

**Minería de Datos**  
1er parcial – Primavera 2023

Nombre: \_\_\_\_\_ Matricula: \_\_\_\_\_  
Calificación: \_\_\_\_\_

**Instrucciones:** Analiza cada una de las preguntas que se exponen a continuación. Realiza las implementaciones correspondientes, las cuales se deberán de acuerdo a las instrucciones del profesor. En caso de duda, contactar al profesor.

El primer parcial consiste en entrenar un modelo de aprendizaje basado en árboles de clasificación para la detección de incendios forestales en imágenes digitales. Para ello considerar los siguientes puntos:

1. (1.0 pts) Realiza una caracterización de la población, en la cual se señale con precisión aquellos aspectos que se han considerado de la población para la construcción de una muestra.
2. (1.0 pts) Construir una muestra (conjunto de imágenes) de incendios forestales, a partir de una estrategia de muestreo propuesta para el problema. Definir con claridad la estrategia seleccionada, estratos considerados, número de instancias a recolectar (tomando en cuenta que serán 3 clases para trabajar: incendio, no incendio, humo), entre otros factores relevantes que consideren para la construcción de esta.
3. (1.0 pts) Realizar un proceso de extracción de zonas de interés “ROI” (asimétrico), usando un etiquetador de imágenes digitales. Se deberán identificar y etiquetar tres regiones de interés: Incendio (I), Humo (H), no incendio (N). Las etiquetas de clase deberán de recuperarse al momento de generar el vector de características de cada imagen local. El número de regiones ROI por clase deberá ser aproximadamente igual al número de instancias definido en la caracterización.
4. (1.0 pts) Definir un conjunto de valores característicos a extraer de las imágenes locales a partir de una imagen de entrada. Enlistar cada uno de ellos, definiendo con claridad cómo se calculan y argumentar brevemente porqué cada valor puede aportar a un proceso de entrenamiento. Se deberán considerar descriptores de 1er orden, de 2º orden. Calcular por lo menos 10 tipos de descriptores diferentes.
5. (1.0 pts) Para el proceso de experimentación, deberán de definir con claridad el tipo de particionamiento inicial de training – test generado. Además, indicar como se trabajará la validación cruzada sobre el conjunto training inicial. Se deberá argumentar el valor “k” de validación cruzada, garantizando la representatividad de las particiones. Generar las particiones pertinentes para el proceso de experimentación y validar que las particiones no tengan sesgo.
6. Implementar un proceso de entrenamiento basado en árboles, en el cual se logre identificar un modelo que se ajuste lo mejor posible a la estructura de los datos de entrenamiento.
  - a. (2.5 pts) Realizar una serie de experimentos que permita ir encontrando los resultados para cada pliegue de la validación cruzada, sin que los resultados se encuentren sobre - ajustados. Se deberá especificar y mostrar con gráficas como se detecta el sobre – ajuste en los experimentos, indicando con claridad cual es la desviación estándar y media que se obtiene de la validación cruzada.
  - b. (1.5 pts) Realizar el experimento final sobre el conjunto training y test inicial, para obtener el resultado global de su fase de experimentación.
7. (1.0 pts) Hacer un análisis de los resultados obtenidos, donde se explique e interprete los resultados que se han obtenido en las diferentes fases.

FECHA DE ENTREGA: LA SEÑALADA EN CLASE