

Proyecto Final

MAT281

Sebastian Rosas

Héctor Silva

December 12, 2024



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Contexto



Visualización descriptiva

ID



Destination



Name



Home Planet



Age



Cabine



Cryo Sleep



VIP



Luxury Amenities



Distribución Variables Numéricas

Visualización descriptiva

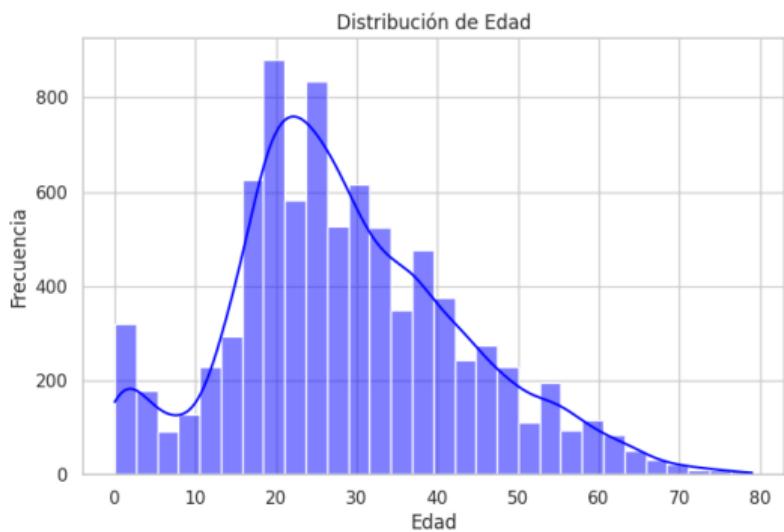


Figure: Distribución de edades

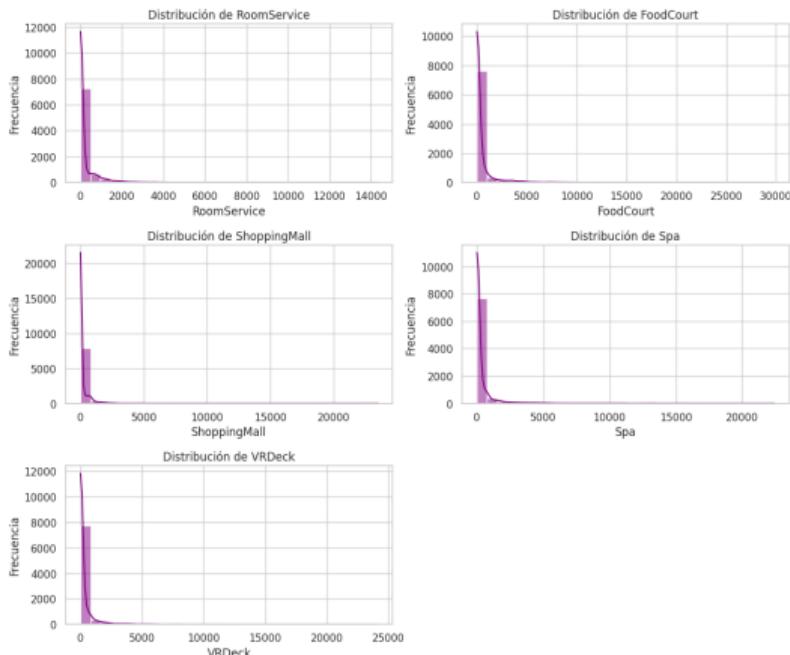


Figure: Distribución Luxury Amenities

Distribución Variables Categóricas

Visualización descriptiva

Distribución de HomePlanet

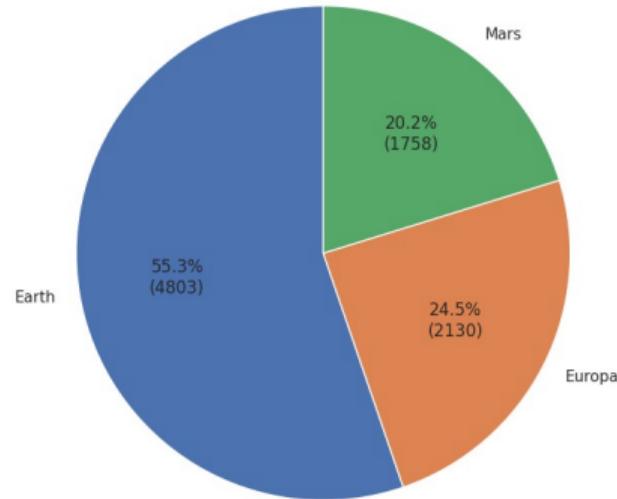


Figure: Distribución de HomePlanet

Distribución de CryoSleep

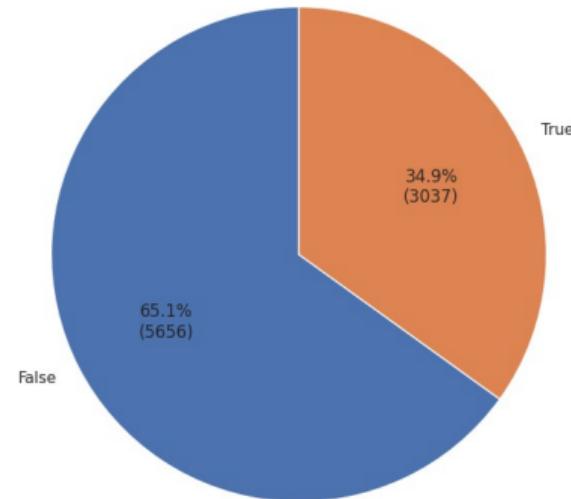


Figure: Distribución de CryoSleep

Distribución Variables Categóricas

Visualización descriptiva

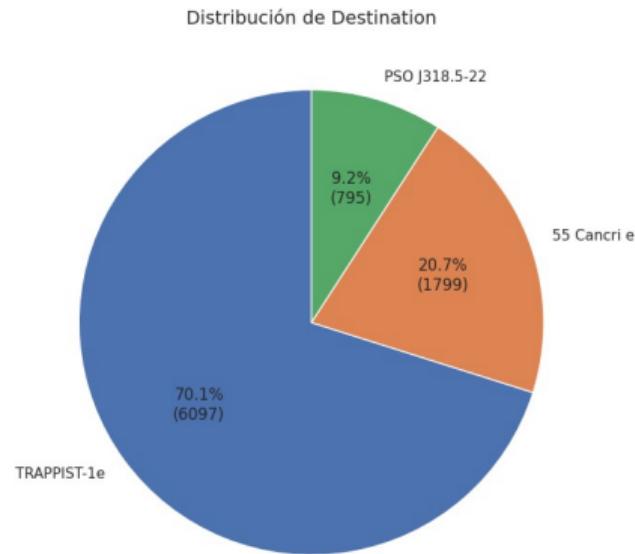


Figure: Distribución de Destination

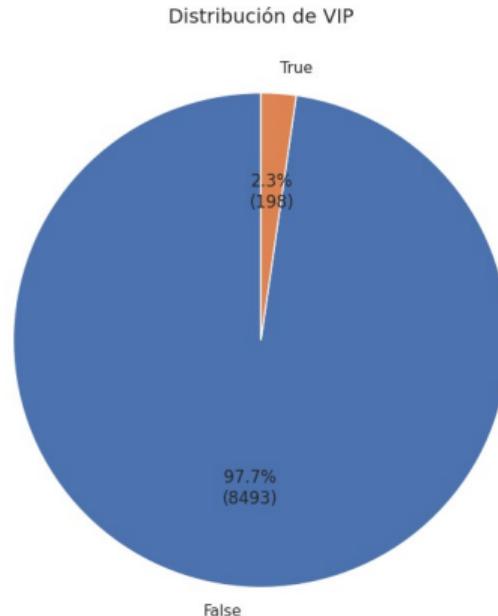


Figure: Distribución de VIP

Distribución Variables combinadas

Visualización descriptiva

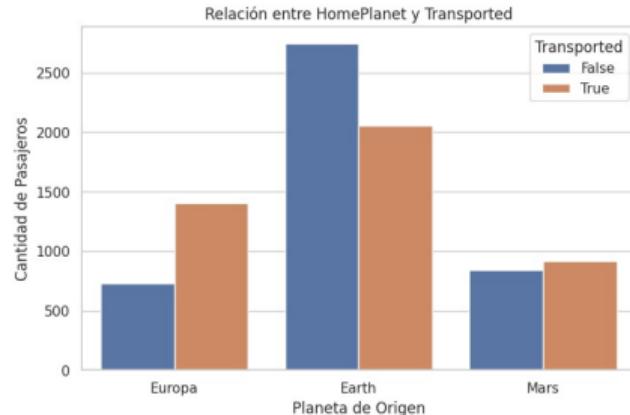


Figure: Relación
HomePlanet-Transported

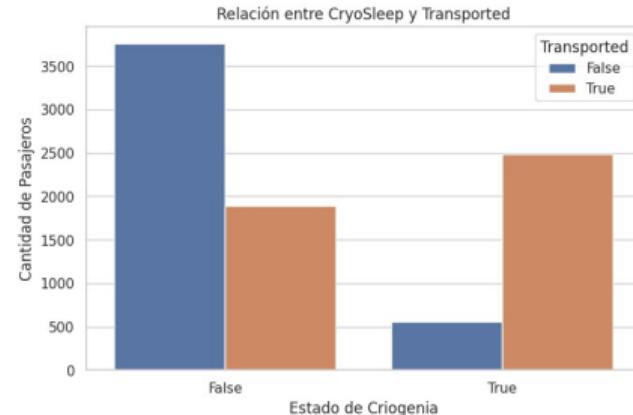


Figure: Relación
CryoSleep-Transported

Distribución Variables Combinadas

Visualización descriptiva

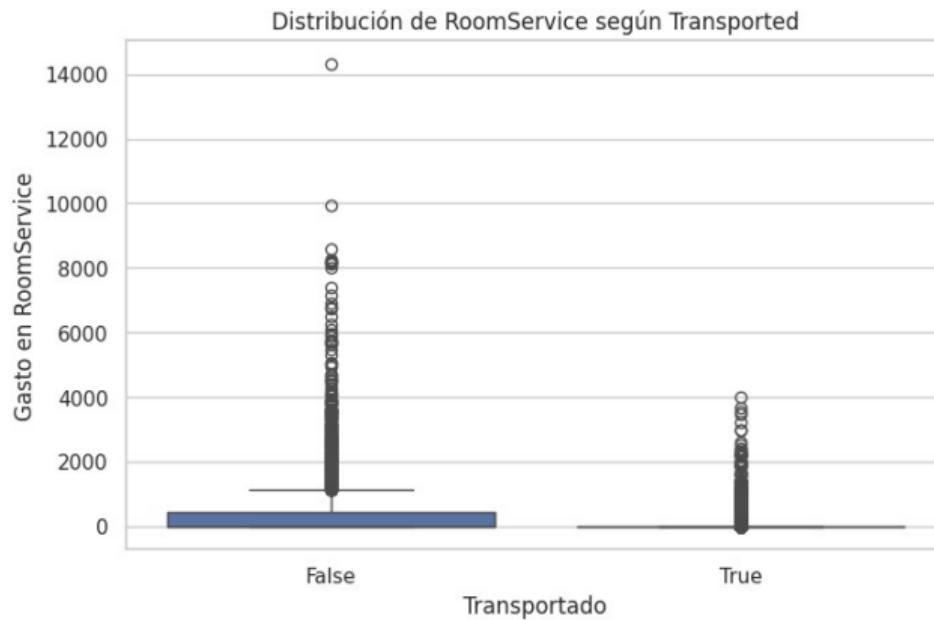


Figure: Distribución de RoomService según Transported

Preprocesamiento de datos



Preprocesamiento

Valores faltantes antes del preprocesamiento

Columna	Valores Faltantes
PassengerId	0
HomePlanet	201
CryoSleep	217
Cabin	199
Destination	182
Age	179
VIP	203

Columna	Valores Faltantes
RoomService	181
FoodCourt	183
ShoppingMall	208
Spa	183
VRDeck	188
Name	200
Transported	0

Preprocesamiento

Relleno de datos

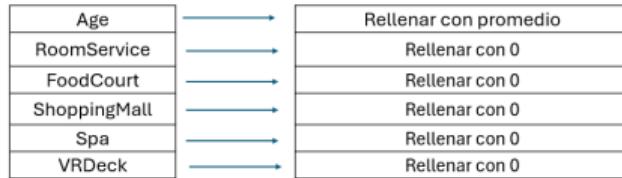


Figure: Atributos numéricicos



Figure: Atributos categóricos

Preprocesamiento

Relleno de datos

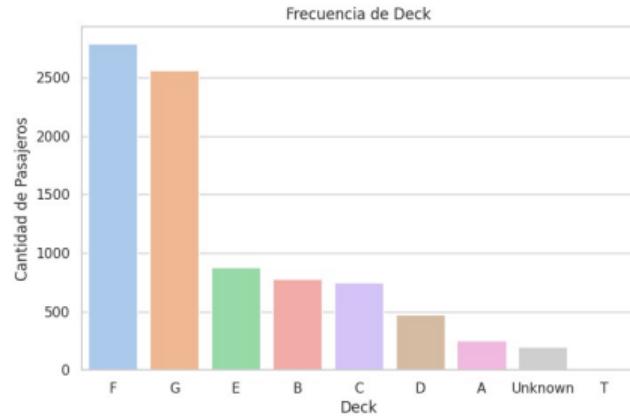


Figure: cantidad de pasajeros por Deck

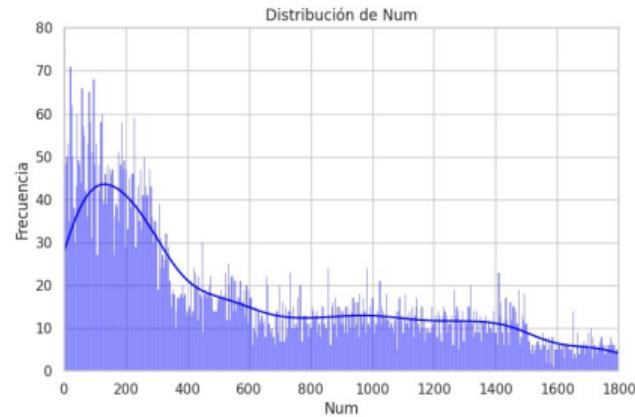


Figure: Distribución de num

Preprocesamiento

Relleno de datos

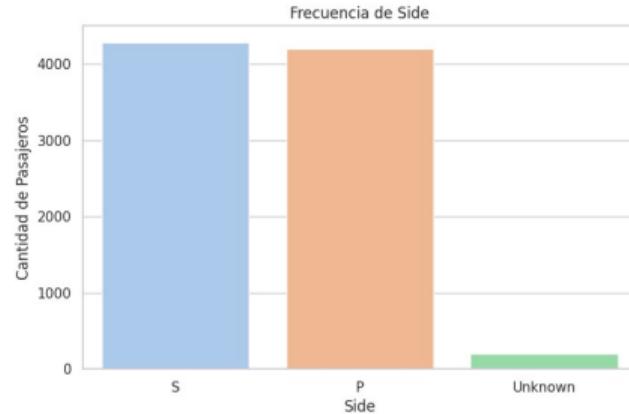


Figure: cantidad de pasajeros por Side

Selección de modelos

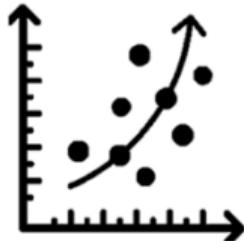


Selección de modelos

modelos aplicados



RF



RL



SVM



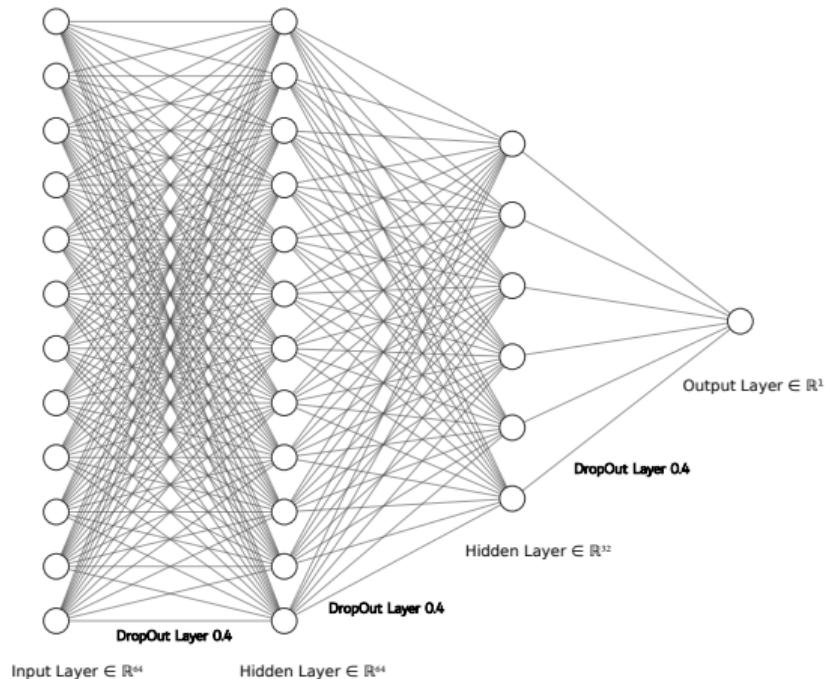
XGB



RN

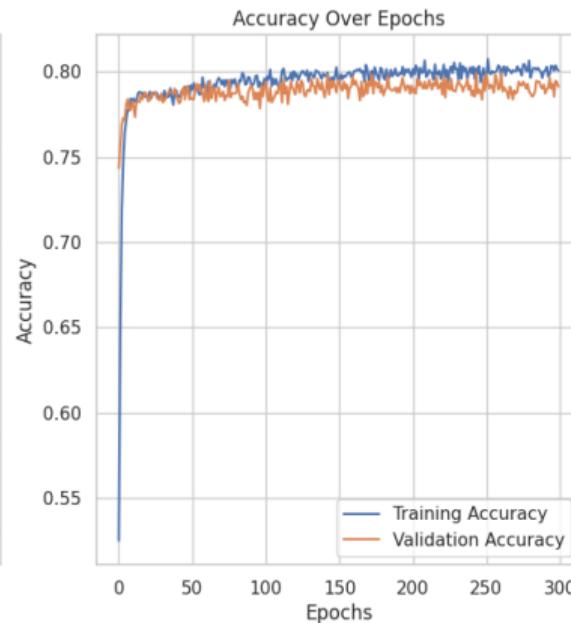
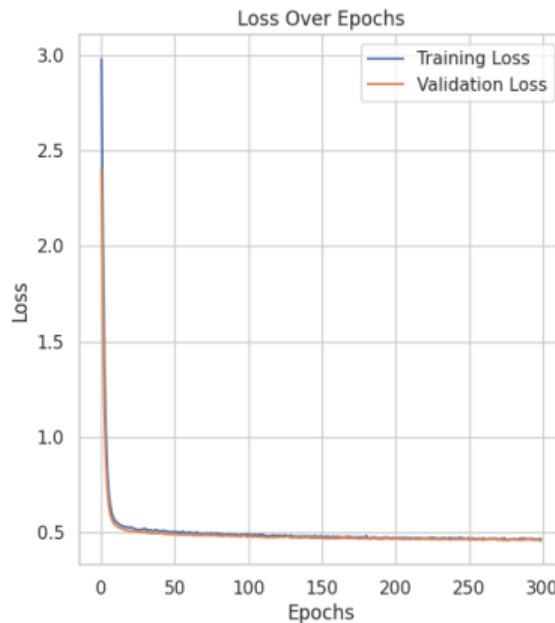
Selección de modelos

Red Neuronal



Selección de modelos

Red Neuronal



Selección de modelos

modelos aplicados

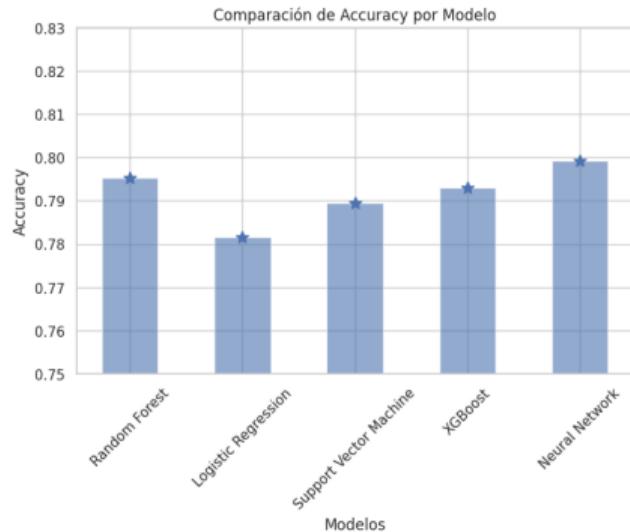


Figure: Accuracy

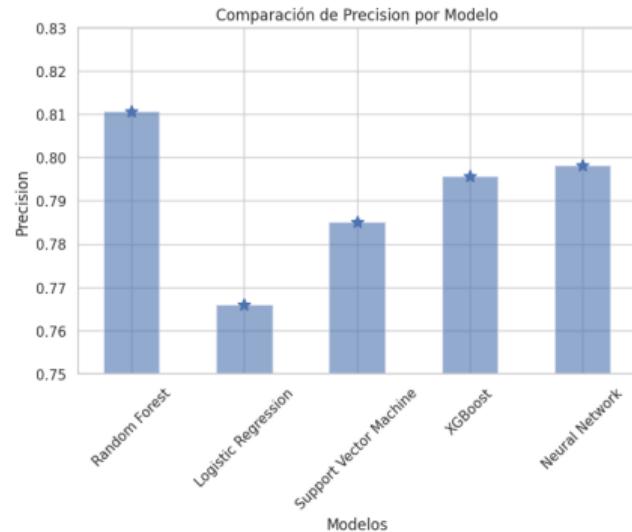


Figure: Precisión

Selección de modelos

modelos aplicados

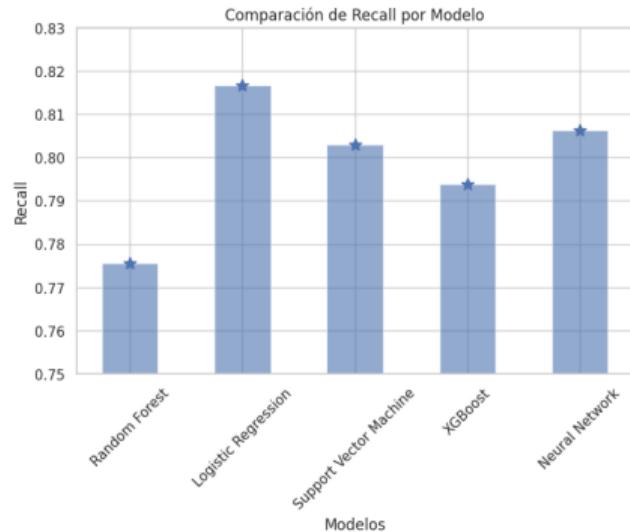


Figure: Recall

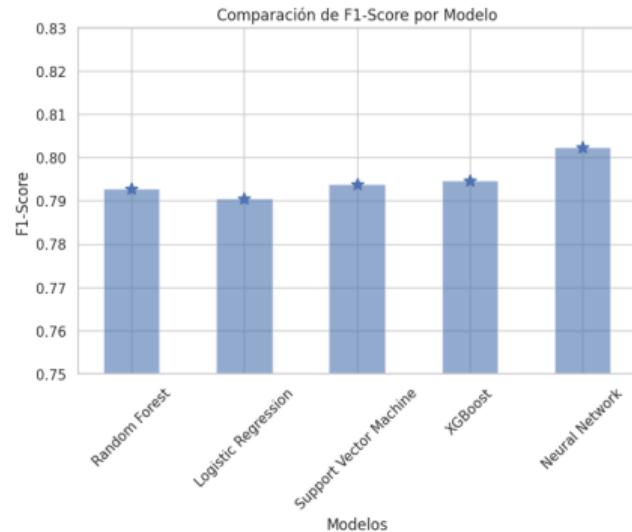


Figure: F1-Score

Selección de modelos

Mejores modelos



Selección de modelos

Importancia de Variables

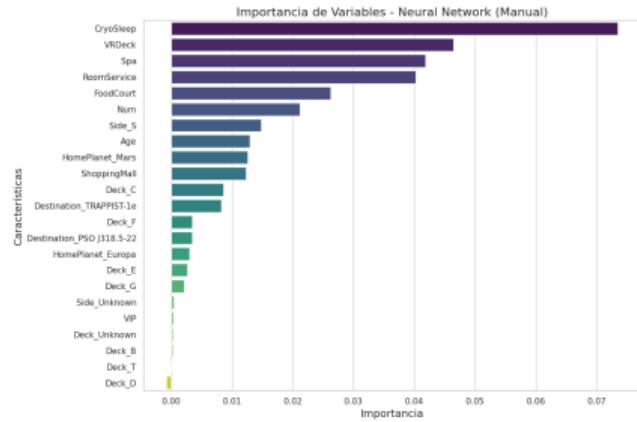


Figure: Red Neuronal

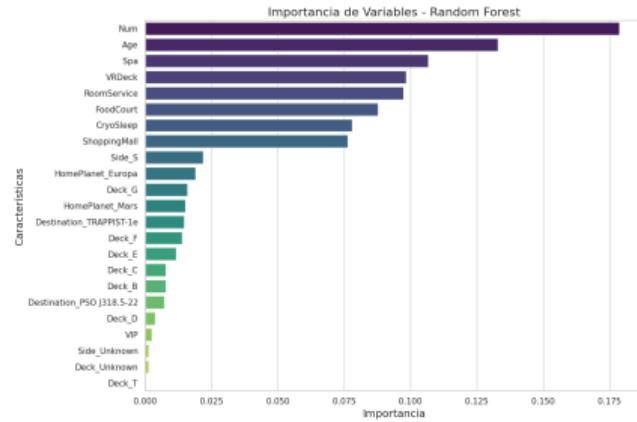


Figure: Random Forest

Selección de modelos

Importancia de Variables

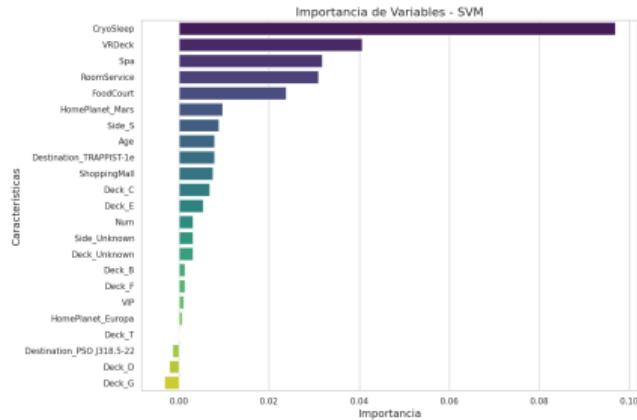


Figure: SVM

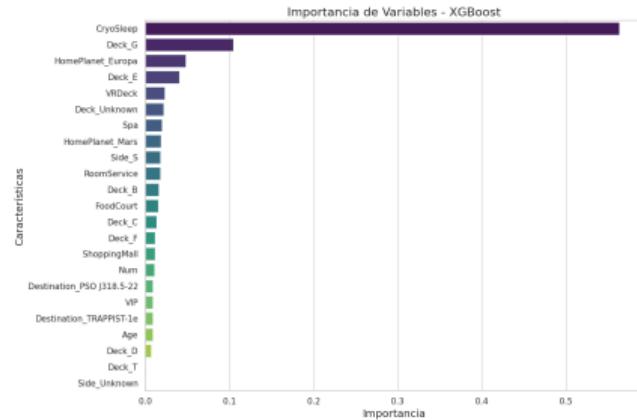


Figure: XGBoost

Optimización de hiperparametros



Optimización de hiperparámetros

GridSearchCV y RandomizedSearchCV

Parámetro	Valor
Modelo	Random Forest
max_depth	10
min_samples_leaf	1
min_samples_split	5
n_estimators	300

Table: Mejores parámetros del modelo con
precisión de validación: 0.8085

Parámetro	Valor
Modelo	Red Neuronal
model_optimizer	adam
model_neurons	32
model_dropout_rate	0.2
epochs	100
batch_size	16

Table: Parámetros de configuración y
precisión de validación: 0.7913

Visualización y métricas



Selección de modelos

modelos aplicados

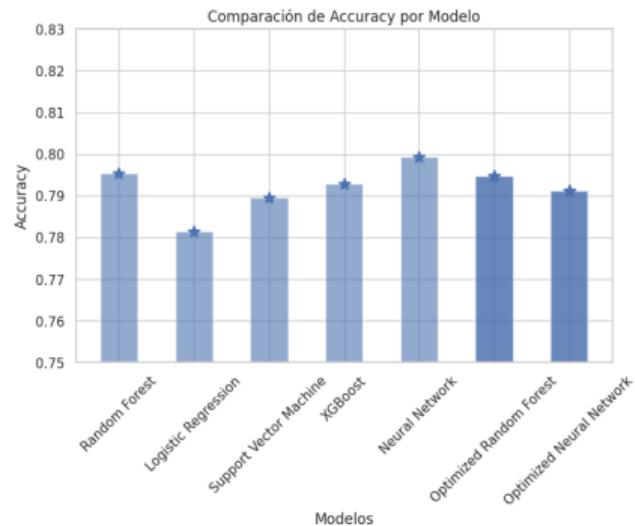


Figure: Accuracy

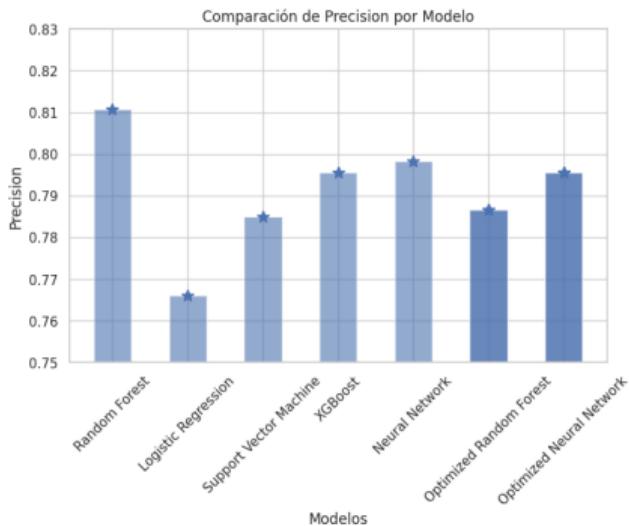


Figure: Precisión

Selección de modelos

modelos aplicados

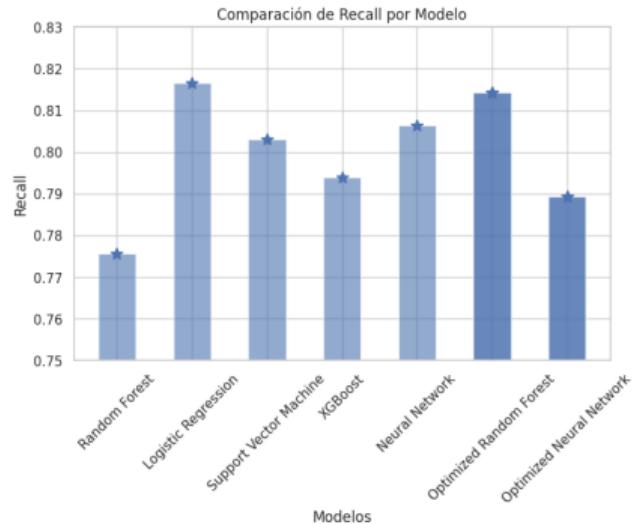


Figure: Recall

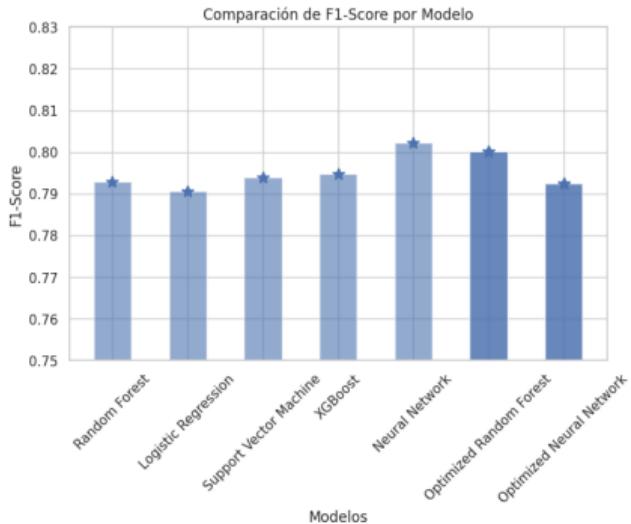


Figure: F1-Score

Selección de modelos

Importancia de Variables

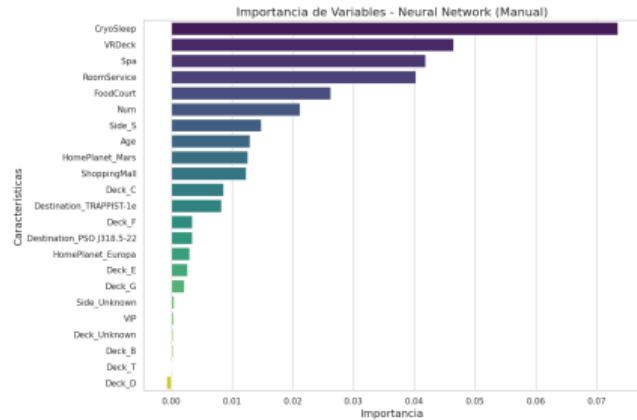


Figure: Red Neuronal

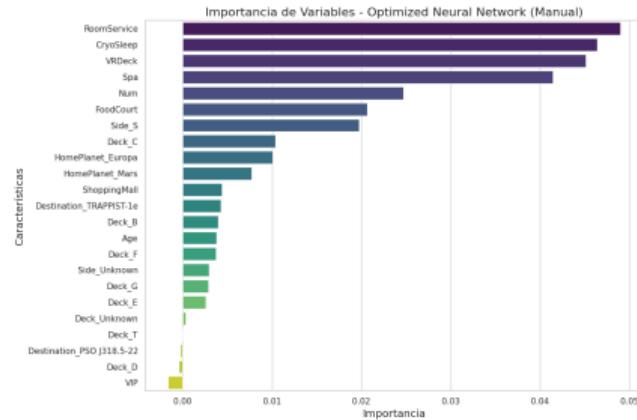


Figure: Red Neuronal Optimizado

Selección de modelos

Importancia de Variables

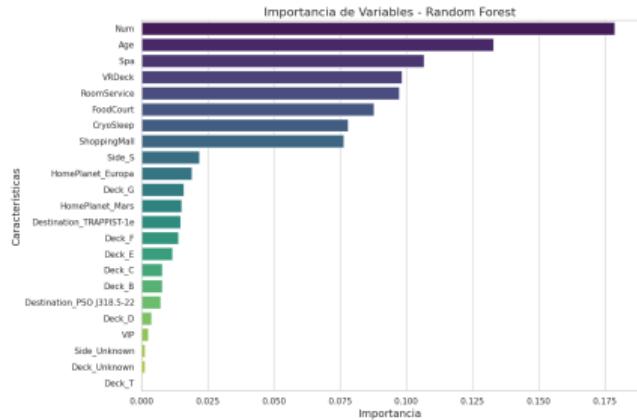


Figure: Random Forest

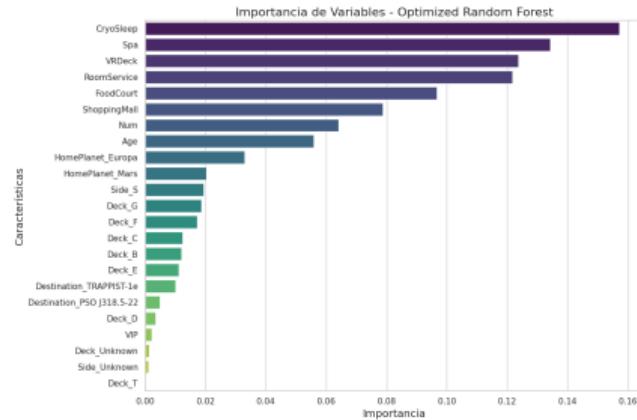


Figure: Random Forest optimizado

Resultados y Conclusiones



Resultados

Model	Score
Random Forest	0,7914
Logistic Regression	0,7921
Support Vector Machine	0,8031
XGBoost	0,7942
Neural Network	0,8024
Optimized Random Forest	0,8036
Optimized Neural Network	0,8064

Table: Comparación de puntajes de modelos



Figure: Mejores modelos

Thank you!