

## Laboratorio 4

Utilizando la base de datos de flores y el método de KNN categorizar cada una de ellas para encontrar el valor óptimo de K siguiendo los siguientes parámetros o requerimientos de diseño:

- NO SE PERMITE UTILIZAR LIBRERIAS DE SCKLEARN, TODO EL PROCEDIMIENTO TIENE QUE SER ESCRITO DESDE CERO
- La cantidad de datos utilizados por cada categoría para evaluarlos debe ser el 20 % de esta. La cual debe ser seleccionada de forma aleatoria. El restante 80 % tiene que ser utilizado para estimar el mejor valor de K
- Deben aplicar la distancia euclidiana con mínimo 2 variables. NO SE ACEPTAN SOLAMENTE 1.
- Deben almacenar los vectores de diferencia por cada K probado.
- Deben reportar por cada K probado la cantidad de aciertos y fallas realizadas por el algoritmo. Seleccione el K que mejor desempeño obtuvo mirando solamente la cantidad de aciertos y fallos.
- Deben probar con mínimo 5 valores de K. iniciando desde K=3. NO SE ACEPTA UN K INFERIOR A 3
- Con el mejor K seleccionado, ahora deben seleccionar el 50% de cada categoría y probar el desempeño en la clasificación.
- Con el mejor K seleccionado, ahora deben seleccionar el 75% de cada categoría y probar el desempeño en la clasificación.
- Analicen los resultados obtenidos y den una razón el por qué se podría obtener desempeños diferentes al probarlos con diferentes cantidades de datos.
- Reportar todo lo realizado, conclusiones, justificaciones y procedimientos en un documento en LATEX.

NOTA: EL MEJOR TRABAJO OBTENDRA UN 0.3 EN UN QUIZ DENTRO DEL SEGUIMIENTO.