**PRIMER PARCIAL — INF 354**

Universitaria: Tarqui Machaca Selena Leydi

Ci: 8348032 LP RU: 1749954

LINK GITHUB: <https://github.com/Seltam475/PRIMER-PARCIAL-354.git>

PREGUNTAS:

1. Seleccione un dataset de su interés (de cualquier repositorio, que contenga datos tabulares mínimo de 1500 filas y al menos 10 columnas). Realice el cálculo de la media, moda, cuartiles de datos y de percentiles por columna; explique qué significa en cada caso graficando los resultados (sin numpy y pandas).

2. Realice lo mismo del inciso anterior con el uso de numpy y pandas

3. Del dataset elegido realice la imputación por columnas sin scikit-learn.

4. Aplique a su dataset el etiquetado simple y binario.

5. Del dataset elegido, migre el mismo a WEKA y utilice cuatro técnicas de preprocesamiento (realice la captura de pantallas de estos por fases). Explique la razón de aplicar estas técnicas.

6. El dataset elegido en PYTHON, realice tres tecnicas de preprocesamiento. Explique la razón de aplicar estas técnicas.

1. **Manejo de valores faltantes**:
   * Técnica: Rellenar los valores faltantes en el conjunto de datos.
   * Razón: Los datos reales suelen tener valores faltantes. Rellenarlos puede ayudar a evitar problemas en el modelado y el análisis posterior. La imputación puede realizarse utilizando la media, la mediana u otros métodos adecuados.
2. **Escalado de características**:
   * Técnica: Escalar las características para que tengan la misma escala.
   * Razón: Algunos algoritmos de aprendizaje automático, como la regresión lineal y las máquinas de soporte vectorial, son sensibles a la escala de las características. Escalarlas asegura que todas tengan un efecto similar en el modelo.
3. **División de datos en conjuntos de entrenamiento y prueba**:
   * Técnica: Dividir el conjunto de datos en un conjunto de entrenamiento y un conjunto de prueba.
   * Razón: Esta técnica es fundamental para evaluar el rendimiento del modelo. El conjunto de entrenamiento se utiliza para ajustar el modelo, mientras que el conjunto de prueba se utiliza para evaluar su rendimiento. Evita el overfitting y permite una evaluación imparcial del rendimiento del modelo.

7. Con el uso de librerías en PYTHON, construya la dependencia de Abuelos, tíos, padres, primos e hijos de su familia.

8. Del dataset de citas obtenida en la clase, realice el cálculo en excel para obtener el arbol de decisió, grafique el el arbol.

9. Con PYTHON sin librerías construya los índices de al menos dos ciclos para un Split de Train 80 y test de 20, apliquelo en el dataset iris.