

Feedback

1. Aprendizaje personal:

¿Qué conceptos nuevos aprendiendieron al trabajar con clustering y la distancia de Gower?

Aprendimos que la distancia de Gower es muy útil cuando se trabaja con datos mixtos, porque permite combinar variables numéricas y categóricas. También entendimos que la normalización depende del tipo de variable: usar MinMax cuando la distribución está sesgada y StandardScaler cuando es más simétrica. En cuanto al clustering, comprendimos a evaluar los resultados con métricas como Silhouette, que mide qué tan compactos son los grupos, y el Davies-Bouldin, que muestra qué tan separados están entre sí. Por otro lado, descubrimos que el uso de la ponderación muestral (fnlwgt) para ajustar los resultados reflejan mejor la población real. En general, aprendimos que siempre hay un equilibrio entre interpretabilidad y métricas numéricas, y que a veces optimizar demasiado las métricas puede hacer que los resultados sean menos claros o útiles para el análisis.

¿Qué parte del proyecto les resultó mas desafiante?

Lo más desafiante fue decidir entre usar linkage='average' o linkage='complete' en el clustering jerárquico. Con el método average, las métricas mostraban mejores resultados y sugerían un k=2, pero los clusters quedaban muy desbalanceados (99,98% frente a 0,02%), lo que dificultaba interpretar las desigualdades entre grupos. En cambio, con complete los clusters eran más equilibrados y fáciles de interpretar, aunque las métricas numéricas eran algo peores. Por eso, fue necesario documentar y justificar la elección, priorizando la interpretabilidad sobre la optimización estricta de las métricas.

¿Qué estrategias usaron para superar esas dificultades?

Las estrategias utilizadas fue crear varias versiones e ir cambiando el contexto, además, hicimos distintas pruebas hasta llegar al resultado que nos parecia mas politicamente correcto.

2. Utilidad percibida:

¿Consideran que este proyecto les aportó herramientas prácticas aplicables a otros problemas?

Si consideramos que el proyecto nos aporoto nuevas herramientras practicas, ya que nos permite decidirnos por un tema de tesina y comprender mejor ciertos conceptos.

¿En qué situaciones reales creen que podrían aplicar lo trabajado en este proyecto?

Algunas situaciones reales donde se podría aplicar lo trabajado en este proyecto son, por ejemplo:

- Análisis economico de mercado
- En el area de recursos humanos de una empresa
- En análisis medicos

Nos serviría en cualquier ámbito en el cual tengamos que analizar datos para entrenar un modelo que reconozca determinados patrones.

¿Qué parte del trabajo consideran que fue más valiosa para su formación?

La parte más valiosa del trabajo para nuestra formación fue el análisis comparativo del ejercicio 6, porque nos permitió entender el trade-off metodológico entre interpretabilidad y optimización de métricas, además de su relevancia social. También destacamos la implementación de la distancia de Gower, que nos ayudó a manejar datos mixtos de forma adecuada. El análisis de desigualdades por género y raza nos pareció especialmente importante, ya que aportó una comprensión más profunda del impacto social de los datos.

3. Feedback para la cátedra:

¿Qué aspectos de la consigna estuvieron claros y cuáles podrían mejorarse?

El aspecto que no estuvo del todo claro en la consigna fue el objetivo del trabajo. Según lo que comprendimos, el propósito principal planteado era una tarea de predicción, específicamente determinar si el ingreso anual de una persona era superior o inferior a \$50K. Sin embargo, luego entendimos que el objetivo real era definir nuestro propio contexto de análisis y aplicar las técnicas de clustering en función de eso. Este punto podría haberse explicado con mayor claridad en la consigna, ya que al inicio generó confusión sobre el enfoque que debíamos darle al proyecto.

¿El tiempo asignado fue suficiente?

El tiempo asignado fue un poco justo.

¿Qué sugerencias tienen para hacer este proyecto más interesante o útil el próximo año?

Estaría bueno poder hacer el modelado completo incluyendo la validación y los test. Además, de realizar un breve trabajo práctico casi al nivel del proyecto con el docente para saber conocer que decisiones se podrían haber tomado mejor o peor.