

Reporte de resultados 4

Incorporación de variables del censo económico

Asistente de investigación: Erick Gabriel Fajardo Martínez

Investigador: Dr. Gabriel Purón Cid

2022-08-12

Avances

Recapitulando, en los reportes anteriores se habían desarrollado y entrenado modelos de aprendizaje máquina utilizando el algoritmo de *Gradient Boosting*, donde la variable de interés es la clasificación de cada municipio (corrupto y no corrupto) y las variables explicativas son los datos sobre las finanzas públicas de cada municipio. Todos los datos sobre las finanzas públicas fueron estandarizados por la población total del municipio. Recordemos que fueron entrenados 4 modelos, los cuales pueden ser clasificados en dos grandes categorías: egresos e ingresos, es decir, modelos entrenados con los respectivos datos de las finanzas públicas; y en subcategorías que corresponden a la construcción de la variable de interés, ya que la clasificación de corrupción fue realizada con la pregunta de percepción y la pregunta de incidencia de la ENCIG 2019. La clasificación de los modelos es resumida de la siguiente manera:

Modelo de egresos

- Utilizando la variable de interés con base en la percepción
- Utilizando la variable de interés con base en la incidencia

Modelo de ingresos

- Utilizando la variable de interés con base en la percepción
- Utilizando la variable de interés con base en la incidencia

Descripción de los datos del Censo Económico

La novedad de estos nuevos modelos es que incorporan los datos del censo económico, lo cual proporciona una mayor robustez en cuanto a las características que definen los outputs de cada modelo. Las variables elegidas de este conjunto de datos fueron las siguientes:

Table 1:

Variables censo económico 2019
Producción bruta total (millones de pesos)
Inversión total (millones de pesos)
Total de gastos (millones de pesos)
Total de ingresos (millones de pesos)
Horas trabajadas por personal dependiente de la razón social (miles de horas)
Personal ocupado total
Horas trabajadas por personal remunerado (miles de horas)
Total de remuneraciones (millones de pesos)
Contribuciones patronales a regímenes de seguridad social (millones de pesos)
Utilidades repartidas a los trabajadores (millones de pesos)
Gastos por consumo de papelería y artículos de oficina (millones de pesos)