

## LABORATORIO 06

### K-NN Y ÁRBOLES DE DECISIÓN

Docentes: Edward Hinojosa C. - Yesseina D. Yari R.

04 de Junio del 2021

#### 1 CONCEPTOS BÁSICOS

- k-NN
- Entropía
- Árboles de Decisión

#### 2 EQUIPOS Y MATERIALES

- Un computador.
- Material del curso.
- Bibliografía del curso [1].

#### 3 EJERCICIOS

1. Implemente (en cualquier lenguaje de programación) el Algoritmo k-NN para clasificar los elementos en el archivo fruit.csv (8 puntos):
  - Defina (aleatoriamente y proporcionalmente por cada clase) 80% de elementos como el conjunto de entrenamiento y 20% de elementos como conjunto de test.
  - Cree un archivo .txt que muestre el nombre del alumno, el algoritmo utilizado, el conjunto de entrenamiento y el conjunto de test.
  - Para cada elemento de test muestre sus datos, su clase real, sus k vecinos más cercanos, la distancia a los k vecinos más cercanos y su clase estimada. Dicha información se debe mostrar en el archivo .txt
  - Defina la tasa de acierto del algoritmo k-NN. Dicha información se debe mostrar en el archivo .txt
2. Implemente (en cualquier lenguaje de programación) el Algoritmo ID3 para generar el árbol de decisión para los elementos del conjunto de entrenamiento en el archivo virus\_train.csv (12 puntos):
  - Considere el archivo virus\_test como elementos del conjunto de test.
  - Cree un archivo .txt que muestre el nombre del alumno, el algoritmo utilizado, el conjunto de entrenamiento y el conjunto de test.
  - Muestre cada ganancia de información obtenida en cada paso, indicando su posición en el árbol de decisión (Por ejemplo: ganancia\_fatigue(mild), ganancia\_fatigue(mild)\_cough(moderate),...). Dicha información se debe mostrar en el archivo .txt
  - Para cada elemento de test muestre sus datos, su clase real y su clase estimada usando el árbol de decisión generado. Dicha información se debe mostrar en el archivo .txt
  - Defina la tasa de acierto del árbol de decisión. Dicha información se debe mostrar en el archivo .txt
  - Realice una imagen en formato .jpg, .png o .pdf (a mano) del árbol de decisión generado.

#### 4 ENTREGABLES

Al finalizar el estudiante deberá:

1. Generar los archivos .txt y/o imágenes solicitados.
2. Comprimir en un archivo .zip todos los archivos anteriores (además de todo código fuente, sin el cual no se revisará el laboratorio) y subirlo al aula virtual hasta 5 minutos después de finalizar la hora del laboratorio con el nombre:

Laboratorio\_XX\_ApellidoPaterno\_ApellidoMaterno\_PrimerNombre\_IA\_2021A\_EPCC\_UNSA.zip

#### 5 RÚBRICA DE EVALUACIÓN

Criterios	Muy Bueno	Bueno	Regular	Malo
Resolución del Laboratorio	Resuelve los ejercicios sin errores mostrando cada uno de los puntos solicitados. <b>Puntaje: 16-20 puntos</b>	Resuelve los ejercicios con pocos errores mostrando casi o todos todos los puntos solicitados. <b>Puntaje: 11-15 puntos</b>	Resuelve los ejercicio con varios errores y mostrando todos o pocos de los puntos solicitados. <b>Puntaje: 1-10 puntos</b>	No resuelve los ejercicios o no entrega el laboratorio. <b>Puntaje: 0 puntos</b>

- **IMPORTANTE** En caso de copia o plagio o similares todos los alumnos implicados tendrán sanción en toda la evaluación del curso.

#### BIBLIOGRAFÍA

- [1] SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Database System Concepts**. 7th. Edition: McGraw-Hill Book Company, 2020. ISBN 9780078022159.