

Problema predictivo a resolver

El problema predictivo a resolver es predecir si una pareja que se conoce en una cita rápida de cuatro minutos volverá a verse.

Dataset

El dataset utilizado es el conjunto de datos de citas rápidas de la Columbia Business School, que contiene información sobre los participantes en eventos de citas rápidas entre 2002 y 2004. Los datos incluyen información demográfica, atributos de personalidad y calificaciones de atractivo, sinceridad, inteligencia, diversión, ambición e intereses compartidos.

Métricas de desempeño

Las métricas de desempeño requeridas son:

Precisión: La precisión es la proporción de veces que el modelo predice correctamente si una pareja volverá a verse.

Recall: El recall es la proporción de veces que el modelo predice que una pareja volverá a verse, de todas las parejas que realmente volverán a verse.

Desempeño deseable

El desempeño deseable en producción es una precisión y un recall de al menos 70%.

Perspectiva de negocio

Desde una perspectiva de negocio, se busca que a partir de la caracterización sociodemográfica, intereses y otros atributos de los usuarios; se puedan encontrar parejas afines con las cuales puedan establecer relaciones, garantizando por lo menos un 80% de éxito.

Si el algoritmo no recomienda al usuario parejas o este no coincide con más del 40% de las parejas, se dirá que el modelo no es apto para la recomendación de personas. Luego de poner el modelo en funcionamiento, las personas podrán dar un feedback de cómo fue su experiencia, con el fin de garantizar un aumento en la satisfacción del 10% y la calidad de los matches aumente un 15%.