**Universidad de San Carlos de Guatemala**

**Facultad de Ingeniería**

**Escuela de Estudio de Postgrado**

**Maestría en Ingeniería para la Industria con Especialización en Ciencias de la Computación**

**Introducción a la minería de datos**

**Nombre:** Luis Fernando Culajay Sandoval

La investigación fue realizada por Madhar Taamneh, Sharaf Alkheder y Salah Taamneh quienes analizan accidentes en Abu Dhabi (específicamente 5973 entre 2008 y 2013, el objetivo del análisis fue predecir que tan fuertes podrían llegar a ser las lesiones y determinar los factores que más influyen en lo grave que pueda ser un accidente.

Para realizar el análisis utilizaron minería de datos, a través del software WEKA, aplicando 4 algoritmos de clasificación principales:

* Decision Tree (J48)
* Rule Induction (PART)
* Naive Bayes
* Multilayer Perceptron

Por otra parte, para medir la efectividad de cada modelo fue evaluado de 3 maneras:

1. Se utilizó el dataset completo para entrenar el algoritmo
2. Se realizó validación cruzada de 10 pliegues
3. Se volvió a muestrear el conjunto de datos para acoplar los valores hacia una distribución uniforme.

Finalmente, la investigación llegó a las siguientes conclusiones:

* Personas entre 18 a 30 años son más vulnerables a tener accidentes
* Los hombres tienen más accidentes que las mujeres
* Las colusiones surgen principalmente en ángulos rectos
* Existe una tendencia de disminución de accidentes con el paso de los años
* Las nacionalidades con más accidentes son UAE, asiáticos y árabes

