

Modelo de agrupamiento utilizando indicadores económicos

Introducción

El PIB (Producto Interno Bruto) es la suma del valor de todos los bienes y servicios de uso final que genera un país o entidad federativa durante un periodo (comúnmente un año o trimestre).

Es muy importante saber si la economía de un país está creciendo o no, es decir, si se produjo más o menos que el año anterior. El cambio del PIB a lo largo del tiempo es uno de los indicadores más importantes del crecimiento económico. **Un crecimiento en el PIB** significa que hay más dinero para construir edificios, casas o comprar maquinaria y que se producirán más bienes y servicios. Esto es beneficioso para todos porque habrá más empleo y más oportunidades para hacer negocios. Por el contrario, **una disminución en el PIB** significa que la producción y actividad económica del país disminuirá; en estas condiciones, es probable que haya desempleo y que esto afecte a muchas familias.

Una de las contribuciones al crecimiento/decrecimiento del PIB de México, se debe en gran parte a su actividad industrial; recordemos que la actividad industrial se define como la transformación de materias primas en productos de consumo final o intermedio. Las principales industrias en México son la automotriz, la petroquímica, la construcción y el cemento, la textil, la industria alimenticia y de bebidas, la minería y el turismo.

Cuando hablamos de actividad industrial, también hablamos de niveles de productividad; usualmente se nos da dicha información en unidades monetarias o en unidades porcentuales. México cuenta con una herramienta estadística llamada: **índice de volumen físico**, que se utiliza para medir los cambios en la cantidad de bienes y servicios producidos en una economía, independientemente de las fluctuaciones en los precios. Este índice es crucial para evaluar mensualmente el crecimiento económico real, ya que permite a los analistas y decisores entender cómo varía la producción sin que los resultados estén distorsionados por la inflación o deflación.

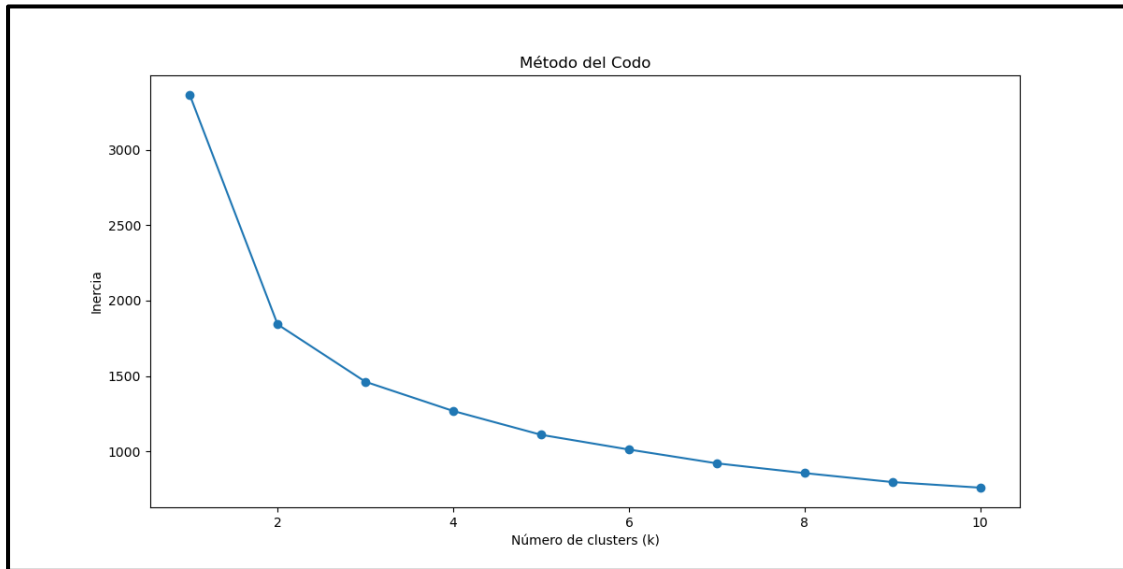
Para propósitos de este ejercicio. Se analizará la relación entre el **Índice del total de la actividad industrial mexicana y el índice extracción de petróleo y gas** para saber cuál es la k más óptima para ser aplicada en el método de clasificación KNN. Los datos de las variables se encuentran en las mismas unidades, por lo que no fue necesario normalizarlos, y fueron recopilados desde enero de 1993 hasta febrero del 2024 (índice base 2018 = 100) y descargados del banco de información económica (BIE) del Instituto Nacional de Estadística y Geografía (INEGI).

Modelo de agrupamiento

$$\text{Índice base 2018} = 100 (\% \text{porcentaje}) \rightarrow \frac{\text{Cantidad producida en el año "n"}}{\text{Cantidad producida en el año 2018}} * 100$$

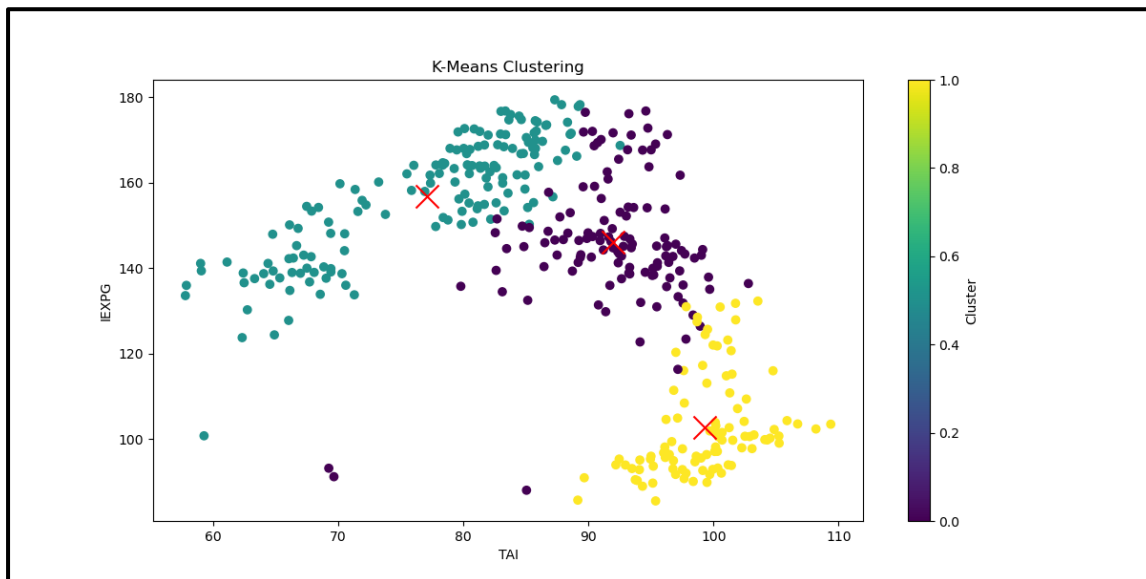
Variable	Representación	Tipo
TAI	Total de la actividad industrial (%porcentaje) <u>Valores tomados:</u> 5.788261 a 245.3426 puntos porcentuales	Continua
IEXPG	Índice de extracción de petróleo y gas (%porcentaje) <u>Valores tomados:</u> 5.788261 a 245.3426 puntos porcentuales	Continua
IMMN	Índice de minería de minerales metálicos y no metálicos excepto petróleo y gas (%porcentaje) <u>Valores tomados:</u> 5.788261 a 245.3426 puntos porcentuales	Continua
IE	Índice de edificación (%porcentaje) <u>Valores tomados:</u> 5.788261 a 245.3426 puntos porcentuales	Continua
ICOI	Índice de construcción de obras de ingeniería civil (%porcentaje) <u>Valores tomados:</u> 5.788261 a 245.3426 puntos porcentuales	Continua
IEA	Índice de elaboración de azúcares, chocolates, dulces y similares (%porcentaje) <u>Valores tomados:</u> 5.788261 a 245.3426 puntos porcentuales	Continua
IFVGO	Índice de conservación de frutas, verduras, guisos y otros alimentos preparados (%porcentaje) <u>Valores tomados:</u> 5.788261 a 245.3426 puntos porcentuales	Continua
IIB	Índice de industria de las bebidas (%porcentaje) <u>Valores tomados:</u> 5.788261 a 245.3426 puntos porcentuales	Continua
IIT	Índice de industria del tabaco (%porcentaje) <u>Valores tomados:</u> 5.788261 a 245.3426 puntos porcentuales	Continua

Modelo del codo



La gráfica nos indica que nuestro conjunto de datos empieza a perder variabilidad a partir del tercer punto del codo, lo que significa que el k más óptimo para trabajar es el 3.

Agrupamiento



A partir del k encontrado, podemos observar como se comportan las interacciones, junto con sus centroides, entre la información del IEXPG y el TAI. Se observa como los puntos están divididos en 3 grupos de diferentes colores debido al valor k.

Modelo de clasificación con cross validation

En evidencias pasadas encontramos que el k óptimo para el modelo de knn de nuestro conjunto de datos es de: 5; estaba representada por esta información:

Accuracy (KNN): 0.7964601769911505

	precision	recall	f1-score	support
0	0.93	0.90	0.92	30
1	0.79	0.90	0.84	29
2	0.74	0.59	0.65	29
3	0.71	0.80	0.75	25
accuracy			0.80	113
macro avg	0.79	0.80	0.79	113
weighted avg	0.80	0.80	0.79	113

Interpretación:

El modelo nos indica que el 79.65% de las predicciones hechas son correctas. Es un buen modelo.

Modelo de clasificación con método del codo

Veamos ahora como se comporta dicho modelo usando el k óptimo encontrado por el método de codo (k = 3).

Accuracy (KNN): 0.8053097345132744

	precision	recall	f1-score	support
0	0.88	0.93	0.90	30
1	0.77	0.83	0.80	29
2	0.84	0.55	0.67	29
3	0.74	0.92	0.82	25
accuracy			0.81	113
macro avg	0.81	0.81	0.80	113
weighted avg	0.81	0.81	0.80	113

Interpretación:

El modelo nos indica que el 80.54% de las predicciones hechas son correctas. Es un buen modelo.

Conclusión: el método del codo nos da un mejor k para un mejor accuracy en el modelo KNN.