

Propuesta de Investigación:

El actual nivel del desarrollo de Latinoamérica y su proyección a futuro por intermedio de
herramientas modernas de la ciencia de los datos

Rovai, Marcelo José

Universidad del Desarrollo – UDD

Escuela de Ingeniería – Instituto de Data Science

Nota del autor

Esta Propuesta de Investigación es parte del proceso de evaluación del tramo de
Macroeconomía y Ciencia de los Datos - 2018, ministrado por la profesora Mercedes Haga
para el Magister en Data Science de la Escuela de Ingeniería de la Universidad del Desarrollo
– UDD – Campus de Santiago, Chile.

Resumen

¿Cuales son las causas del *subdesarrollo* de Latinoamérica y las variables macroeconómicas que lo miden? ¿Por que la región no consigue ingresar en el bloque de los llamados países *desarrollados*? ¿Como se puede vencer esta barrera? Este trabajo pretende, por intermedio de la analice de datos, utilizándose de herramientas modernas del *data science* como *machine learning*, comparar los principales factores comunes a algunos países selectos de Latinoamérica, sus comparables en otros continentes y del primero mundo (los *países desarrollados*) explorando sus actuales condiciones y posibles proyecciones a futuro. El resultado de este trabajo será la propuesta de un manejo mas adecuado de las variables (o políticas) macroeconómicas que llevarían los países a se desarrollar mas rápidamente.

Palabras clave: macroeconomía, desarrollo, subdesarrollo, *data science*, *machine learning*

Propuesta de Investigación:

El actual nivel del desarrollo de Latinoamérica y su proyección a futuro por intermedio de herramientas modernas de la ciencia de los datos

Identificación y descripción del problema macroeconómico

Hay muchos factores que deben tenerse en cuenta al tratar de entender las causas del subdesarrollo de los latinoamericanos, como comenta Tsegaamlak, Asfaw Keraga¹. Además, cuando se observa los diversos indicadores macroeconómicos de un país, se concluye que hay mucha confusión (o controversia) sobre cómo clasificar a los países por su grado de desarrollo. ¿Cuales son los países desarrollados¹? ¿Cuales son subdesarrollados²? ¿Cuáles no encajan ni en una cosa o en otra, como aquellos en *vía de desarrollo o emergentes*³?

Por lo tanto, este trabajo va a tratar de entender cuáles son los principales indicadores o criterios de clasificación utilizados hoy en día, tales como el PIB⁴, el PNB⁵, además del IDH⁶, las tasas de alfabetización, nivel de renta y longevidad, que son los adoptados pela OCDE⁷ en su intento de clasificar los grados de desarrollo de los países.

Además de los índices más comunes ya descritos, también se intentará cuantificar informaciones más subjetivas, como los aspectos históricos de la colonización de los países latinoamericanos frente a sus equivalentes asiáticos y africanos; tipos de regímenes políticos que se enfrentaron a partir de sus procesos de independencia, entre otros.

Y para completar las variables a ser utilizadas en el trabajo, se cruzarán datos históricos de los países con índices más específicos, tales como inflación y desempleo.

¹ Recuperado de: https://es.wikipedia.org/wiki/Pa%C3%ADs_desarrollado

² Recuperado de: <https://es.wikipedia.org/wiki/Subdesarrollo>

³ Recuperado de: https://es.wikipedia.org/wiki/Pa%C3%ADs_en_v%C3%ADa_de_desarrollo

⁴ Producto Interno Bruto, utilizado por el Banco Mundial.

⁵ Producto Nacional Bruto, utilizado por la ONU.

⁶ Índice de Desarrollo Humano también de la ONU.

⁷ Organización para la Cooperación y el Desarrollo Económicos.

Una vez en posesión de todos estos *parámetros* anuales (desde la segunda mitad del siglo XX hasta la fecha lo más reciente posible), los mismos serán utilizados como *features* en el entrenamiento de un algoritmo de clasificación a ser aplicado a los países. Las etiquetas por aplicar a estos datos pueden ser, por ejemplo: *subdesarrollados*, *emergentes* y *desarrollados*⁸. La ilustración 1, ejemplifica lo descrito anteriormente.

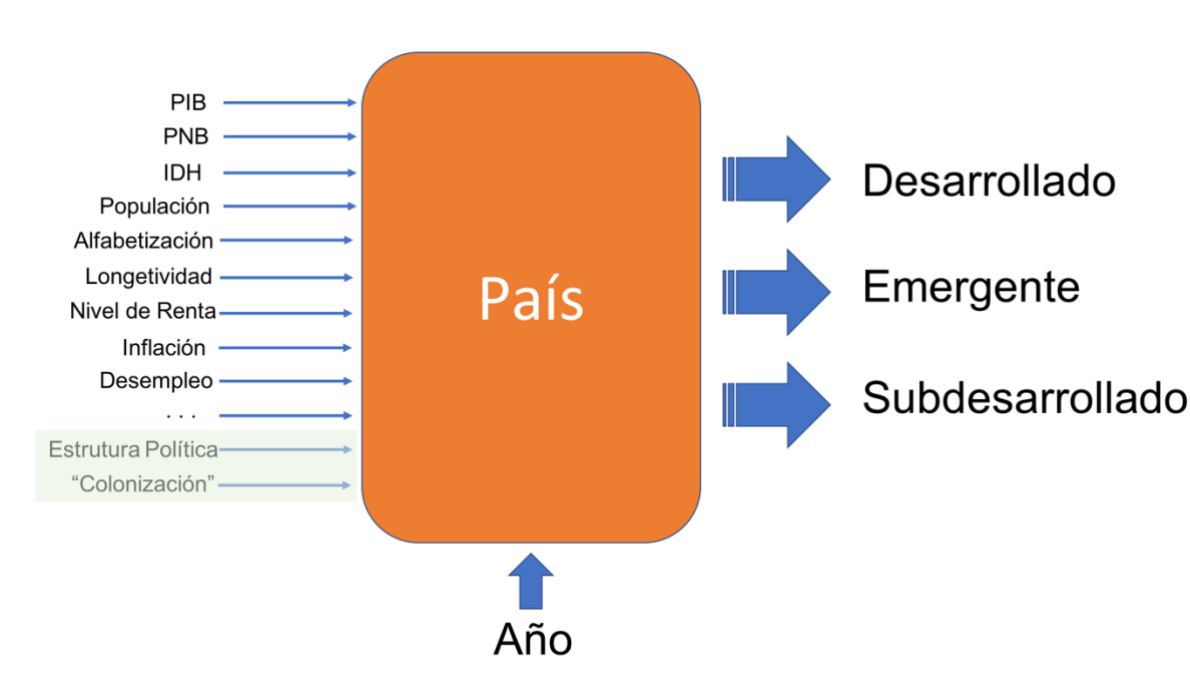


Ilustración 1 - ML aplicado a la clasificación de los países

La evolución de la Macroeconomía con el *Data Science*

Hoy en día se utiliza para series de datos macroeconómicos, técnicas de previsión basadas en redes neuronales, como muestran Cook, Thomas R., and Aaron Smalter Hallⁱⁱ. Pero, tradicionalmente, los trabajos de proyección de esos índices se obtienen de manera individual, es decir, uno por uno y con pocos cruces de parámetros, como se puede observar en el ejemplo de la figura 2, donde una proyección a partir del 2016 del desempleo en

⁸ Tenga en consideración que la decisión de la asignación de estos *labels* es un poco arbitraria, como ya mencionado y por tanto la clasificación a ser la adoptada acá y utilizada al largo de todo el análisis, podrá o no seguir una clasificación tradicional.

América Latina es estimado pelo IMF (Fondo Monetario Internacional) y graficado por Rovai, Macelo ⁱⁱⁱ.

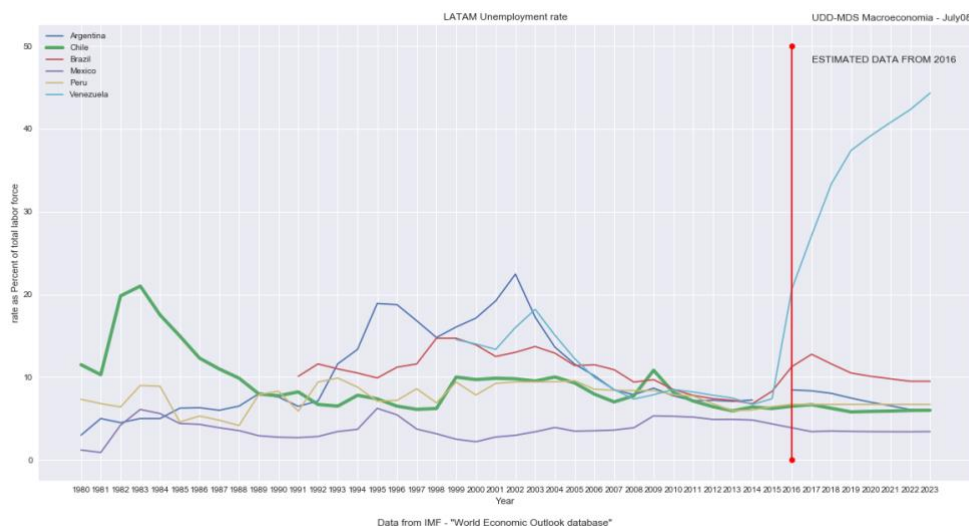


Ilustración 2 - Tasa de desempleo - Latam (select)

A partir del uso de herramientas más modernas de Data Science (DS), como los algoritmos de aprendizaje de máquina (ML), es posible hacer previsiones más precisas a partir de mayores volúmenes de datos y múltiples parámetros. Así, tomándose un gran número de países, cada uno con su conjunto de datos históricos y teniendo en consideración su estado de desarrollo en cada uno de los años específicos, sería posible en teoría, estimar cuando un país pasaría de subdesarrollado a desarrollado.

Como explicado por Woloszkoa, Nicolas ^{iv} “A medida que las capacidades del aprendizaje automático continúen aumentando, también lo harán las oportunidades para que los algoritmos funcionen junto con la teoría económica para hacer proyecciones del PIB y evaluar el impacto de las políticas. Cerrar la brecha entre la economía y el aprendizaje automático es un desafío científico que creo que aportará nuevos conocimientos a los problemas clave de política”.

Así, que al aplicar técnicas de ML, podría ser posible resolver el problema detectado, pues una vez que el entrenado el algoritmo y obtenido un modelo, se podrían hacer estudios que implican dos hipótesis:

1. Para un determinado país de Latinoamérica, el modelo podría ser alimentado con un conjunto de *features* estimado para un año futuro, determinando la probabilidad de que el sea considerado *desarrollado*.
2. Encontrar sobre la base de datos históricos del país, en qué año pasaría a ser considerado *desarrollado*.

En posesión de estos resultados, se podrían desarrollar propuestas para una gestión más apropiada de políticas macroeconómicas (que por su vez actúan directamente en los *features*) que llevarían a los países a desarrollarse más rápidamente.

Etapas para el desarrollo del trabajo

1. Definición de un grupo selecto de países a ser estudiados, englobando al menos Latinoamérica, miembros del OCDE, África, Oceanía y Asia. Una lista preliminar puede ser observada a seguir:
 - Argentina, Brasil, Chile, Colombia, Venezuela, México
 - Portugal, España, EE. UU, Reino Unido, Australia
 - Republica Democrática del Congo, Camarones
 - Singapur, China, Corea del Sur, Japón, China, India, Rusia
2. Estudio en mas profundidad del grupo definido en la etapa 1, cruzando datos históricos, como por ejemplo lo observado en la ilustración 3, construida a partir de datos del Banco Mundial por Rovai, Marcelo ^v.

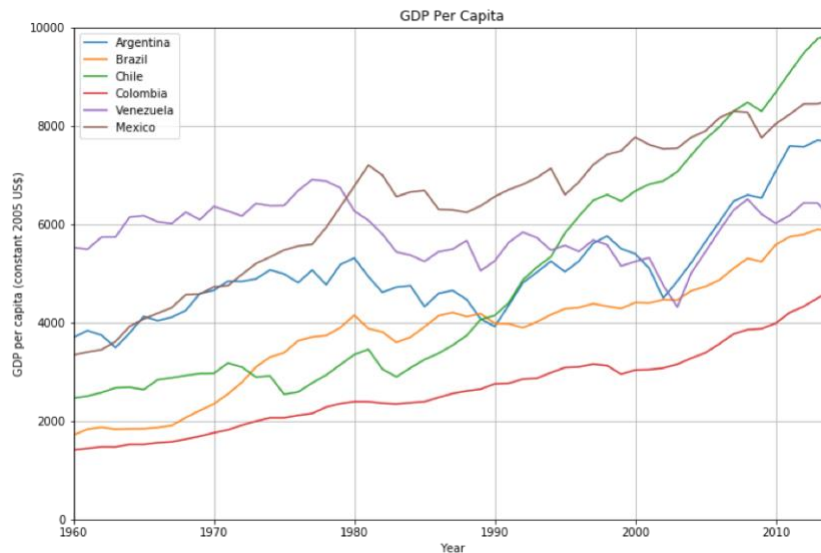


Ilustración 3 - Tasa de crecimiento GDP per capita - Latam (select)

3. Cruzamiento de diferentes indicadores, buscando “padrones” que puedan llevar a diferentes situaciones o niveles de desarrollo como se muestra en la ilustración 4, donde se examina la dependencia entre el crecimiento del GDP per capita de Chile y el de Venezuela, llevándose en consideración los gastos del gobierno versus sus exportaciones.

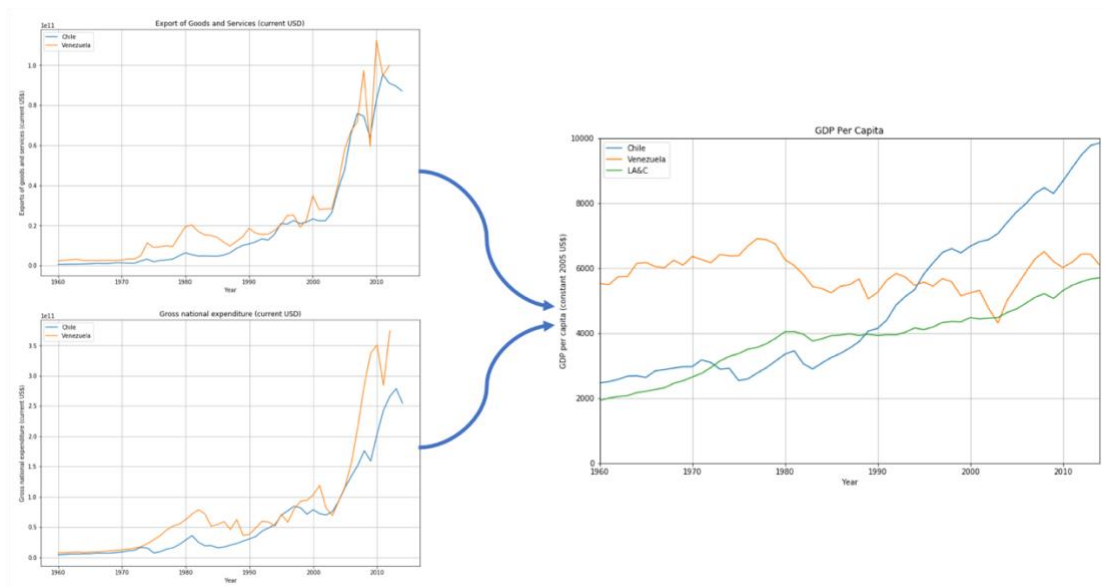


Ilustración 4 - crecimiento del GDP per capita de Chile y Venezuela

4. Estudios del pasado histórico y político de los países, definiendo *features* relacionados a esos eventos. Es muy posible que esta sea ser una de las etapas más complejas del trabajo, por lo que se debe reservar un tiempo significativo a ese modelaje.
5. Para cada país y su conjunto de *features* definidos en la etapa anterior, se deberá asociar un *label* a este particular momento de la historia. Por ejemplo, la tabla 1 muestra algunos *features* y el *label* ⁹ asociado a los EE. UU. en 2017:

Populación	327.000.000
GDP	USD 19.290.000.000.000
IDH	0.92
...	
LABEL	Desarrollado

6. Definición del mejor modelo para *Machine Learning* para trabajar los *features/labels* definidos en las etapas anteriores.
7. Entrenamiento de los datos para se llegar al modelo teórico optimo.
8. Examen y estudio de las hipótesis para llegar al mejor conjunto de *features* que llevarían el país a desarrollarse más rápidamente y cuales deberían ser las propuestas de políticas económicas a ser utilizadas para se llegar a este resultado.

Herramientas por utilizar en la investigación

1. Bases de datos principales:
 - a. IMF Database (<https://www.imf.org/en/Data>)
 - b. World bank Database (<https://data.worldbank.org>)
 - c. Our World in Data (<https://ourworldindata.org>)
 - d. OECD Database (<https://data.oecd.org>)

⁹ Como punto de partida, se podrá utilizar los *labels* definidos por las principales organizaciones tales como *OECD (Development Assistance Committee)*, *World Bank (High Income Economies)*, *IMF (advanced Economies)*, etc.

- e. Knoema (<https://knoema.com>)
 - f. Kaggle (<https://www.kaggle.com/datasets>)
- 2. Extracción y limpieza de datos:
 - a. Pandas
- 3. Manipulación de datos:
 - a. NumPy
 - b. SciPy
 - c. StatsModel
- 4. Visualización de datos:
 - a. Matplotlib
 - b. Seaborn
- 5. Machine Learning:
 - a. Scikit-learn
 - b. TensorFlow / Keras (Deep Learning)

Referencias

ⁱ Asfaw Keraga Tsegaamlak. (2014, abril 28). Países desarrollados y subdesarrollados. Causas del subdesarrollo. Recuperado de <https://www.gestiopolis.com/paises-desarrollados-y-subdesarrollados-causas-del-subdesarrollo/>

ⁱⁱ Cook, Thomas R., and Aaron Smalter Hall. "Macroeconomic Indicator Forecasting with Deep Neural Networks." Federal Reserve Bank of Kansas City, Research Working Paper 17-11, September. Available at <https://doi.org/10.18651/RWP2017-11>

ⁱⁱⁱ Rovai, Marcelo. (Julio, 2018). *IMF - World Economic Outlook database*. Recuperado de: <https://github.com/Mjrovai/Python4DS/tree/master/IMF%20-%20World%20Economic%20Outlook>

^{iv} Woloszko, Nicolas (2017, noviembre 8). MAKING BETTER ECONOMIC FORECASTS WITH MACHINE LEARNING. Recuperado de: <http://dataconomy.com/2017/11/making-better-economic-forecasts-machine-learning/>

^v: Rovai, Marcelo (Abril de 2018). *Latin America GDP Analysis* Recuperado de: <https://github.com/Mjrovai/Python4DS/tree/master/Latin%20America%20GDP%20Analysis>