

## Trabajo Práctico Final – Data Mining y Machine Learning

El TP consta de 2 Partes y ambas Partes deberán entregarse cada una en un único archivo pdf. No se aceptarán otros formatos de archivo que no sean pdf.

En cada uno de los 2 archivos pdf deberá figurar la siguiente información:

- Nombre del alumno y Asignaciones (en carátula)
- Informe con los resultados de los análisis
- Anexo con el código en R utilizado (anexo como una sección dentro del mismo pdf)

En el nombre del archivo pdf colocar “NombreDelAlumno DM TP Final Parte1.pdf” y “NombreDelAlumno DM TP Final Parte2.pdf”

**Aclaraciones:** En el caso de agregar imágenes de pantalla, las mismas tienen que estar recortadas a fin de mostrar solamente lo pedido en cada ejercicio. Por ejemplo, para indicar los centroides se puede insertar una imagen de pantalla, pero no de toda la pantalla, sino solamente mostrando los centroides.

Aunque se indique el código R utilizado en las respuestas a los ejercicios, igualmente se espera un anexo con todo el código utilizado.

## Trabajo Práctico Final - Parte 1

### Parte A

- 1) Abra la base *shuttle* de la librería *MASS*. Se quiere predecir el uso o no del autolander (variable a predecir: “*use*”). Renombre la base como “*base*”.
- 2) Indique de qué trata el problema.
- 3) Muestre un *head* de la base.
- 4) Comente las variables, cantidad de registros, **cantidad de registros de cada clase**.
- 5) Realice un gráfico de barras de la variable “*use*”.

### Parte B

- 6) Cargue la librería *caret*, setee la semilla=8 y particione la base en un conjunto de entrenamiento y uno de testeo, utilizando la instrucción `createDataPartition` de la librería *caret*. Setee **p=(asignado)**. Indique el código R utilizado.
- 7) Muestre un *head*, *str* y un *summary* del conjunto de entrenamiento y del conjunto de testeo. ¿Cuántos registros quedaron en cada conjunto? **¿Cuántos registros de cada clase quedaron en cada conjunto?**

**Parte C**

- 8) Cree un Árbol de Decisión (con librería *rpart*) para modelar el problema planteado. Indique el código R utilizado.
- 9) Grafique el árbol de decisión resultante con la instrucción *rpart.plot* de la librería *rpart.plot*.
- 10) Calcule la matriz de confusión utilizando la instrucción *confusionMatrix* de la librería *caret*. Muestre una imagen de los resultados obtenidos.
- 11) ¿Cuántos registros quedaron bien y mal clasificados?
- 12) ¿Cómo fue la performance del modelo? (accuracy, sensibilidad y especificidad)
- 13) Según el árbol de decisión creado, ¿cuándo debería usarse el autolander?

**Parte D**

- 14) Setee la semilla=8 y cree una Red Neuronal (con librería *nnet*) para modelar el problema planteado, con *size*=6 y *maxit*=5000. Indique el código R utilizado.
- 15) Grafique la Red Neuronal resultante.
- 16) Calcule la matriz de confusión utilizando la instrucción *confusionMatrix* de la librería *caret*. Muestre una imagen de los resultados obtenidos.
- 17) ¿Cuántos registros quedaron bien y mal clasificados? ¿Cómo fue la performance del modelo? (accuracy, sensibilidad y especificidad)

**Trabajo Práctico Final - Parte 2****Ejercicio 1**

- 1) Abra la base *Vehicle* de la librería *mlbench* y renómbrela como "base".
- 2) Indique de qué trata el problema, comente las variables, cantidad de registros.
- 3) Muestre un *head* de la base.
- 4) Borre la variable *Class* y muestre nuevamente un *head* de la base.
- 5) Setee la semilla=8 y realice un Agrupamiento K-means con **cantidad de grupos = (asignado2)**. Indique el código R utilizado.
- 6) Muestre una imagen de los centroides.
- 7) ¿Cuántos elementos quedaron en cada grupo?
- 8) ¿A qué grupo pertenece el cuarto elemento de la base?
- 9) Realice un gráfico con dos variables coloreado por los grupos formados.
- 10) ¿Puede determinar alguna característica de alguno de los grupos?