

Reservaciones de hotel

Machine Learning
Perla Conchita Peña Campos





Industria hotelera Portugal

4, 983 millones de euros

VENTAS DE ALOJAMIENTO 2018

67%

PROVIENE DE LA INDUSTRIA
HOTELERA

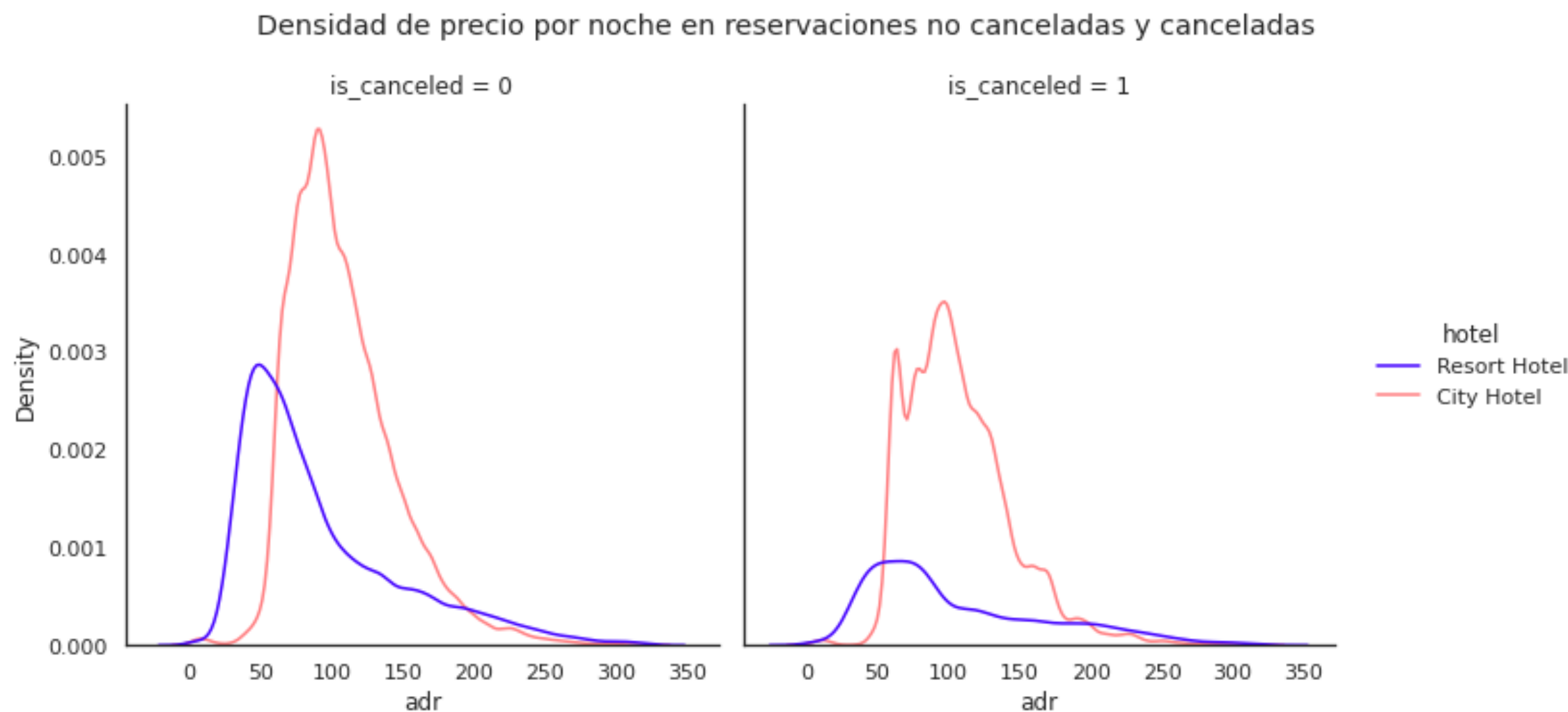
Gestión hotelera

PRINCIPAL PROBLEMA

Contenido

El data set contiene información de dos hoteles de Portugal, resort (H1) y de ciudad (H2). Cuenta con 31 variables que describen 40,060 observaciones para el de resort y 79,330 observaciones para el de ciudad (H2). Comprende información del 01 julio del 2015 al 31 de agosto del 2017

<https://www.sciencedirect.com/science/article/pii/S2352340918315191>



Variables numéricas

- `['lead_time',
'stays_in_weekend_nights',
'stays_in_week_nights',
'adults',
'children',
'babies',
'is_repeated_guest',
'previous_cancellations',
'previous_bookings_not_canceled',
'adr',
'required_car_parking_spaces',
'total_of_special_requests']`

Variables categóricas

- `['hotel',
'arrival_date_month',
'meal',
'market_segment',
'distribution_channel',
'reserved_room_type',
'deposit_type',
'customer_type']`



Modelo predictivo

Predecir la cancelación de una reservación



Training test split y preprocessing

1. Outliers
2. Data set de test y train
3. Preprocessing test y train (Pipeline):
 - a. StandarScaler (numéricas)
 - b. OneHotEncoder (categóricas)



ML

1. Regresión logística
2. Árboles de decisión
3. Random Forest
4. Support Vector Classification
5. Redes neuronales artificiales

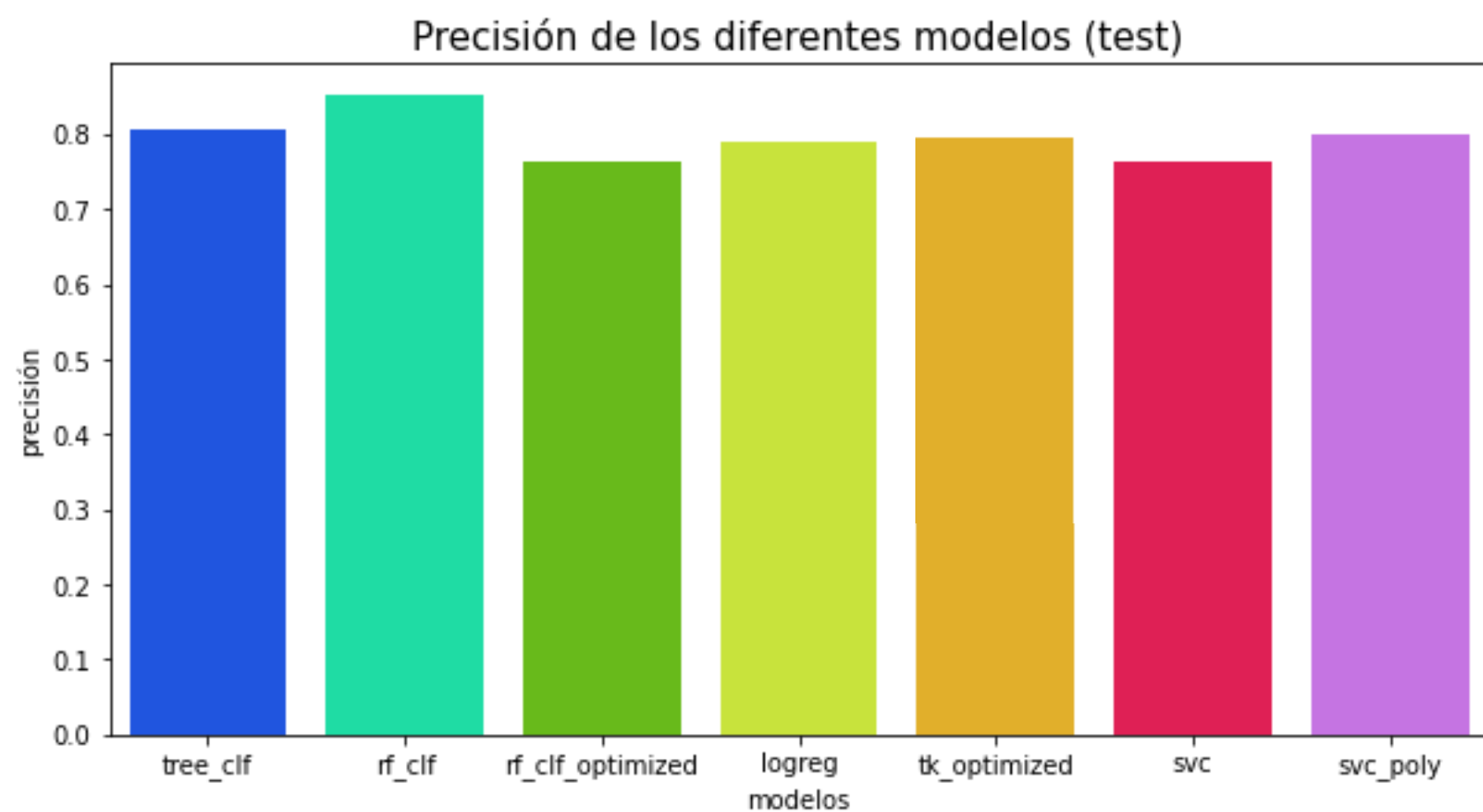
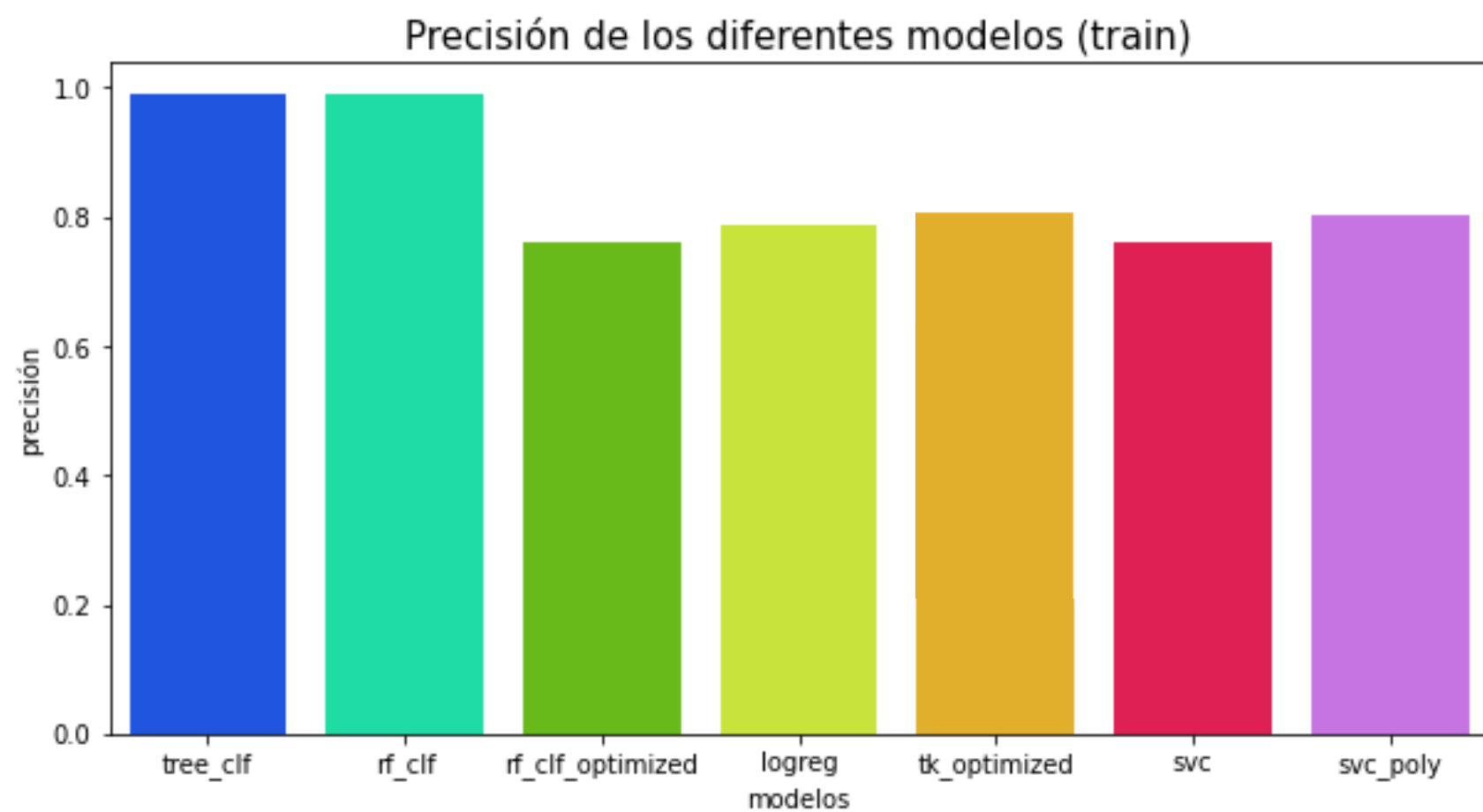


Conclusiones

1. ML extra (K medias, imágenes)
2. Personales

ML
accuracy





Accuracy

Matriz de confusión

Modelos de clasificación

Cancelación de las reservaciones



Modelo	Sensibilidad	Especificidad
Regresión Logística	0.58	0.93
Árbol de decisión	0.74	0.84
Random Forest	0.74	0.91
Random Forest (Tuning)	0.35	0.99
Árbol de decisión	0.74	0.84

Modelo	Sensibilidad	Especificidad
Support Vector Classification	0.60	0.93
Support Vector Classification (Kener Poly)	0.60	0.93
Support Vector Classification (Tuning)	0.63	0.94
Red neuronal artificial (Tuning)	0.61	0.88
Red neuronal artificial (layers Tuning)	0.63	0.91

Modelos ML

Conclusiones



Repositorio