

## Trabajo de laboratorio para Sistemas de Soporte de Decisión – Octubre 2020

### Parte 1

Utilizando los conceptos de big data y bases de datos NoSQL, presentar una posible solución para el siguiente problema. Se espera como solución, mencionar posibles bases de datos y la arquitectura a alto nivel para la solución de este problema. Nota: para la arquitectura, basarse en lambda como comentamos en la clase.

**Objetivo:** emitir alertas a conductores de una compañía (transporte de pasajeros) ante posibles congestiones en el tránsito.

La compañía TuTransporte cuenta con varios móviles operativos durante todo el día, en la búsqueda de agregar valor al servicio, TuTransporte desea alertar a los conductores activos ante posibles congestiones en el tránsito para que así otros conductores puedan desviarse de dichas rutas impactando en un mejor servicio al cliente.

El proceso que será necesario implementar debe considerar los siguientes puntos:

- Recolección de datos (datos emitidos por los GPS de los vehículos (lat, long, tiempo), recolectar estos datos y almacenarlos para poder ser útiles para el análisis (recolección de ubicación cada 30 segundos para cada móvil operativo).
- El negocio considera que si no hubo variación en los últimos 5 reportes de la posición de un móvil, se debe alertar por posible congestión dicha ubicación (lat, long).
- Además de las alertas para los móviles operativos en un momento dado, es necesario también que cualquier conductor que iniciará el servicio, pueda consultar si hubo congestiones en el día.
- Por otra parte, la compañía requiere almacenar los puntos donde ocurrieron congestiones para mantener un histórico de congestiones para futuros análisis y toma de decisiones.

### Parte 2

Una entidad bancaria, El Banco, cuenta con una cartera de clientes premium que corresponden a los clientes vip de dicha entidad.

Es de interés para esta institución, en base a datos históricos del comportamiento de dichos clientes, poder determinar qué clientes dejarán El Banco o alguno de los beneficios que utilizan. Se debe determinar que clientes dejarán la institución en los dos meses posteriores al análisis.

Para esto, la institución requiere optimizar (maximizar) la función de ganancia que contiene los componentes que implicarán la campaña de marketing a realizar de acuerdo a la estimación de baja de los clientes premium. Dicha función luego del análisis de los Departamentos involucrados en el proyecto, se corresponde con:

$ganancia = \$ 9.750 * 'BAJA+2' - \$ 250 * ( 'BAJA+1' + 'CONTINUA' )$

Previo al análisis que requiere la institución, es de interés responder algunas inquietudes relacionadas al análisis de variables presentes en los datos obtenidos

### **Análisis de algunas variables**

Se pide:

- 1) La variable del monto de consumo total de la tarjeta1, ¿sigue una distribución normal?.  
¿Obtiene alguna otra información con este análisis?
- 2) Realice el histograma de la edad de los clientes. ¿Qué información le brinda dicho histograma?
- 3) Para el caso del histograma del monto del consumo en pesos de la tarjeta1 de los clientes, ¿le brinda alguna información?
- 4) Compare la clase del dataset con respecto a la edad de los clientes de la institución, ¿qué información encuentra en esto, o que le permite asegurar o no asegurar el resultado obtenido sobre estas variables?

### **Parte 3**

Objetivo: búsqueda de los parámetros óptimos de Rpart. Agregar variables al dataset de origen.

Se pide:

- 1) Mencione brevemente que utilidad tienen los árboles de decisión en este tipo de análisis.
- 2) En el modelo3.r visto en clase. Mencione brevemente que estrategia siguió para la optimización de los parámetros de rpart.
- 3) Agregue una imagen de la salida resultante de cada corrida de Rpart(al menos 3 corridas del modelo 3) en la búsqueda de parámetros óptimos. Alcanza con que se visualice en las imágenes los primeros 4 o 5 registros de cada salida obtenida (por ejemplo, un Excel de la salida obtenida ordenado de mayor a menor por la columna de ganancia).

Además de las 3 imágenes, mencione cuáles son los parámetros óptimos que encontró luego de las distintas pruebas.

- 4) Luego de identificados los parámetros óptimos de Rpart, agregar algunas variables al dataset y volver a ejecutar el modelo3.r usando el nuevo dataset y utilizando los valores óptimos de los parámetros. ¿Qué resultado obtuvo para la ganancia?

Ejecutar el modelo2.r y verificar en las 5 salidas resultantes si las variables que agregó al dataset

aparecen dentro de las variables más significativas.

Justificar brevemente las decisiones que tomó para agregar las nuevas variables.