

Base Trabajo práctico

Explicación

Les proponemos que a partir del dataset de campañas de marketing utilizado en CD1 busquemos cómo podemos agruparlos y comenzar a explorar cómo podríamos determinar ciertas acciones comerciales en función a cada grupo de cliente que identifiquemos y cómo lograr que sea sostenible en el tiempo.

Fuente del dataset:

<https://www.kaggle.com/datasets/rodsaldanha/arketing-campaign?resource=download>

1. Explorando datos. Buscar alguna relación entre las variables.
Al analizar 2 o más variables entre sí. ¿Cuántos grupos identificas y cómo los perfilarías?
2. Utilizando K-Means. ¿En cuántos clusters dividirías a los clientes? ¿Por qué?
3. Utilizando Clustering Jerárquico. ¿En cuántos clusters dividirías a los clientes? ¿Por qué?
4. Explorar los grupos resultantes a partir de diferentes técnicas de análisis y visualización de datos.
 - a. ¿Cómo caracterizarías/perfilarías a cada grupo desde un punto de vista del negocio a partir de K-Means?
 - b. ¿Que diferencias hay con Jerárquico en cuanto a la caracterización de los grupos?
 - c. ¿Qué acciones comerciales personalizadas para cada grupo realizarías?
 - d. Utilizando la variable target, ¿Cuán bien funcionó la clusterización realizada?
5. Pensa como transformar las variables “Education”, “Marital” y “AcceptedCmp” para mejorar la identificación de grupos. Convertílas y volvé a correr k-means y jerárquico. ¿Encontraron grupos distintos? ¿Hubo diferencias con la agrupación anterior previa a la transformación de las variables? Comparar nuevamente utilizando la variable target.
6. La información que tuvimos disponible fue un trabajo de recopilación puntual y muy manual. ¿Cuál sería la mejor forma de capturar los datos de los clientes para seguir alimentando este registro de manera sostenible en el tiempo?
7. Si no contases con la variable target: ¿Cómo podrías confirmar, transcurrido algunos meses si la agrupación propuesta por el algoritmo es razonable?

Entregables

Se deberá entregar el archivo de Orange con el flujo creado para la ejecución de los modelos y un archivo pdf con las respuestas a las diferentes preguntas que plantea el TP. La entrega será grupal, es decir, sólo se entregará un conjunto de archivos por grupo.

Fecha de entrega

Punto 1 y 2: **Sábado 9/11/2023** a las 13hrs

Puntos 3, 4, 5, 6 y 7. A más tardar, el trabajo debe ser entregado el día **22 de Septiembre** hasta las 23:59 hs.