

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.

F1-score: El F1-score es una combinación de la precisión y el recall.

Desempeño deseable

El desempeño deseable en producción es una precisión y un recall de al menos 65-70%.

El objetivo de esta tarea es desarrollar un modelo que pueda predecir si una pareja que se conoce en una cita rápida de cuatro minutos volverá a verse. El modelo se entrenará en un conjunto de datos de citas rápidas que contiene información sobre los participantes en eventos de citas rápidas entre 2002 y 2004. Las métricas de desempeño requeridas son la precisión, el recall y el F1-score. El desempeño deseable en producción es una precisión y un recall de al menos el 80%.