

Predição de Vendas de Supermercado

Tassiane Barros
Aian Shay



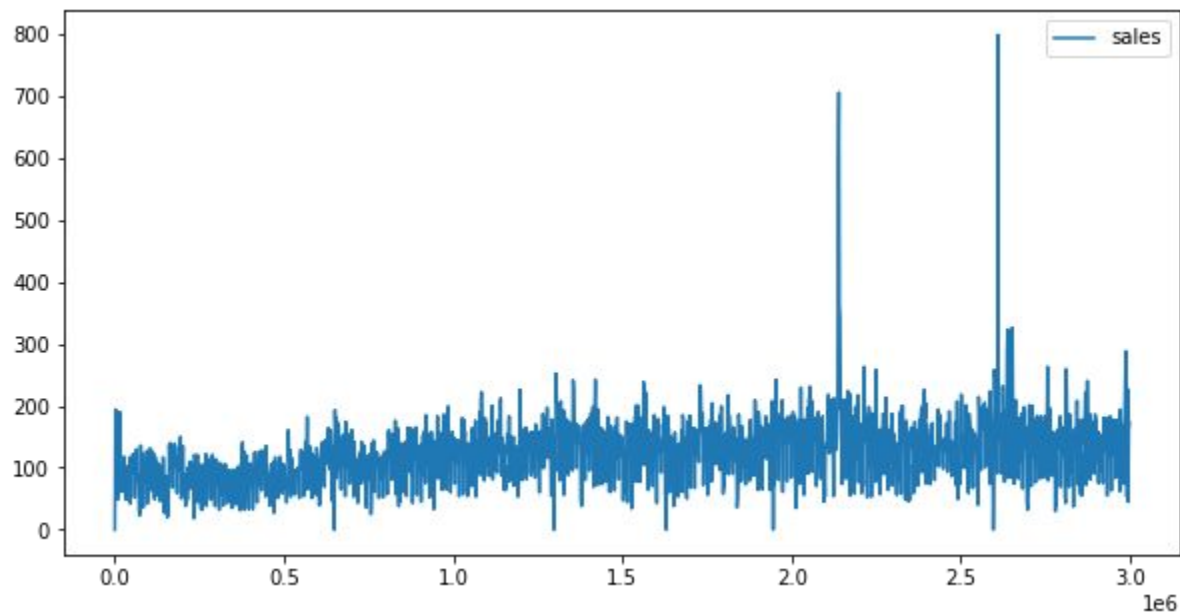


Dataset

- Store Sales
(https://www.kaggle.com/competitions/store-sales-time-series-forecasting/data?select=sample_submission.csv)
 - 2013-2017
 - 1684 registros
- Pré-processamento
 - PERSONAL CARE
 - Remoção de colunas
 - Conversão de data para datetime
 - Sem dados nulos

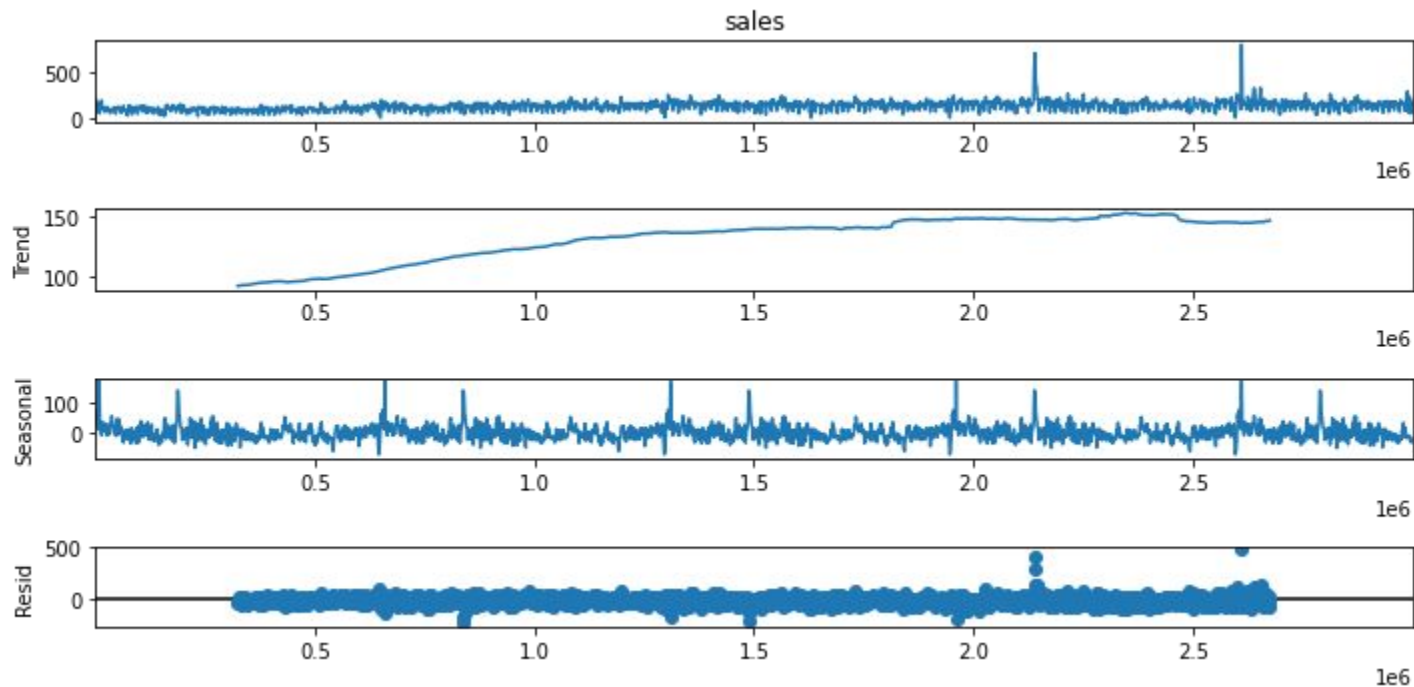


Série Temporal



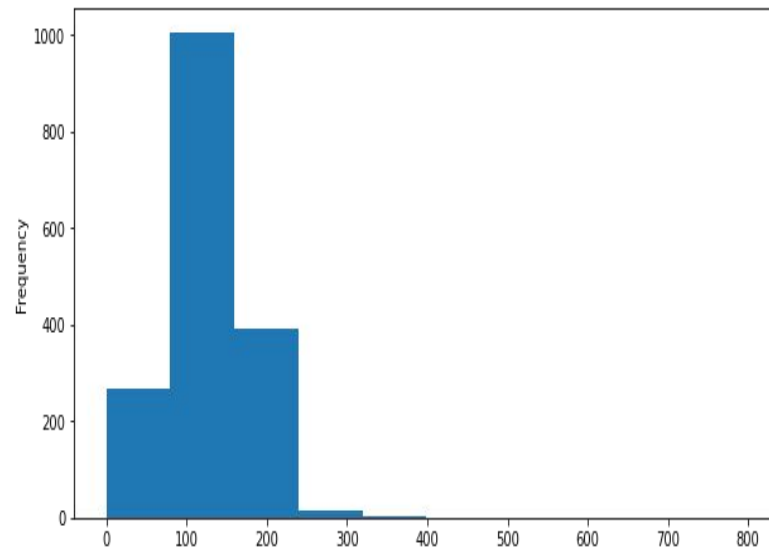
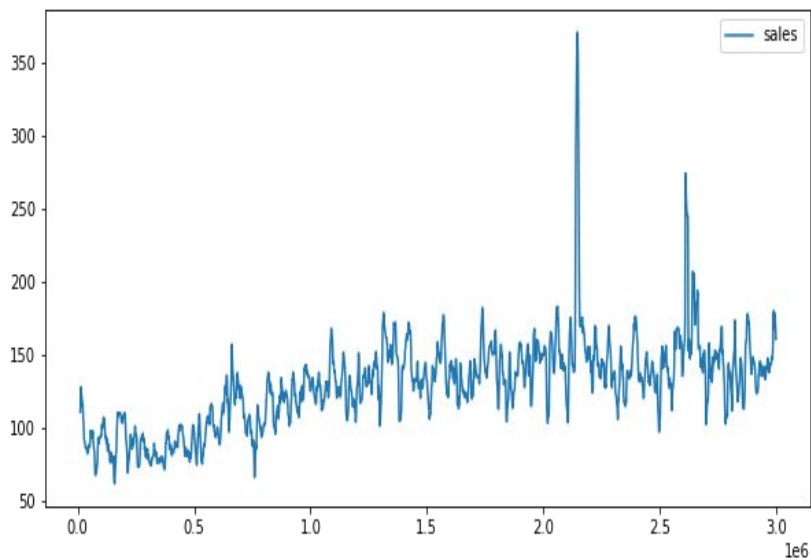


Tendência, sazonalidade e resíduo



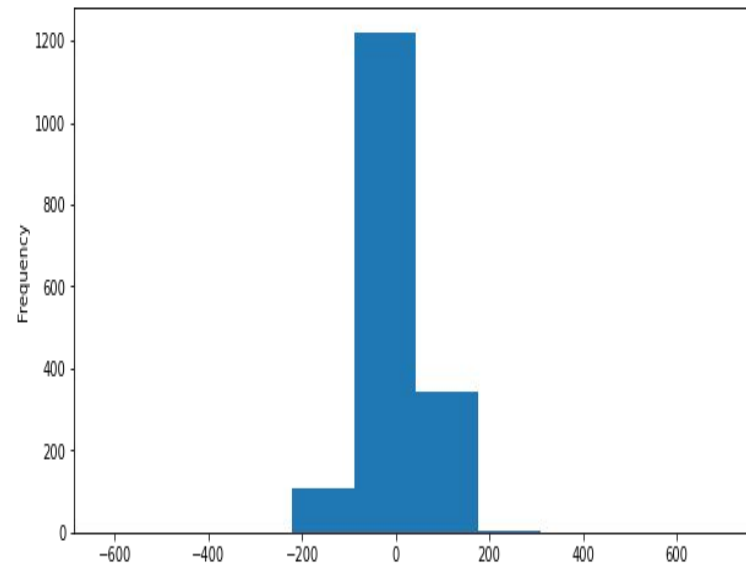
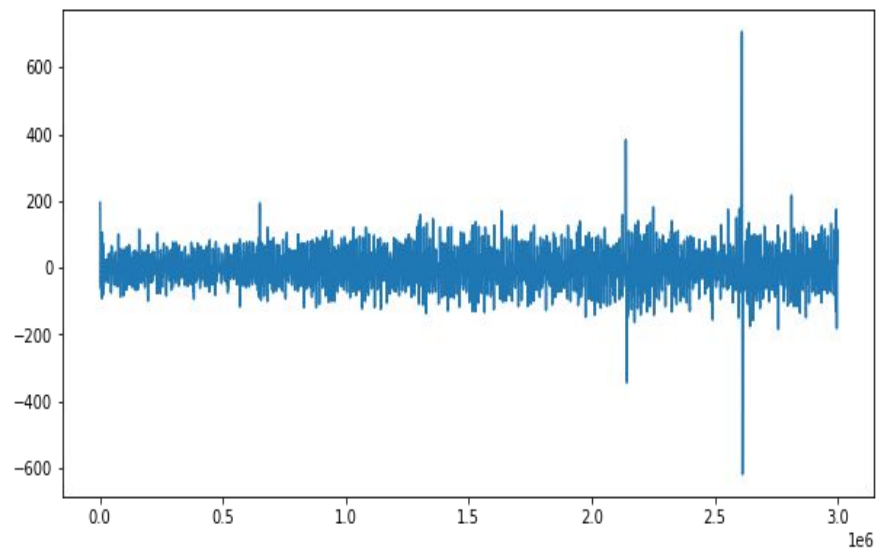


Estacionariedade





Diferenciação



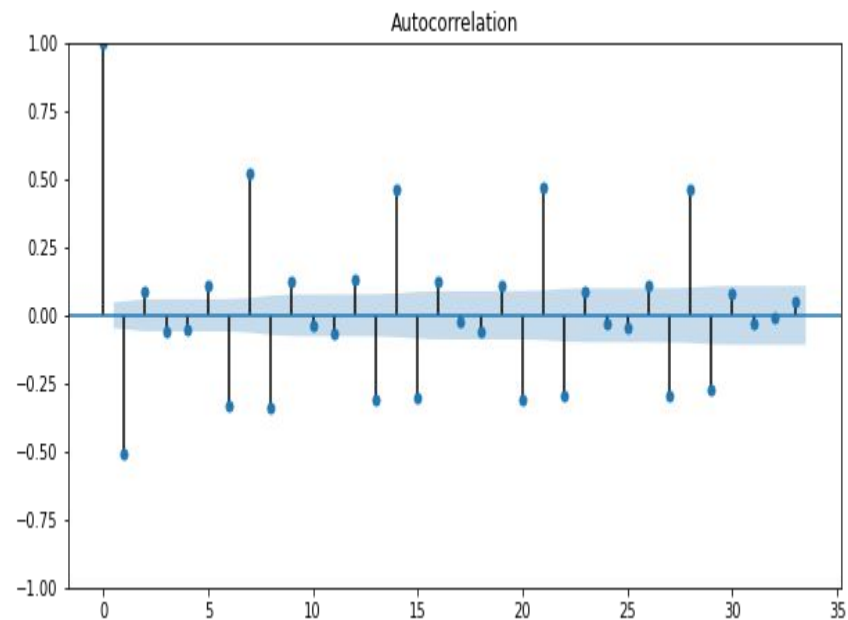
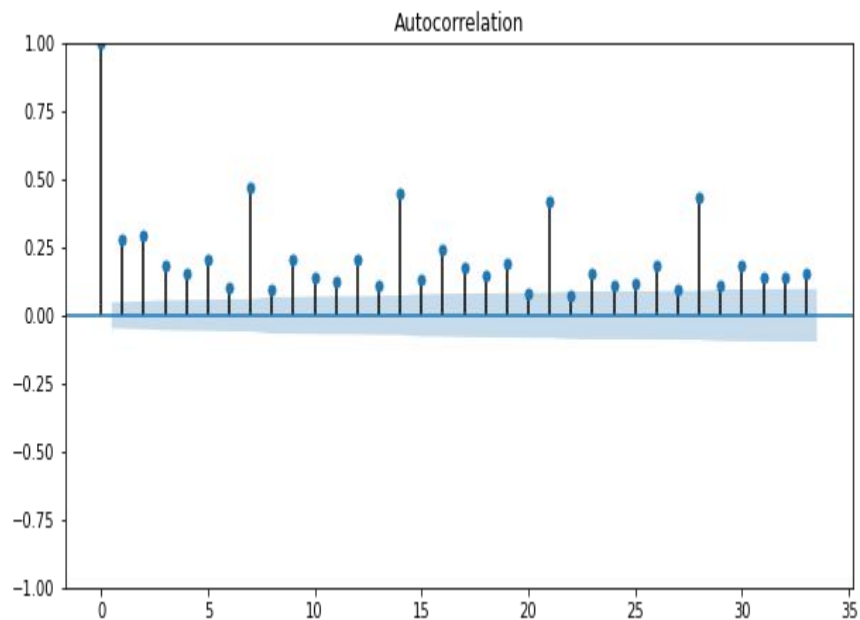


Teste de estacionariedade

- O teste Dickey-Fuller (ADF) rejeita a hipótese nula de não estacionariedade usando 22 lags, o p-valor foi inferior 1%, portanto há evidência estatística de que a série diferenciada é estacionária.
 - p-value = 4.15e-07
 - 22 lags

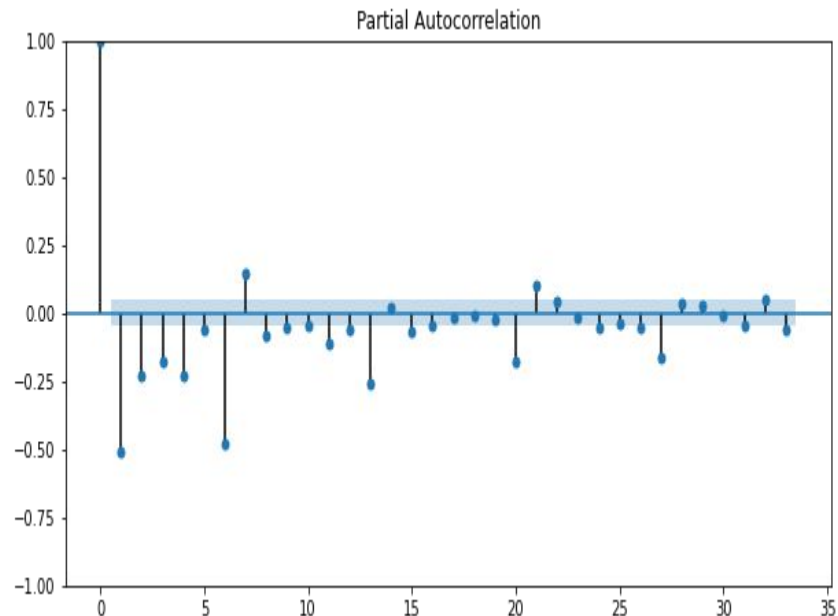
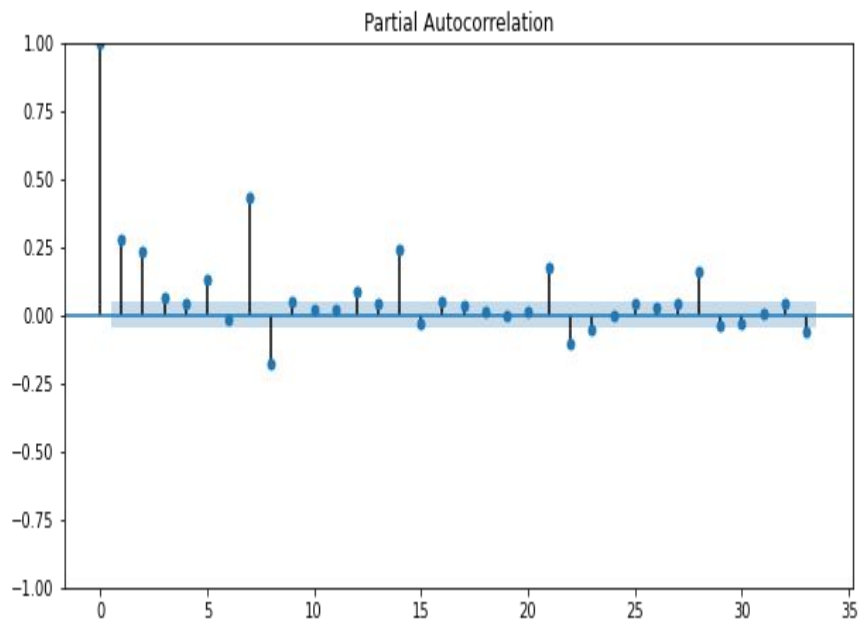


Autocorrelação





Autocorrelação parcial





Modelos ARIMA

- Dados: 75% treino e 25% teste
- Modelos
 - AR
 - (3,0,0)
 - (4,0,0)
 - MA
 - (0,0,7)
 - (0,0,2)
 - ARIMA
 - (3,1,7)
 - (4,1,2)
 - AutoArima
 - (5,1,5)



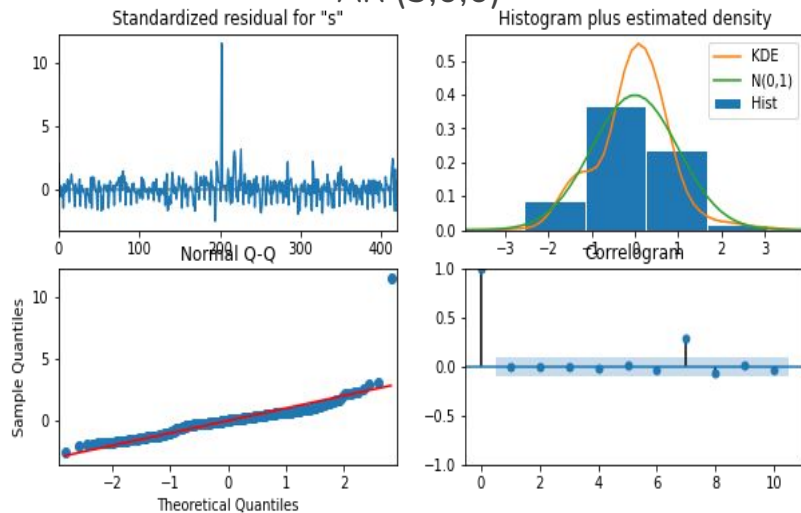
Comparação dos resultados

	AR	MA	ARIMA	Auto ARIMA
mse	3653.744519	3703.701249	3585.841880	3.246770e+03
mae	40.515488	39.482992	39.064980	3.258716e+01
mape	0.307999	0.375769	0.302731	2.759926e+15

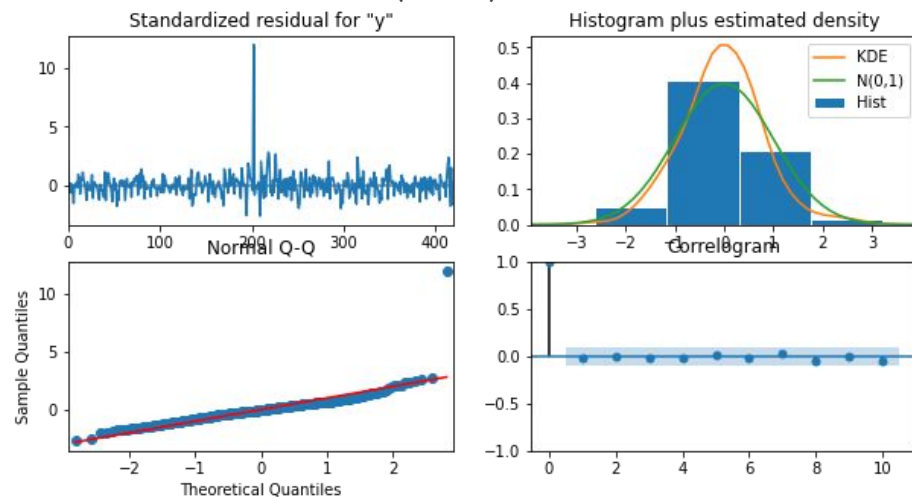


Comparação dos resultados - Resíduos

AR (3,0,0)

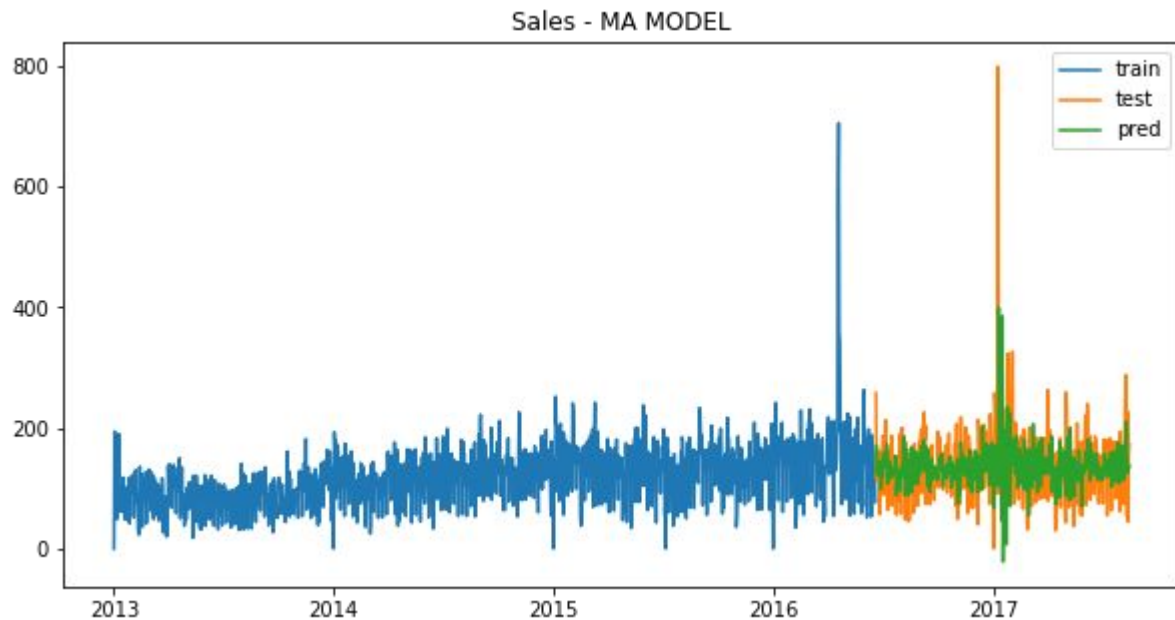


MA (0,0,7)





Predições MA (0,0,7)



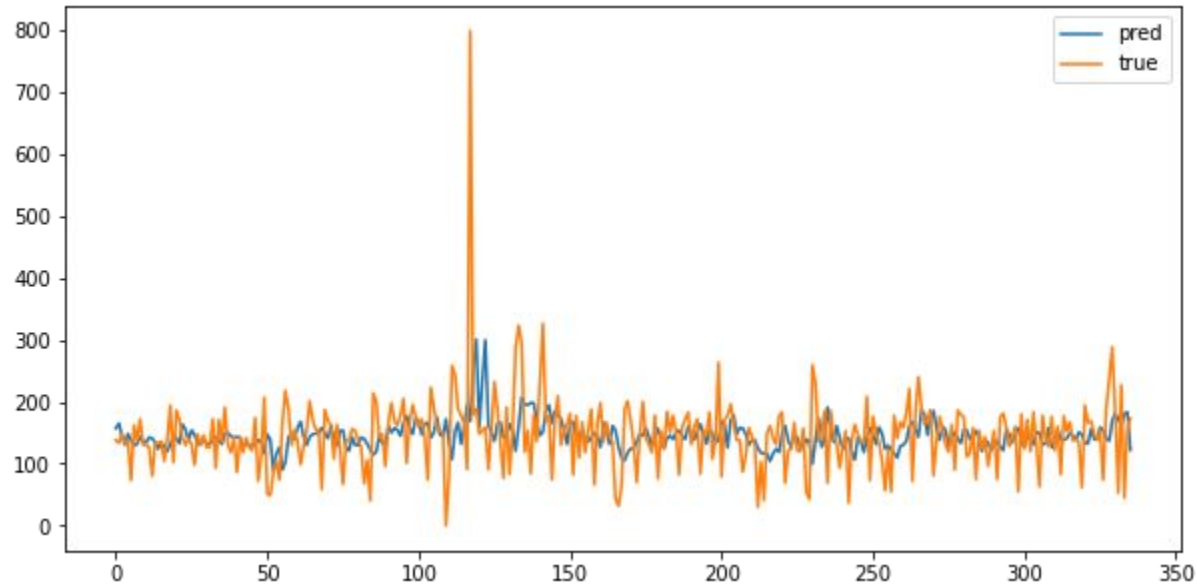


Aprendizado de Máquina

- Inclusão das lags (7)
- SVR
- Redes Neurais
- KNN



Support Vector Regression (SVR)



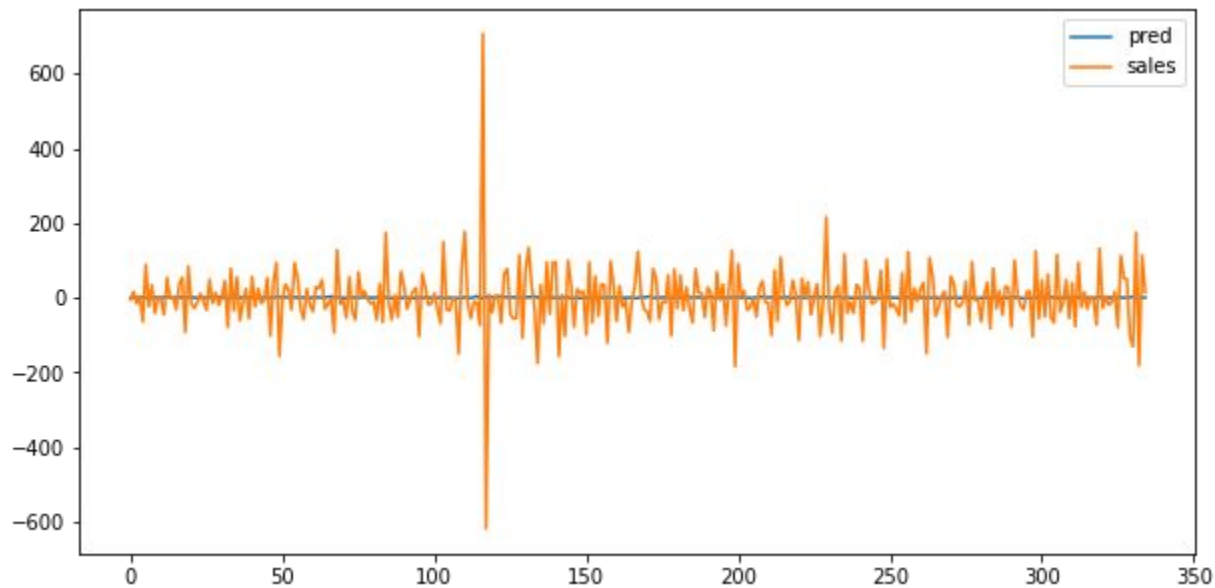


Support Vector Regression (SVR)

	SVR	SVR Multi
mape	0.284605	0.342910
mse	4022.839105	5059.942811
mae	41.331127	46.975467



Redes Neurais





Redes Neurais

MLP

mae	53.747619
mape	578.746037
mse	6801.967028



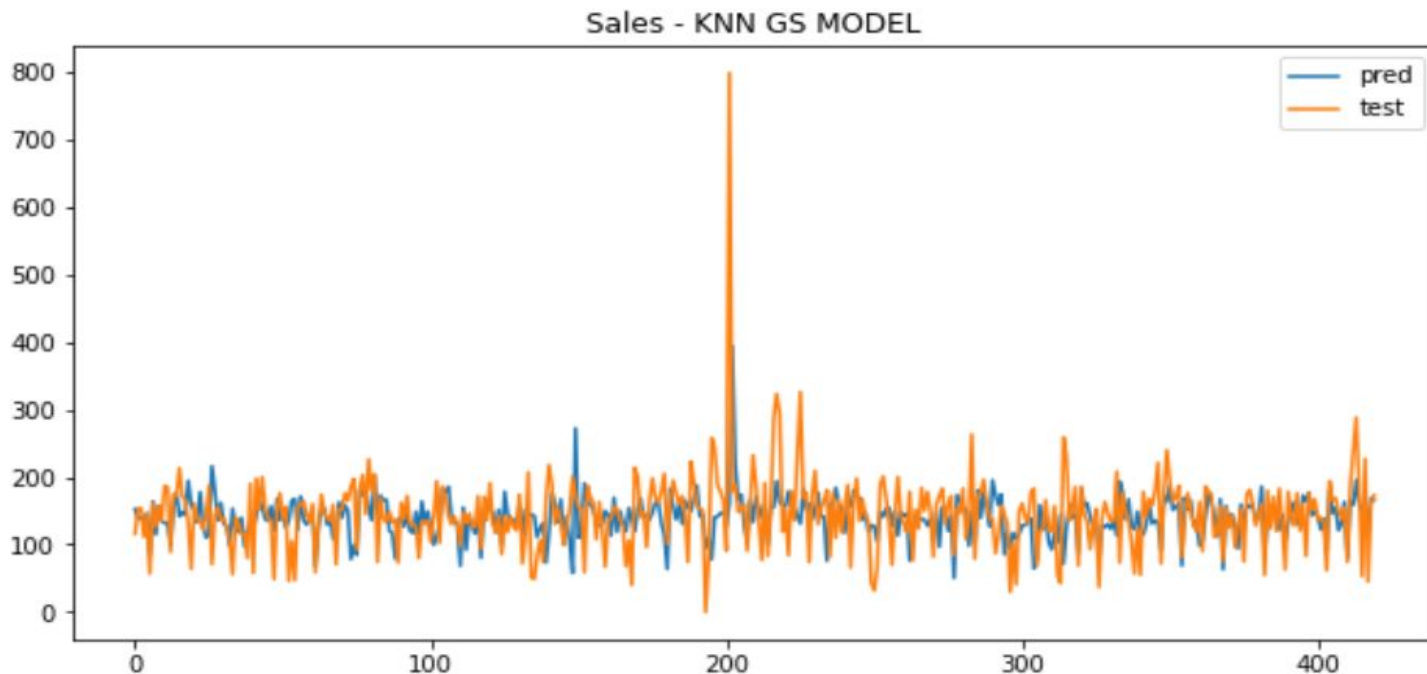
K-NN – Best model

- Parâmetros
 - 'n_neighbors': [5, 7],
 - 'weights': ['uniform', 'distance']
 - 'algorithm': ['ball_tree', 'kd_tree', 'brute', 'auto']



K-NN – Best model

'algorithm': 'kd_tree', 'n_neighbors': 5, 'weights': 'distance'



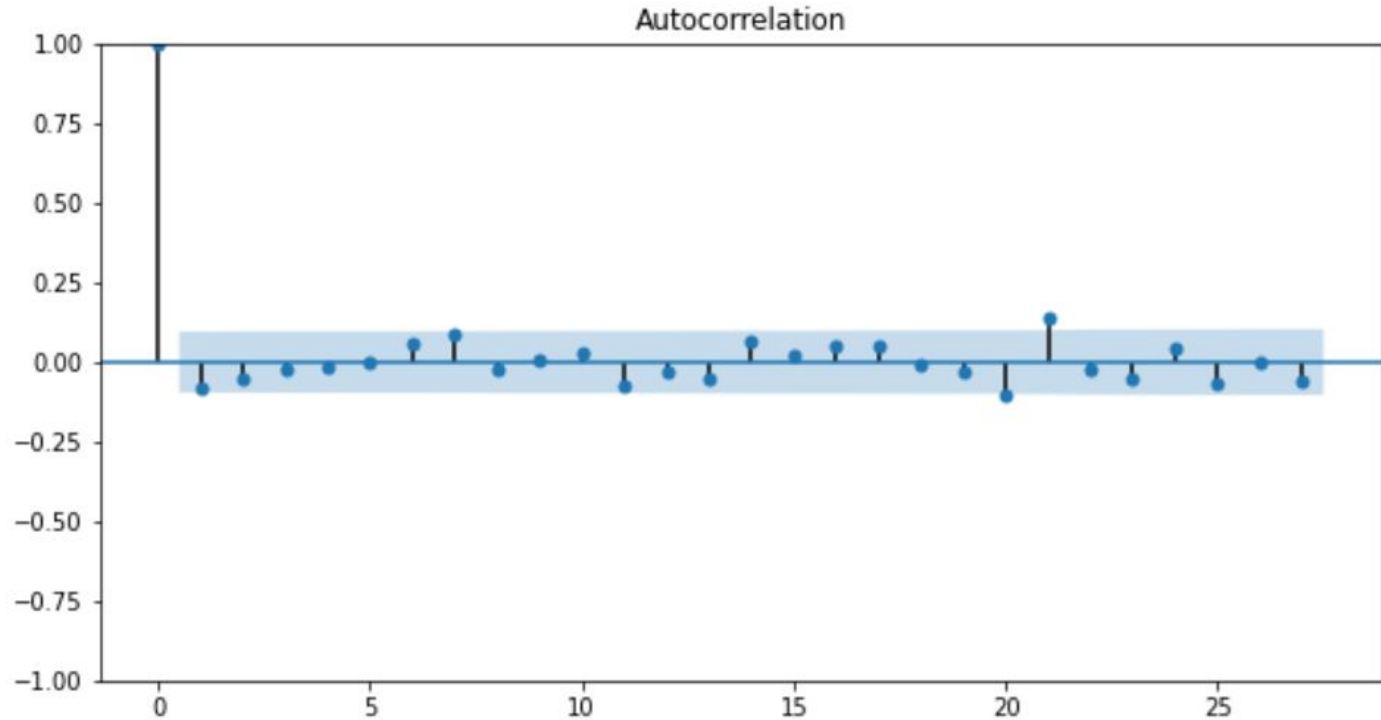


K-NN

	KNN	KNN D1	KNNGS	KNN D1 ONE
mse	3476.005879	43141.353943	3341.105751	11560.754216
mae	36.911038	197.170384	36.549368	76.812800
mape	0.285554	5.748773	0.281892	1.069949



K-NN – Best model (resíduos)





Comparação entre os modelos

	SVR	SVR Multi	MLP		KNN	KNN D1	KNNGS	KNN D1 ONE
mape	0.284605	0.333272	578.746037	mse	3476.005879	43141.353943	3341.105751	11560.754216
mse	4022.839105	5217.499000	6801.967028	mae	36.911038	197.170384	36.549368	76.812800
mae	41.331127	48.061604	53.747619	mape	0.285554	5.748773	0.281892	1.069949

	AR	MA	ARIMA	Auto ARIMA
mse	3653.744519	3703.701249	3585.841880	3.246770e+03
mae	40.515488	39.482992	39.064980	3.258716e+01
mape	0.307999	0.375769	0.302731	2.759926e+15



Obrigado!