

Speed dating data set – Primera entrega

Por:

Jose Franco Arroyave

Isaac Esteban Uribe

Jordan Estiven Marulanda

Profesor: Raul Ramos Pollan

Universidad de Antioquia

24/09/2023

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

La necesidad de predecir si dos personas harán match se fundamenta en la búsqueda de la calidad de las personas que se le recomiendan a los usuarios. Al garantizar un 80% de éxito en la predicción de compatibilidad a partir de la caracterización sociodemográfica, intereses y otros atributos de los usuarios, se mejora significativamente la experiencia de estos en la aplicación. Esta justificación se respalda en orientación profesional de psicólogos y se traduce en un aumento del 60% en el tiempo que los usuarios pasan en la plataforma, lo que hace que se dé un incremento en promedio del 10% en los nuevos usuarios y la percepción de la aplicación mejore en un 12% en las plataformas de distribución.