

Asignación 2: Aprendizaje Supervisado

F.C/W.G GD M.D

Becas Crema 2.0

Objetivo:

La Organización de Becas Crema ya ha obtenido una vista minable que permite su análisis para la detección de patrones subyacentes, usando dicho dataset, se requiere que compare el rendimiento de algoritmos de clasificación vistos en el curso que permitan etiquetar el modo de ingreso de la persona usando un subconjunto de las variables restantes seleccionado por usted.

Requerimientos:

- Escoger las variables predictoras que usará en cada modelo.
- Use modelos basados en **k vecinos mas cercanos**, **árboles de decisión** y **reglas de clasificación** cuya clase objetivo será el método de ingreso.
- Evalúe la calidad de cada modelo usando matrices de confusión.
- **(Adicional)** Use el análisis ROC/AUC para comparar la calidad de sus modelos y especificar cuál recomienda y sus argumentos, se insta a usar visualizaciones.

Consideraciones de contenido:

- Puede agregar visualizaciones que soporten sus análisis y/o decisiones.
- Se considerarán puntos adicionales por evaluar y comparar los modelos usados.
- Puede recomendar decisiones enmarcadas en el contexto del problema.

Escogiendo un hogar

Objetivo

El amigo de un amigo tiene pensado estudiar en Roma durante 2 años a partir de algún momento. En vista de esto, ha decidido usar la plataforma [sturent](#) para encontrar un hogar durante sus estudios al ser la única compañía certificada para estudiantes en Roma. Nuestro amigo, al tener poco capital disponible, necesita escoger inteligentemente el hogar en base a una lista de disponibles provista por **sturent**. Para ello, necesita de su ayuda tratando de predecir el precio de los apartamentos en base a los apartamentos disponibles y escoger, de esta manera, el mejor entre ellos.

Requerimientos

- Culminar la función `parse_data()` de `google_api.R`.
- Encuentre un modelo de regresión apropiado para hombres y otro para mujeres para tratar de predecir el precio mensual de un hogar en base a los features que usted escoja.
- Tiene libertad plena para añadir o remover features o **instancias** al dataset.

- Es obligatorio transformar la columna dirección por, al menos, otra columna a juicio de usted usando el API de *Google Maps Directions* y específicamente la sección *Distance Matrix*. Información adicional en **Usando el API de Google Maps**.
- Escoja, usando el modelo del primer punto, el mejor apartamento disponible.
- Debe utilizar como entrada el archivo **hogares.xlsx** obligatoriamente.

Consideraciones de contenido

- Puede usar regresión lineal simple o múltiple.
- Se considerarán puntos adicionales por evaluar y comparar los modelos usados.
- Puede recomendar decisiones enmarcadas en el contexto del problema.
- Recomendamos usar el método **lm** de la librería **stats** en R.

Información sobre hogares.txt

De Tipo de Inmueble

- *appartamento*: apartamento.
- *monolocale*: estudio.

De Descripción:

- *ingresso*: entrada.
- *camera*: habitación.
- *cucina*: cocina.
- *bagno*: baño.
- *riscaldamento*: calefacción.
- *condominio*: debe pagar condominio.

De Habitaciones disponibles:

- *singola*: simple.
- *doppia*: doble.
- *posto letto PL*: cama en habitación doble.
- *intero appartamento*: apartamento entero.

Precio Mensual:

- *spese escluse*: no incluye cargos por servicios.

Notas:

- *disponibile per ragazzi/ragazze*: Disponible para hombres/mujeres.

Usando el *API de Google Maps Directions*

Para poder usar el API deben generar un API Key en la página principal de Google Maps. Recomendamos el siguiente [link](#).

Pasos para generar la llave (*Key*)

1. Hacer click en el botón azul *GET A KEY*.
2. Registrar su aplicación para usar el API. En la mayoría de los casos, deberán generar un proyecto nuevo.
3. Habilitar su *key* para el uso específico de *Google Maps Distance Matrix API*.
4. Hacer pruebas usando un URL como especifica la documentación de [Google Maps Distance Matrix](#)

Usando archivos de referencia

Se proveen dos archivos base llamados **google_api.R** y **usage.R** para facilitar el uso del API. En él están las funciones básicas de lo necesitado para transformar la dirección en al menos una columna en base a su criterio.

- *google_api.R*: Provee una interfaz para usar el API dónde dado un conjunto de strings de origen, un conjunto de destinos y el API key devuelve un json con la salida.
- *usage* es un ejemplo de cómo usar **google_api** intuitivo y es recomendable leer la documentación de **jsonlite** de R para poder culminar la función **parse_data()** en **google_api** cuya objetivo es obtener los datos que usted desee del json que retorna **get_data()**.

Consideraciones generales:

Del Repositorio

Ingresa a la dirección [DataMiningUCV](#) y siga las instrucciones para crear un repositorio en [GitHub](#) perteneciente a la organización **DataMiningUCV**.

Este repositorio será propiedad de la organización pero solo usted puede realizar cambios en el mismo. El repositorio debe poseer lo siguiente:

1. Script (.Rmd) intradocumentado y reproducible llamado **becas.Rmd** referente a **Becas Crema 2.0**.
 2. Script (.Rmd) intradocumentado y reproducible llamado **hogar.Rmd** referente a **Escogiendo un Hogar**.
 3. README.md explicando la configuración del ambiente en el cual trabajó:
- Ejemplo: [README.md de Bootstrap](#)

De la entrega:

- La asignación es **estrictamente** individual.
- Fecha de entrega: 11 de marzo de 2016.
- Recuerde leer las condiciones de entrega de asignaciones de la materia.