ALGORITMO PREDICCIÓN Primero se cargaran las librerías necesarias In [1]: import pandas as pd import numpy as np from datetime import datetime, timedelta from sklearn.model_selection import train_test_split, RandomizedSearchCV from sklearn.ensemble import StackingRegressor from sklearn.linear_model import ElasticNet from xqboost import XGBRegressor from lightgbm import LGBMRegressor from catboost import CatBoostRegressor from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score from scipy.stats import randint, uniform import plotly.graph_objects as go A continuación se carga el ficheo desde GITHUB SalesOrder.csv In [2]: # Cargar el dataset url = 'https://raw.githubusercontent.com/annaalfaro/TFM/main/SalesOrder.csv' Sales_order = pd.read_csv(url) En el siguiente código se convertián los formatos de los campos en fecha, se obtendrán las características de fecha en Año, mes, día y día de la semana. Se copnvertiran los códigos en nuevos códigos númericos para poder trabajarlos. También se generar las "lag Features" són columnas añadidas: lag_1 -> sería la cantidad definida para el día anterior lag_7 -> cantidad definida para la semana anterior lag_30 -> cantidad definida para el mes anterior Creación de las "rolling" Featuures): también son columnas añadidas rolling_7 -> genera la media de las cantidades durante una semana Rolling_30 -> genera la media de las cantidades durante un mes In [3]: # Convertir las columnas de fechas a formato datetime Sales_order['S0_Date'] = pd.to_datetime(Sales_order['S0_Date'], errors='coerce') Sales_order = Sales_order[['S0_Date', 'S0_CustomerItemid', 'S0_Quantity']] # Crear características adicionales a partir de la fecha Sales_order['Year'] = Sales_order['S0_Date'].dt.year Sales_order['Month'] = Sales_order['SO_Date'].dt.month Sales_order['DayOfYear'] = Sales_order['SO_Date'].dt.dayofyear Sales_order['DayOfWeek'] = Sales_order['SO_Date'].dt.dayofweek # Asegurar que SO_CustomerItemid sea numérico (convertir a códigos) Sales_order['SO_CustomerItemid'] = Sales_order['SO_CustomerItemid'].astype('category').cat.codes # Crear "lag features" Sales_order['Lag_1'] = Sales_order['S0_Quantity'].shift(1) Sales_order['Lag_7'] = Sales_order['S0_Quantity'].shift(7) Sales_order['Lag_30'] = Sales_order['S0_Quantity'].shift(30) # Crear rolling features (promedios móviles de demanda) Sales_order['Rolling_7'] = Sales_order['SO_Quantity'].rolling(window=7).mean() Sales order['Rolling 30'] = Sales order['SO Quantity'].rolling(window=30).mean() # Eliminar filas con valores nulos que se introducen al crear las lag features Sales_order = Sales_order.dropna() Se definen las variables X e Y y los datos de la base de datos se dividen en datos de entrenamiento y de test. In [4]: # Definir las características (X) y la variable objetivo (y) X = Sales_order[['S0_CustomerItemid', 'Year', 'Month', 'DayOfYear', 'DayOfWeek', 'Lag_1', 'Lag_7', 'Lag_30', 'Rolling_7', 'Rolling_30'] y = Sales_order['S0_Quantity'] # Dividir los datos en conjunto de entrenamiento y test X train, X test, y train, y test = train test split(X, y, test size=0.2, random state=42)Modelo de regresión XGBoost In [6]: # Optimización de hiperparámetros para XGBoost param_dist_xgb = { 'n_estimators': randint(50, 300), 'max_depth': randint(3, 10), 'learning_rate': uniform(0.01, 0.2), 'colsample_bytree': uniform(0.6, 0.4) xgb_model = XGBRegressor() #random_search_xgb = RandomizedSearchCV(xgb_model, param_distributions=param_dist_xgb, n_iter=20, scoring='neg_mean_squared_error', cv= random_search_xgb = RandomizedSearchCV(xgb_model, param_distributions=param_dist_xgb, n_iter=20, scoring='neg_mean_squared_error', cv=3 random_search_xgb.fit(X_train, y_train) best_xgb = random_search_xgb.best_estimator_ Modelo LightGBM In [8]: # Optimización de hiperparámetros para LightGBM param_dist_lgb = { 'n_estimators': randint(50, 300), 'num leaves': randint(20, 150), 'learning_rate': uniform(0.01, 0.2), 'feature_fraction': uniform(0.6, 0.4), 'bagging fraction': uniform(0.6, 0.4) lgb_model = LGBMRegressor() #random_search_lgb = RandomizedSearchCV(lgb_model, param_distributions=param_dist_lgb, n_iter=20, scoring='neg_mean_squared_error', cv= random_search_lgb = RandomizedSearchCV(lgb_model, param_distributions=param_dist_lgb, n_iter=20, scoring='neg_mean_squared_error', cv=3 random_search_lgb.fit(X_train, y_train) best_lgb = random_search_lgb.best_estimator_ [LightGBM] [Warning] bagging fraction is set=0.749816047538945, subsample=1.0 will be ignored. Current value: bagging fraction=0.749816 047538945 [LightGBM] [Warning] feature fraction is set=0.9802857225639665, colsample bytree=1.0 will be ignored. Current value: feature fraction= 0.9802857225639665 [LightGBM] [Warning] bagging_fraction is set=0.749816047538945, subsample=1.0 will be ignored. Current value: bagging_fraction=0.749816 047538945 [LightGBM] [Warning] feature_fraction is set=0.9802857225639665, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9802857225639665 [LightGBM] [Warning] bagging_fraction is set=0.6624074561769746, subsample=1.0 will be ignored. Current value: bagging_fraction=0.66240 74561769746 [LightGBM] [Warning] feature_fraction is set=0.662397808134481, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.662397808134481 [LightGBM] [Warning] bagging fraction is set=0.6624074561769746, subsample=1.0 will be ignored. Current value: bagging fraction=0.66240 74561769746 [LightGBM] [Warning] feature_fraction is set=0.662397808134481, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.662397808134481 [LightGBM] [Warning] bagging fraction is set=0.6624074561769746, subsample=1.0 will be ignored. Current value: bagging fraction=0.66240 74561769746 [LightGBM] [Warning] feature_fraction is set=0.662397808134481, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.662397808134481 [LightGBM] [Warning] bagging_fraction is set=0.8404460046972835, subsample=1.0 will be ignored. Current value: bagging_fraction=0.84044 60046972835 [LightGBM] [Warning] feature_fraction is set=0.8832290311184181, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8832290311184181 [LightGBM] [Warning] bagging_fraction is set=0.8404460046972835, subsample=1.0 will be ignored. Current value: bagging_fraction=0.84044 60046972835 [LightGBM] [Warning] feature_fraction is set=0.8832290311184181, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8832290311184181 [LightGBM] [Warning] bagging_fraction is set=0.8404460046972835, subsample=1.0 will be ignored. Current value: bagging fraction=0.84044 60046972835 [LightGBM] [Warning] feature_fraction is set=0.8832290311184181, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8832290311184181 [LightGBM] [Warning] bagging fraction is set=0.9329770563201687, subsample=1.0 will be ignored. Current value: bagging fraction=0.93297 70563201687 [LightGBM] [Warning] feature_fraction is set=0.6849356442713105, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6849356442713105 [LightGBM] [Warning] bagging_fraction is set=0.9329770563201687, subsample=1.0 will be ignored. Current value: bagging_fraction=0.93297 70563201687 [LightGBM] [Warning] feature_fraction is set=0.6849356442713105, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6849356442713105 [LightGBM] [Warning] bagging fraction is set=0.9329770563201687, subsample=1.0 will be ignored. Current value: bagging fraction=0.93297 70563201687 [LightGBM] [Warning] feature_fraction is set=0.6849356442713105, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6849356442713105 [LightGBM] [Warning] bagging_fraction is set=0.8099025726528951, subsample=1.0 will be ignored. Current value: bagging_fraction=0.80990 25726528951 [LightGBM] [Warning] feature_fraction is set=0.7727780074568463, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.7727780074568463 [LightGBM] [Warning] bagging_fraction is set=0.8099025726528951, subsample=1.0 will be ignored. Current value: bagging_fraction=0.80990 25726528951 [LightGBM] [Warning] feature_fraction is set=0.7727780074568463, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.7727780074568463 [LightGBM] [Warning] bagging_fraction is set=0.8099025726528951, subsample=1.0 will be ignored. Current value: bagging fraction=0.80990 25726528951 [LightGBM] [Warning] feature_fraction is set=0.7727780074568463, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.7727780074568463 [LightGBM] [Warning] bagging_fraction is set=0.7824279936868144, subsample=1.0 will be ignored. Current value: bagging_fraction=0.78242 79936868144 [LightGBM] [Warning] feature_fraction is set=0.9140703845572055, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9140703845572055 [LightGBM] [Warning] bagging_fraction is set=0.7824279936868144, subsample=1.0 will be ignored. Current value: bagging_fraction=0.78242 79936868144 [LightGBM] [Warning] feature_fraction is set=0.9140703845572055, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9140703845572055 [LightGBM] [Warning] bagging fraction is set=0.7824279936868144, subsample=1.0 will be ignored. Current value: bagging fraction=0.78242 79936868144 [LightGBM] [Warning] feature_fraction is set=0.9140703845572055, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9140703845572055 [LightGBM] [Warning] bagging_fraction is set=0.786705157299192, subsample=1.0 will be ignored. Current value: bagging_fraction=0.786705 157299192 [LightGBM] [Warning] feature_fraction is set=0.9439761626945282, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9439761626945282 [LightGBM] [Warning] bagging_fraction is set=0.786705157299192, subsample=1.0 will be ignored. Current value: bagging_fraction=0.786705 157299192 [LightGBM] [Warning] feature_fraction is set=0.9439761626945282, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9439761626945282 [LightGBM] [Warning] bagging_fraction is set=0.786705157299192, subsample=1.0 will be ignored. Current value: bagging_fraction=0.786705 157299192 [LightGBM] [Warning] feature_fraction is set=0.9439761626945282, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9439761626945282 [LightGBM] [Warning] bagging_fraction is set=0.9795542149013333, subsample=1.0 will be ignored. Current value: bagging_fraction=0.97955 42149013333 [LightGBM] [Warning] feature_fraction is set=0.9862528132298237, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9862528132298237 [LightGBM] [Warning] bagging_fraction is set=0.9795542149013333, subsample=1.0 will be ignored. Current value: bagging_fraction=0.97955 42149013333 [LightGBM] [Warning] feature fraction is set=0.9862528132298237, colsample bytree=1.0 will be ignored. Current value: feature fraction= 0.9862528132298237 [LightGBM] [Warning] bagging_fraction is set=0.9795542149013333, subsample=1.0 will be ignored. Current value: bagging fraction=0.97955 42149013333 [LightGBM] [Warning] feature_fraction is set=0.9862528132298237, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9862528132298237 [LightGBM] [Warning] bagging_fraction is set=0.6390688456025535, subsample=1.0 will be ignored. Current value: bagging_fraction=0.63906 88456025535 [LightGBM] [Warning] feature fraction is set=0.8736932106048627, colsample bytree=1.0 will be ignored. Current value: feature fraction= 0.8736932106048627 [LightGBM] [Warning] bagging fraction is set=0.6390688456025535, subsample=1.0 will be ignored. Current value: bagging fraction=0.63906 88456025535 [LightGBM] [Warning] feature_fraction is set=0.8736932106048627, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8736932106048627 [LightGBM] [Warning] bagging fraction is set=0.6390688456025535, subsample=1.0 will be ignored. Current value: bagging fraction=0.63906 88456025535 [LightGBM] [Warning] feature_fraction is set=0.8736932106048627, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8736932106048627 [LightGBM] [Warning] bagging_fraction is set=0.6137554084460873, subsample=1.0 will be ignored. Current value: bagging_fraction=0.61375 54084460873 [LightGBM] [Warning] feature_fraction is set=0.9637281608315128, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9637281608315128 [LightGBM] [Warning] bagging_fraction is set=0.6137554084460873, subsample=1.0 will be ignored. Current value: bagging_fraction=0.61375 54084460873 [LightGBM] [Warning] feature_fraction is set=0.9637281608315128, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9637281608315128 [LightGBM] [Warning] bagging fraction is set=0.6137554084460873, subsample=1.0 will be ignored. Current value: bagging fraction=0.61375 54084460873 [LightGBM] [Warning] feature_fraction is set=0.9637281608315128, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9637281608315128 [LightGBM] [Warning] bagging_fraction is set=0.7700623497964979, subsample=1.0 will be ignored. Current value: bagging_fraction=0.77006 23497964979 [LightGBM] [Warning] feature_fraction is set=0.6831766651472755, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6831766651472755 [LightGBM] [Warning] bagging_fraction is set=0.7700623497964979, subsample=1.0 will be ignored. Current value: bagging_fraction=0.77006 23497964979 [LightGBM] [Warning] feature_fraction is set=0.6831766651472755, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6831766651472755 [LightGBM] [Warning] bagging fraction is set=0.7700623497964979, subsample=1.0 will be ignored. Current value: bagging fraction=0.77006 23497964979 [LightGBM] [Warning] feature_fraction is set=0.6831766651472755, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6831766651472755 [LightGBM] [Warning] bagging_fraction is set=0.9757995766256756, subsample=1.0 will be ignored. Current value: bagging_fraction=0.97579 95766256756 [LightGBM] [Warning] feature_fraction is set=0.9579309401710595, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9579309401710595 [LightGBM] [Warning] bagging_fraction is set=0.9757995766256756, subsample=1.0 will be ignored. Current value: bagging_fraction=0.97579 95766256756 [LightGBM] [Warning] feature_fraction is set=0.9579309401710595, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9579309401710595 [LightGBM] [Warning] bagging fraction is set=0.9757995766256756, subsample=1.0 will be ignored. Current value: bagging fraction=0.97579 95766256756 [LightGBM] [Warning] feature_fraction is set=0.9579309401710595, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9579309401710595 [LightGBM] [Warning] bagging_fraction is set=0.6353970008207678, subsample=1.0 will be ignored. Current value: bagging_fraction=0.63539 70008207678 [LightGBM] [Warning] feature_fraction is set=0.6783931449676581, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6783931449676581 [LightGBM] [Warning] bagging fraction is set=0.6353970008207678, subsample=1.0 will be ignored. Current value: bagging fraction=0.63539 70008207678 [LightGBM] [Warning] feature fraction is set=0.6783931449676581, colsample bytree=1.0 will be ignored. Current value: feature fraction= 0.6783931449676581 [LightGBM] [Warning] bagging_fraction is set=0.6353970008207678, subsample=1.0 will be ignored. Current value: bagging fraction=0.63539 70008207678 [LightGBM] [Warning] feature_fraction is set=0.6783931449676581, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6783931449676581 [LightGBM] [Warning] bagging_fraction is set=0.815876852955632, subsample=1.0 will be ignored. Current value: bagging_fraction=0.815876 852955632 [LightGBM] [Warning] feature_fraction is set=0.8347004662655393, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8347004662655393 [LightGBM] [Warning] bagging_fraction is set=0.815876852955632, subsample=1.0 will be ignored. Current value: bagging_fraction=0.815876 852955632 [LightGBM] [Warning] feature_fraction is set=0.8347004662655393, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8347004662655393 [LightGBM] [Warning] bagging fraction is set=0.815876852955632, subsample=1.0 will be ignored. Current value: bagging fraction=0.815876 852955632 [LightGBM] [Warning] feature_fraction is set=0.8347004662655393, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8347004662655393 [LightGBM] [Warning] bagging_fraction is set=0.7103996728090174, subsample=1.0 will be ignored. Current value: bagging_fraction=0.71039 96728090174 [LightGBM] [Warning] feature_fraction is set=0.718509402281633, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.718509402281633 [LightGBM] [Warning] bagging fraction is set=0.7103996728090174, subsample=1.0 will be ignored. Current value: bagging fraction=0.71039 96728090174 [LightGBM] [Warning] feature_fraction is set=0.718509402281633, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.718509402281633 [LightGBM] [Warning] bagging fraction is set=0.7103996728090174, subsample=1.0 will be ignored. Current value: bagging fraction=0.71039 96728090174 [LightGBM] [Warning] feature_fraction is set=0.718509402281633, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.718509402281633 [LightGBM] [Warning] bagging_fraction is set=0.7693605922825478, subsample=1.0 will be ignored. Current value: bagging_fraction=0.76936 05922825478 [LightGBM] [Warning] feature_fraction is set=0.7579526072702278, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.7579526072702278 [LightGBM] [Warning] bagging_fraction is set=0.7693605922825478, subsample=1.0 will be ignored. Current value: bagging_fraction=0.76936 05922825478 [LightGBM] [Warning] feature_fraction is set=0.7579526072702278, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.7579526072702278 [LightGBM] [Warning] bagging_fraction is set=0.7693605922825478, subsample=1.0 will be ignored. Current value: bagging fraction=0.76936 05922825478 [LightGBM] [Warning] feature_fraction is set=0.7579526072702278, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.7579526072702278 [LightGBM] [Warning] bagging_fraction is set=0.88453678109946, subsample=1.0 will be ignored. Current value: bagging_fraction=0.8845367 8109946 [LightGBM] [Warning] feature_fraction is set=0.9160702162124823, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9160702162124823 [LightGBM] [Warning] bagging_fraction is set=0.88453678109946, subsample=1.0 will be ignored. Current value: bagging_fraction=0.8845367 8109946 [LightGBM] [Warning] feature_fraction is set=0.9160702162124823, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9160702162124823 [LightGBM] [Warning] bagging fraction is set=0.88453678109946, subsample=1.0 will be ignored. Current value: bagging fraction=0.8845367 8109946 [LightGBM] [Warning] feature_fraction is set=0.9160702162124823, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9160702162124823 [LightGBM] [Warning] bagging_fraction is set=0.9659838702175123, subsample=1.0 will be ignored. Current value: bagging_fraction=0.96598 38702175123 [LightGBM] [Warning] feature_fraction is set=0.9400154311159197, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9400154311159197 [LightGBM] [Warning] bagging_fraction is set=0.9659838702175123, subsample=1.0 will be ignored. Current value: bagging_fraction=0.96598 38702175123 [LightGBM] [Warning] feature_fraction is set=0.9400154311159197, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9400154311159197 [LightGBM] [Warning] bagging fraction is set=0.9659838702175123, subsample=1.0 will be ignored. Current value: bagging fraction=0.96598 38702175123 [LightGBM] [Warning] feature_fraction is set=0.9400154311159197, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.9400154311159197 [LightGBM] [Warning] bagging_fraction is set=0.7483273008793065, subsample=1.0 will be ignored. Current value: bagging_fraction=0.74832 73008793065 [LightGBM] [Warning] feature_fraction is set=0.8675365010654429, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8675365010654429 [LightGBM] [Warning] bagging_fraction is set=0.7483273008793065, subsample=1.0 will be ignored. Current value: bagging_fraction=0.74832 73008793065 [LightGBM] [Warning] feature_fraction is set=0.8675365010654429, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8675365010654429 [LightGBM] [Warning] bagging_fraction is set=0.7483273008793065, subsample=1.0 will be ignored. Current value: bagging fraction=0.74832 73008793065 [LightGBM] [Warning] feature_fraction is set=0.8675365010654429, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8675365010654429 [LightGBM] [Warning] bagging_fraction is set=0.7098887171960256, subsample=1.0 will be ignored. Current value: bagging_fraction=0.70988 87171960256 [LightGBM] [Warning] feature_fraction is set=0.8244973703390804, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8244973703390804 [LightGBM] [Warning] bagging_fraction is set=0.7098887171960256, subsample=1.0 will be ignored. Current value: bagging_fraction=0.70988 87171960256 [LightGBM] [Warning] feature_fraction is set=0.8244973703390804, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8244973703390804 [LightGBM] [Warning] bagging fraction is set=0.7098887171960256, subsample=1.0 will be ignored. Current value: bagging fraction=0.70988 87171960256 [LightGBM] [Warning] feature_fraction is set=0.8244973703390804, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.8244973703390804 [LightGBM] [Warning] bagging_fraction is set=0.9329770563201687, subsample=1.0 will be ignored. Current value: bagging_fraction=0.93297 70563201687 [LightGBM] [Warning] feature_fraction is set=0.6849356442713105, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6849356442713105 Modelo CatBoost In [9]: # Entrenar CatBoost cat_model = CatBoostRegressor(iterations=500, learning_rate=0.1, depth=6, silent=True) cat_model.fit(X_train, y_train) Out[9]: <catboost.core.CatBoostRegressor at 0x164e32d90> Meta-modelo ElasticNet In [10]: # Definir los modelos base para el stacking estimators = [('xgb', best_xgb), ('lgbm', best_lgb), ('cat', cat_model) # Meta-modelo ElasticNet para el stacking stacking_model = StackingRegressor(estimators=estimators, final_estimator=ElasticNet(alpha=0.1, l1_ratio=0.7)) stacking model.fit(X train, y train) [LightGBM] [Warning] bagging_fraction is set=0.9329770563201687, subsample=1.0 will be ignored. Current value: bagging_fraction=0.93297 70563201687 [LightGBM] [Warning] feature_fraction is set=0.6849356442713105, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6849356442713105 [LightGBM] [Warning] bagging_fraction is set=0.9329770563201687, subsample=1.0 will be ignored. Current value: bagging_fraction=0.93297 70563201687 [LightGBM] [Warning] feature_fraction is set=0.6849356442713105, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6849356442713105 [LightGBM] [Warning] bagging fraction is set=0.9329770563201687, subsample=1.0 will be ignored. Current value: bagging fraction=0.93297 70563201687 [LightGBM] [Warning] feature_fraction is set=0.6849356442713105, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6849356442713105 [LightGBM] [Warning] bagging_fraction is set=0.9329770563201687, subsample=1.0 will be ignored. Current value: bagging_fraction=0.93297 70563201687 [LightGBM] [Warning] feature_fraction is set=0.6849356442713105, colsample_bytree=1.0 will be ignored. Current value: feature_fraction= 0.6849356442713105 Out[10]: | > StackingRegressor lgbm xgb cat ▶ XGBRegressor ► LGBMRegressor ► CatBoostRegressor final_estimator ▶ ElasticNet Predicciones de las demandas futuras desde hoy a un año vista. Generar un rango de fechas, se define una estimación entre 15-60 días despues de la fecha de recepción del la órden. Se calculan con los últimos 30 días las varables lag y rolling (retardos y promedios moviles), estas se añaden al conjunto de datos. PAra la previsión se utiliza un modelo stacking. In [15]: # Predecir demandas futuras (desde hoy hasta un año después) today = datetime.now() futuro df = pd.DataFrame() # Generar un rango de fechas futuras (un año desde hoy) futuro_df['S0_Date'] = pd.date_range(start=today, periods=365, freq='D') # Generar la fecha estimada de entrega para cada SO (entre 15 y 60 días después de SO_Date) futuro_df['SO_EstDate'] = futuro_df['SO_Date'] + pd.to_timedelta(np.random.randint(15, 61, size=len(futuro_df)), unit='D') # Asignar CustomerItemid a las fechas futuras futuro_df['SO_CustomerItemid'] = np.random.choice(data['SO_CustomerItemid'].unique(), len(futuro_df)) # Crear las mismas características para el conjunto de datos futuro futuro_df['Year'] = futuro_df['S0_Date'].dt.year futuro df['Month'] = futuro df['SO Date'].dt.month futuro_df['DayOfYear'] = futuro_df['SO_Date'].dt.dayofyear futuro_df['DayOfWeek'] = futuro_df['SO_Date'].dt.dayofweek # Usar los últimos datos de demanda histórica para crear las lag y rolling features ultimos_datos = data.tail(30) # Tomamos los últimos 30 días de datos históricos # Calcular las lag features para las fechas futuras for i in range(len(futuro_df)): lag_1 = ultimos_datos.iloc[-1]['S0_Quantity'] if not ultimos_datos.empty else 0 lag 7 = ultimos datos.iloc[-7]['SO Quantity'] if len(ultimos datos) >= 7 else 0 lag_30 = ultimos_datos.iloc[-30]['S0_Quantity'] if len(ultimos_datos) >= 30 else 0 rolling_7 = ultimos_datos['SO_Quantity'].rolling(7).mean().iloc[-1] if len(ultimos_datos) >= 7 else 0 rolling_30 = ultimos_datos['SO_Quantity'].rolling(30).mean().iloc[-1] if len(ultimos_datos) >= 30 else 0 futuro df.loc[i, 'Lag 1'] = lag 1 futuro_df.loc[i, 'Lag_7'] = lag_7 futuro_df.loc[i, 'Lag_30'] = lag_30 futuro_df.loc[i, 'Rolling_7'] = rolling_7 futuro_df.loc[i, 'Rolling_30'] = rolling_30 # Actualizar los datos históricos con la predicción para el siguiente día ultimos_datos = pd.concat([ultimos_datos, pd.DataFrame({'SO_Quantity': [lag_1]})], ignore_index=True) # Seleccionar las columnas para la predicción X_futuro = futuro_df[['S0_CustomerItemid', 'Year', 'Month', 'DayOfYear', 'DayOfWeek', 'Lag_1', 'Lag_7', 'Lag_30', 'Rolling_7', 'Rolling # Hacer predicciones de la cantidad solicitada para el año futuro predicciones_futuras = stacking_model.predict(X_futuro) # Añadir las predicciones al DataFrame futuro futuro_df['Predicted_Quantity'] = predicciones_futuras # Ordenar las predicciones por fecha antes de guardarlas futuro_df = futuro_df.sort_values(by='S0_Date') # Guardar las predicciones futuras en un fichero CSV futuro_df[['S0_Date', 'S0_EstDate', 'S0_CustomerItemid', 'Predicted_Quantity']].to_csv('prediccionesS0.csv', index=False) Evaluación del modelo, usando: Error cuadratico medio MSE Raíz del error cuadrático medio RMSE Error absoluto medio MAE error porcentual absoluto medio MAPE • Coeficiente de determinación R^2 In [16]: # Evaluación de las métricas con el conjunto de prueba y_pred_test = stacking_model.predict(X_test) # Calcular las métricas mse = mean_squared_error(y_test, y_pred_test) rmse = np.sqrt(mse) mae = mean absolute error(y test, y pred test) r2 = r2_score(y_test, y_pred_test) mape = np.mean(np.abs((y_test - y_pred_test) / y_test)) * 100 print(f'Métricas de Validación en el Conjunto de Prueba:') print(f'Error Cuadrático Medio (RMSE): {rmse:.2f}') print(f'Error Absoluto Medio (MAE): {mae:.2f}') print(f'Error porcentual absoluto medio (MAPE): {mape:.2f}%') print(f'R-cuadrado (R²): {r2:.2f}') Métricas de Validación en el Conjunto de Prueba: Error Cuadrático Medio (RMSE): 6.56 Error Absoluto Medio (MAE): 5.38 Error porcentual absoluto medio (MAPE): 4.82% R-cuadrado (R^2): 0.75 A continuación se genera una gráfica con las cantidades originales y las pedichas, se ha añadido una marca conteniendo la tendencia media suavizada para no obtener picos muy desorbitados. In [17]: # Crear figura fig = go.Figure() # Ordenar los items únicos por su ID (para asegurar el orden en las gráficas) items_ordenados = sorted(data['SO_CustomerItemid'].unique()) # Graficar demanda histórica y predicciones futuras, intercaladas por artículo for item in items_ordenados: # Gráfico de la demanda histórica por artículo fig.add_trace(go.Scatter(x=data[data['S0_CustomerItemid'] == item]['S0_Date'], y=data[data['S0_CustomerItemid'] == item]['S0_Quantity'], mode='lines+markers', name=f'Histórico Item {item}', line=dict(dash='solid'))) # Gráfico de las predicciones futuras por artículo fig.add_trace(go.Scatter(x=futuro_df[futuro_df['S0_CustomerItemid'] == item]['S0_Date'], y=futuro_df[futuro_df['S0_CustomerItemid'] == item]['Predicted_Quantity'], mode='lines+markers', name=f'Predicción Item {item}', line=dict(dash='dash'))) # Calcular la línea de tendencia general usando media móvil # Concatenar los datos históricos y predicciones futuras total_data = pd.concat([data[['S0_Date', 'S0_Quantity']], futuro_df[['S0_Date', 'Predicted_Quantity']].rename(columns={'Predicted_Quantity': 'S0_Quantity'})]) # Ordenar los datos por fecha para el cálculo de la media móvil total data = total data.sort values('SO Date') # Aplicar una media móvil de 7 días (ajustable) para suavizar los datos media_movil = total_data['S0_Quantity'].rolling(window=7).mean() # Añadir la línea de tendencia suavizada a la gráfica fig.add_trace(go.Scatter(x=total_data['S0_Date'], y=media movil, mode='lines', name='Tendencia Media Suavizada', line=dict(color='firebrick', width=4, dash='dot'))) # Actualizar los ejes y el diseño de la gráfica fig.update_layout(title='Demanda por Artículo (Histórica vs Predicha para el Futuro)', xaxis title='Fecha', yaxis_title='Cantidad', hovermode='x unified', legend_title='Artículos') # Mostrar la gráfica interactiva fig.show() In []: