



*Decimosexto Encuentro Internacional en la
Enseñanza de la Probabilidad y la
Estadística*

24-28 de agosto de 2026

Framework para la enseñanza de clustering no jerárquico: Un caso de estudio sobre datos epidemiológicos.

Gael Cortes Olivares, Cecilia Cruz López, Emmanuel Morales García ^c, Julia Aurora Montano Rivas^d

^{a,b,c,d} *Universidad Veracruzana, Facultad de Estadística e Informática, Xalapa-Enríquez, Veracruz, México.*

^a zS22023000@estudiantes.uv.mx, ^b ceccruz@uv.mx, ^c emmorales@uv.mx, ^d julmontano@uv.mx

Resumen

La evolución de la estadística hacia la ciencia de datos potencia el uso de algoritmos de *clustering* no jerárquico para el análisis masivo de información. Esta investigación propone un *framework* metodológico estandarizado que garantiza la reproducibilidad técnica, validado mediante un estudio epidemiológico sobre la propagación del dengue en Veracruz. El proceso inicia con la gobernanza y preprocesamiento de un dataset de 15 millones de registros, empleando un flujo de procesamiento por bloques en Python y Google Colab para optimizar la carga computacional. La fase de ejecución implementa el algoritmo K-means, donde se determina el número óptimo de clústeres mediante el método del codo y métricas de validación como Silueta e índice de Calinski-Harabasz. Finalmente, el marco guía la interpretación de centroides para caracterizar patrones de riesgo no evidentes y fortalecer la toma de decisiones basada en evidencia. Esta propuesta desarrolla competencias avanzadas en el manejo de estructuras de datos masivas, estableciéndose como un modelo pedagógico escalable y reproducible para diversas áreas del conocimiento científico.

Palabra claves: Aprendizaje no supervisado, Ciencia de datos, Estadística.