propiedades convierten a las redes Perceptrón Multicapa en herramientas de propósito general, flexibles y no lineales; mostrando un rendimiento superior respecto a los modelos estadísticos clásicos en numerosos campos de aplicación. Entre estos la investigación económica, en la que es más probable que en los datos y en las series aparezcan relaciones no lineales que lineales (Granger & Teräsvirta, 1993), como las exigidas por los modelos de clasificación tradicionales.

Por lo anterior, para responder las preguntas planteadas en este estudio, se aplica un modelo predictivo de clasificación para las variables tipificadas, basado en redes neuronales artificiales Perceptrón Multicapa con una capa oculta de cuatro nodos, proceso de entrenamiento *batch* y con un algoritmo de optimización de gradiente conjugado (Battiti, 1992), el cual se deriva del algoritmo *backpropagation*. Se fundamenta en el cálculo de la segunda derivada del error con respecto a cada peso, y en obtener el cambio a realizar en los pesos a partir de este valor y el de la primera derivada. Se utilizan funciones de activación logísticas en las salidas de la red, que son utilizadas como una función discriminante no lineal (Biganzoli, Boracchi, Mariani, & Marubini, 1998).

En primer lugar, se determina si el modelo predice correctamente el grupo de nivel de corrupción al que pertenecerá un país dadas las variables planteadas. En segundo lugar, la importancia relativa del nivel de apertura económica con respecto a las demás variables en la clasificación se estima a través de un análisis de sensibilidad que calcula la importancia

de cada predictor en la determinación de la red neuronal; el análisis se basa en el entrenamiento combinado y en las muestras de prueba, evidenciando la importancia y la importancia normalizada para cada predictor. Por último, se determina la importancia que tienen las variables explicativas sobre la variable de respuesta mediante un análisis de sensibilidad numérica (NSA) propuesto por Montaña y Palmer (2002), que se basa en el cálculo de las pendientes que se forman entre entradas y salidas, sin realizar ningún supuesto acerca de la naturaleza de las variables y respetando la estructura original de los datos.

La tabla 2 muestra una descripción de la red entrenada:

**Tabla 2**  
Información sobre la red

Capa de entrada	Covariables	1	IDH POBREZA
		2	GTRADEFINANCFREED
		3	GOVSPENDING
		4	URBANIZACIÓN
	Número de unidades		4
Método de cambio de escala de las covariables		Tipificados	
Capas ocultas	Número de capas ocultas		1
	Número de unidades de la capa oculta 1a		4
	Función de activación		Sigmoidea
Capa de salida	Variables dependientes	1	CUARTILES CORRUPCIÓN
	Número de unidades		4
	Función de activación		Sigmoidea
	Función de error		Suma de cuadrados

**Fuente:** elaborado por el autor.

## 5. RESPUESTAS A LAS PREGUNTAS PLANTEADAS

### a) Características socio-económicas comunes a los países con niveles similares de corrupción

La tabla 3 muestra la clasificación realizada por la red artificial tanto en la etapa de entrenamiento como en la de prueba. Para la primera se tomó una muestra aleatoria correspondiente al 60% de los países para entrenar la red y se corroboró su aprendizaje y capacidad predictiva con el 40% restante. Se observa que al incluir todas las variables el MLP fue capaz de clasificar correctamente el 64,6% de los países. Se obtuvo, como era de esperarse, mejores resultados para los cuartiles 1, con un 90,9% y 4, con 100%, correspondientes a los países con mayor y menor nivel de corrupción, respectivamente, en contraste con los cuartiles centrales, cuyo poder de clasificación no es el deseado.

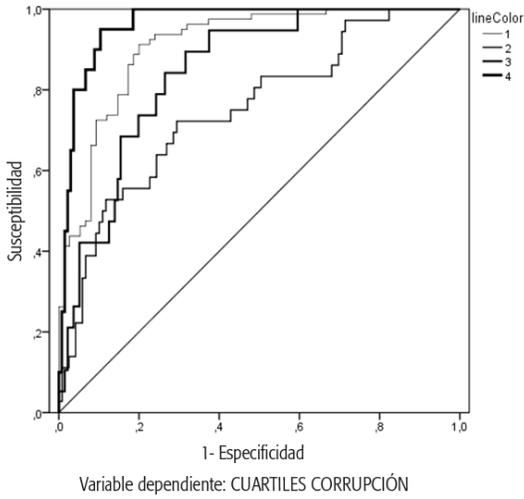
Lo anterior implica que los países con niveles extremos de corrupción presentan características muy comunes relacionadas con las variables explicativas propuestas, las cuales determinan sus niveles de corrupción.

**Tabla 3**  
 Clasificación en cuartiles de corrupción por MLP  
 incluyendo todas las variables independientes

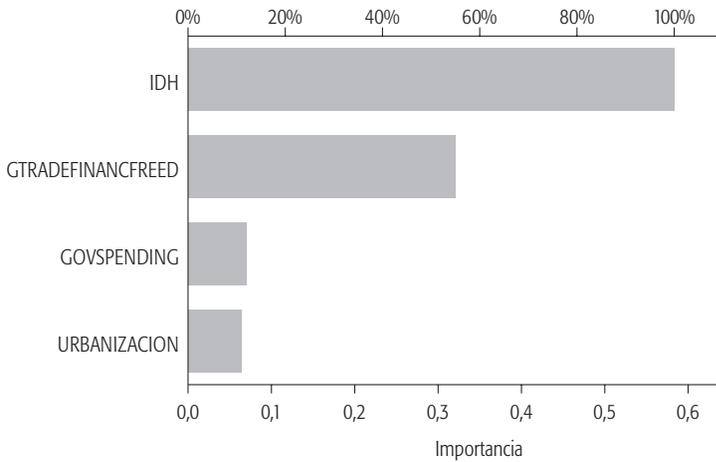
Muestra	Observado	Pronosticado				
		1	2	3	4	Porcentaje correcto
Entrenamiento	1	45	1	0	1	95,7%
	2	7	10	0	2	52,6%
	3	2	2	0	7	0,0%
	4	0	1	0	12	92,3%
	Porcentaje global	60,0%	15,6%	0,0%	24,4%	74,4%
Prueba	1	30	3	0	0	90,9%
	2	9	5	0	3	29,4%
	3	1	2	0	5	0,0%
	4	0	0	0	7	100,0%
	Porcentaje global	61,5%	15,4%	0,0%	23,1%	64,6%

**Fuente:** resultados del modelo.

El gráfico 2 muestra el poder predictivo de la red entrenada, a través de las curvas COR que ejemplifican la relación existente entre la proporción de verdaderos positivos (susceptibilidad: países bien clasificados de acuerdo a su perfil) y la proporción de falsos positivos (especificidad: países mal clasificados de acuerdo con su perfil), para cada uno de los cuartiles correspondientes a niveles de corrupción. Se observa que las curvas 1 y 4 están más alejadas de la diagonal, lo que corrobora lo planteado anteriormente.



**Gráfico 2**  
Curvas COR



**Gráfico 3**  
Importancia relativa de las  
variables independientes

Una vez demostrada la pertinencia de la red, en el gráfico 3 se aprecia que el nivel de apertura económica es relevante como variable clasificadora, lo que significa que los países con diferentes niveles de corrupción comparten niveles semejantes de apertura económica.

### **b) Importancia del nivel de apertura económica como variable clasificadora de los países según sus niveles de corrupción**

Complementando el análisis anterior, en el gráfico 3 se observa que el nivel de apertura económica representado por la variable “GTRADEFINANCFREED” es el segundo en importancia como clasificador luego del Índice de Desarrollo Humano (IDH). Lo anterior revela que los países que comparten políticas similares en cuanto a libertad comercial y financiera tienen características similares en sus niveles de corrupción.

Por otro lado, la poca importancia que parece presentar el nivel de gasto del gobierno o tamaño del Estado se debe a que se incluyen tanto los países desarrollados como los países en vía de desarrollo en una misma base de datos. Cuando se alimentó la red solo con la variable “GTRADEFINANCFREED” se obtuvo un 0% de correctas discriminaciones para los cuartiles 2 al 4, pero un 95,2% de los países en el cuartil 1 fue correctamente clasificado y la relación entre las variables fue negativa. Esto corrobora las hipótesis que plantean que los países más corruptos comparten instituciones gubernamentales más ineficientes que provocan que entre mayor sea el tamaño del Estado mayor sea el nivel de corrupción percibida. Como más adelante se verá, el IDH muestra que estos países se caracterizan por ser los más alejados del desarrollo, por lo

tanto, los países más corruptos (cuartil 1) tienen menores niveles de desarrollo y gobiernos grandes e ineficientes; mientras que el tamaño del Estado pierde importancia en otros niveles de corrupción.

La tabla 4 muestra la importancia relativa de la apertura económica con respecto a las otras variables clasificadoras lo que confirma lo anteriormente planteado.

**Tabla 4**  
Importancia de las variables independientes

VARIABLES	Importancia	Importancia normalizada
IDH pobreza	0,58	100,0%
GTRADEFINANCFREED	0,30	51,7%
GOVSPENDING	0,06	10,7%
URBANIZACIÓN	0,06	10,3%

**Fuente:** resultados del modelo.

Por otra parte, debido a la poca importancia relativa que muestran los niveles de urbanización y el tamaño del Estado para la clasificación de países por sus niveles de corrupción, se entrenó otra red MLP con las mismas características anteriores, pero solo incluyendo el nivel de desarrollo humano y de apertura económica como variables discriminantes.

**Tabla 5**  
 Clasificación en cuartiles de corrupción por MLP incluyendo solamente los niveles de IDH y apertura económica como variables predictoras

Muestra	Observado	Pronosticado				
		1	2	3	4	Porcentaje correcto
Entrenamiento	1	50	4	1	0	90,9%
	2	8	10	0	1	52,6%
	3	2	2	3	6	23,1%
	4	0	2	1	11	78,6%
	Porcentaje global	59,4%	17,8%	5,0%	17,8%	73,3%
Prueba	1	24	1	0	0	96,0%
	2	6	7	1	3	41,2%
	3	1	1	3	1	50,0%
	4	0	0	0	6	100,0%
	Porcentaje global	57,4%	16,7%	7,4%	18,5%	74,1%

**Fuente:** resultados del modelo.

En la tabla 5 se observa que al entrenar la red con estas dos variables, esta fue capaz de identificar relaciones funcionales no lineales que clasificaron correctamente el 74,1 de los países en sus correspondientes niveles de corrupción, se obtuvieron mejores resultados predictivos incluso para los cuartiles intermedios.

**Tabla 6**

Clasificación en cuartiles de corrupción por MLP incluyendo solamente los niveles de apertura económica como variable independiente

Muestra	Observado	Pronosticado				
		1	2	3	4	Porcentaje correcto
Entrenamiento	1	39	7	0	0	84,8%
	2	14	10	0	0	41,7%
	3	1	9	0	0	0,0%
	4	3	10	0	0	0,0%
	Porcentaje global	61,3%	38,7%	0,0%	0,0%	52,7%
Prueba	1	33	1	0	0	97,1%
	2	7	5	0	0	41,7%
	3	4	5	0	0	0,0%
	4	1	6	0	0	0,0%
	Porcentaje global	72,6%	27,4%	0,0%	0,0%	61,3%

**Fuente:** resultado del modelo.

En la tabla 6 se muestra el resultado de entrenar la red MLP con el nivel de apertura como única variable independiente. Se observa que esta característica es más relevante específicamente para clasificar a los países más corruptos de manera adecuada mientras que es irrelevante para los dos cuartiles menos corruptos. Lo anterior indica, como se confirmará en el siguiente punto, que los países más corruptos tienen como una de sus características comunes el ser muy cerrados a los mercados internacionales; y que esta característica no es tan discriminante para países menos corruptos.

### c) Signo de la relación

Ya se demostró anteriormente que los niveles de apertura económica de un país lo caracterizan y permiten clasificarlo, junto con otras variables y con cierto grado de certidumbre, según su nivel de corrupción. No obstante, este análisis estaría incompleto si no se estimara el signo de esta relación.

Por esto se realizó un análisis de sensibilidad numérica (NSA) para determinar la relación entre los niveles de apertura comercial y desarrollo económico, y la probabilidad de pertenecer a niveles clasificatorios más corruptos. Se observó que existe una marcada relación positiva entre la clasificación en cuartiles superiores y ambas variables clasificadoras. Esto indica que entre mayor sea el nivel de desarrollo de un país (por tanto menor pobreza y mayor educación y calidad de vida entre otras características) y mayor sea la integración a los mercados internacionales menor será la probabilidad de ser clasificado como un país corrupto. La tabla 7 muestra los signos correspondientes a estas relaciones.

**Tabla 7**

Relación entre las variables independientes y la pertenencia a cuartiles con menores niveles de corrupción

VARIABLE	SIGNO MEDIANTE NSA
IDH pobreza	POSITIVO (+)
GTRADEFINANCFREED	POSITIVO (+)

**Fuente:** elaborado por el autor a partir de resultados del modelo.



instituciones gubernamentales más ineficientes que provocan que entre mayor sea el tamaño del Estado mayor sea el nivel de corrupción percibida.

En conclusión, la relación entre los niveles de corrupción interna y las variables planteadas muestra que los países más corruptos se caracterizan por tener menores niveles de desarrollo, ser más cerrados a los mercados internacionales y presentar gobiernos grandes e ineficientes. Lo que corrobora en gran medida los planteamientos teóricos. El nivel de apertura económica es una de las características más importantes que tienen en común los países que comparten niveles similares de corrupción.

### Referencias bibliográficas

- Ades, A., & Di Tella, R. (Sep de 1999). Rents, Competition, and Corruption. *The American Economic Review*, 89(4), 982-993.
- Alesina, A., & Weder, B. (2002). Do Corrupt Governments Receive Less Foreign Aid? *The American Economic Review*, 92(4), 1126-1137.
- Bandyopadhyay, S., & Roy, S. (Mayo de 2007). *Corruption and Trade Protection: Evidence from Panel Data*. Obtenido de <http://research.stlouisfed.org/wp/2007/2007-022.pdf>
- Battiti, R. (1992). First and Second Order Methods for Learning: Between Steepest Descent and Newton's Method. *Neural Computation*, 4(2), 141-166.
- Berg, A., & Krueger, a. (septiembre de 2002). Dar vela al comercio. *Finanzas y desarrollo*, 16-19.
- Biganzoli, E., Boracchi, P., Mariani, L., & Marubini, E. (1998). Feed-forward Neural Networks for the Analysis of Censored Survival Data: a Partial Logistic Regression Approach. *Statistics in Medicine*, 17(10), 1169-1186.

- Boehm, F., & Lambsdorff, J. G. (2009). Corrupción y anticorrupción: una perspectiva neo-institucional. *Revista de Economía Institucional*, 45-72.
- Dreher, A., Kotsogiannis, C., & McCorriston, S. (2007). Corruption Around the World: Evidence from a Structural Model. *Journal of Comparative Economics* (35), 443-466.
- Dutt, P., & Traca, d. (2007). Corruption and Bilateral Trade Flows: Extortion or Evasion? *Insead*.
- Friedman, M. (1962). *Price Theory*. Chicago: Aldine Publishing.
- Funahashi, K. (1989). On the Approximate Realization of Continuous Mappings by Neural Networks. *Neural Networks*, 2, 183-192.
- Gerring, J., & Thacker, S. C. (2005). Do Neoliberal Policies Deter Political Corruption? *International Organization*, 59, 233-254.
- Granger, C., & Teräsvirta, T. (1993). Modelling Nonlinear Economic Relationships. *Advanced Texts in Econometrics*. Oxford University Press.
- Jong, E. d., & Bogmans, C. (2010). *Does Corruption Discourage International Trade?* Obtenido de doi:10.1016/j.ejpoleco. 2010. 11.005
- Klitgaard, R. (1988). Controlling Corruption. *Berkeley and Los Angeles: University of California*, 32.
- Lambsdorff, J. G. (2007). *The Institutional Economics of Corruption and Reform. Theory, Evidence and Policy*. Passau, Alemania: Hardback.
- Montaño, M. J., & Palmer, A. (2002). Redes neuronales artificiales: abriendo la caja negra. *Metodología de las ciencias del comportamiento*, 4(1), 77-93.
- Palmer, A., Montaño, J., & Jiménez, R. (2001). Tutorial sobre redes neuronales artificiales: el Perceptrón Multicapa. *Revista electrónica de psicología*.
- Pellegrini, L., & Gerlagh, R. (2007). Causes of Corruption: a Survey of Cross-country Analyses and Extended Results. *Economics of Governance*.
- Peña, D. (2002). *Análisis de datos multivariantes*. Madrid: McGraw-Hill.

- Péres, C. (2008). *Econometría avanzada. Técnicas y herramientas*. Madrid: Pearson.
- Rose-Ackerman, S. (1999). Corruption and Government: Causes, Consequences, and Reform. *Cambridge University Press*.
- Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning Internal Representations by Error Propagation. (D. R. (Eds.), Ed.) *Cambridge, MA: MIT Press*, 318-362.
- Sachs, J. D., & Warner, A. M. (1995). Economic Reform and the Process of Global Integration. *Brooking Papers on Economic Activity*, 1-95.
- Sandholtz, W., & Koetzle, W. (2000). Accounting for Corruption: Economic Structure, Democracy, and Trade. *International Studies Quaterly*, 44, 31-50.
- SPSS Neural Networks 17,0 (2007). Manual del usuario. Chicago.
- Svensson, J. (2005). Eight Questions About Corruption. *Journal of Economic Perspectives*, 19(3), 19-42.
- Tanzi, V. (1998). Corruption Around the World Causes, Consequences, Scope, and Cures. *IMF Staff Papers*, 45(4), 559-594.
- Tavares, S. C. (2007). Do Rapid Political and Trade Liberalizations Increase Corruption? *European Journal of Political Economy*, 23, 1053-1076.
- Treisman, D. (2000). The Causes of Corruption: a Cross-national Study. *Journal of Public Economics*, 76, 399-457.