

OBJETIVOS DE DESARROLLO SOSTENIBLE: PROYECTO DE MACHINE LEARNING

JOSÉ LUIS AGUILAR CHARFÉN

RESUMEN. Este reporte muestra tanto un agrupamiento de los países de acuerdo a las métricas utilizadas para conocer el progreso en las 17 metas de sustentabilidad que proponen las Naciones Unidas, como en estos mismos indicadores. A su vez, también se propone una manera de evaluar el progreso de los países en temas de sustentabilidad.

1. INTRODUCCIÓN

Los Objetivos de Desarrollo Sustentable son 17 metas que se acordaron en la Organización de las Naciones Unidas (ONU) en 2015 **united'nations'development'programme'sustainable'node-1** como una llamada a la acción para proteger al planeta y asegurar que para 2030 la población disfrute de paz y seguridad.

Estos objetivos varían desde reducir la pobreza, hasta temas de salud, equidad de género, higiene, crecimiento económico, energía y cuidado al medio ambiente. Cuando se consideran todos los diferentes objetivos, entonces se construye el SDG Index Score, que califica el progreso de un país en las 17 metas. Para este reporte, se tomarán tanto las calificaciones otorgadas por la ONU, como los datos a partir de los cuales se construyen las calificaciones en cada uno de los criterios.

Se mostrará en primera instancia un agrupamiento jerárquico para determinar qué países son más cercanos en ambos rubros discutidos, y posteriormente, se discutirá brevemente una manera de evaluar el progreso de los países en temas de sustentabilidad.

2. METODOLOGÍA

Los datos utilizados son obtenidos de **united'nations'development'programme'sustainable'node-1**. El algoritmo utilizado puede dividirse en las siguientes partes:

1. Escalamiento.
2. Imputación.
3. Reducción de dimensionalidad.
4. Agrupamiento.

Debido a que no todos los países tienen un SDG Index Score, no se consideran para entrenamiento ni validación ni prueba a estos países. Sin embargo, sí se predicen sus valores dadas las variables usadas para construir los primeros dos SDGs.

Para construir el modelo, se realiza un análisis de correlación de variables, y se observan tanto las distribuciones como las correlaciones con gráficos por pares de variables. Posteriormente, se transforman las variables con la transformación Yeo-Johnson **yeo'new'2000**, preferida sobre la de Box y Cox por la presencia de valores 0 en las variables. Posteriormente, se imputan los valores con metodología de vecinos más cercanos **beretta'nearest'2016**. Después, se codifican las regiones usadas para el SDG Index por un método one-hot. Una vez transformadas las variables de estas maneras, se comparan un modelo de mínimos cuadrados ordinarios, regresión Ridge, y Lasso, ajustando sus regularizaciones por validación cruzada, y se evalúa si se encuentra sobreajustado comparando el error cuadrático medio del conjunto de datos de entrenamiento con uno de prueba por separado, corriendo 20 veces el experimento y promediando los valores.

Las decisiones de construcción del modelo y justificaciones se discuten a mayor profundidad en la sección de resultados y conclusión.

3. RESULTADOS Y DISCUSIÓN

4. CONCLUSIONES

5. REFERENCIAS