



# Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey

**Desarrollo de aplicaciones avanzadas  
de ciencias computacionales (Gpo 501)**

## **Actividad 2.1.**

### **Clasificador Knn**

**Alejandro Daniel Moctezuma Cruz A01736353**

**Profesores:**  
**Candy Yuridiana Alemán Muñoz**

Jueves 3 de abril de 2025

## Conclusiones finales

Una vez realizado el código, pude contemplar que el resultado varía un tanto con diferentes valores de  $k$ . Por ejemplo, parece existir una tendencia en la cual a medida que se utilizan más instancias diferentes, el porcentaje de aciertos parece disminuir. Ejemplo de esto fue el utilizar el archivo llamado “Diabetes-Clasificación”, en donde un valor de  $k$  de 2 da un 81%, uno de 3 un 83%, uno de 7 77%, etc. Si bien este porcentaje no siempre disminuye con respecto al número anterior utilizado para  $k$ , la tendencia parece ser que este porcentaje disminuye rápidamente.

Este comportamiento se repite con la base de datos llamada “Diabetes-Entrenamiento”, que al tener más datos, además de tardar más tiempo, parece ayudar a que el porcentaje disminuya con una tasa mucho más lenta. Es decir, que utilizando esta base de datos más grande, parece que se logra obtener mejores resultados.

Con respecto a la normalización de los datos, por otro lado, consideraría que este parece ayudar mayormente al rendimiento con cantidades bastante grandes de bases de datos, además de que el uso del promedio ayuda a que los datos se encuentren un poco más estandarizados y utilizando como referencia un valor similar.

En el caso de los resultados mínimos y máximos, como se ha mencionado anteriormente, parece existir una tendencia hacia porcentajes cada vez menos acertados a medida que se incrementa el valor de  $k$ . No obstante, este valor parece encontrar una tendencia exponencial a cada vez ser más difícil que se disminuya el porcentaje. Por ejemplo, al utilizar “Diabetes-Clasificación” con un valor de  $k$  de 34, se puede obtener un 70%, mientras que con un valor de  $k$  de 50 se obtiene un porcentaje de 64%. De igual forma, el resultado máximo de tasa de acierto es siempre un 100%, debido a que al utilizar un valor de  $k$  de 1 simplemente se le está diciendo al clasificador que utilice un solo valor.