

Multivariate Analysis & Prediction of Crude Oil Price Using VAR & LSTM

Group Members

- 1. Archana Yadav**
- 2. Ojaswee Tripathi**
- 3. Rishabh Goswami**

Presented in Dr. Biswajit Patra's Class

Introduction



In this project, we are dealing with Multivariate Time Series Data in which our actionable decision would be to determine whether to Invest or Not in Crude Oil. This dataset is of crude oil derivative market.

Objectives

1

Developing Models
for WTI prediction

We have developed
LSTM-based and VAR-
based models with all
and selected features.

2

Evaluating all the
developed models

Using all the developed
models, we have
predicted values for
validation.

3

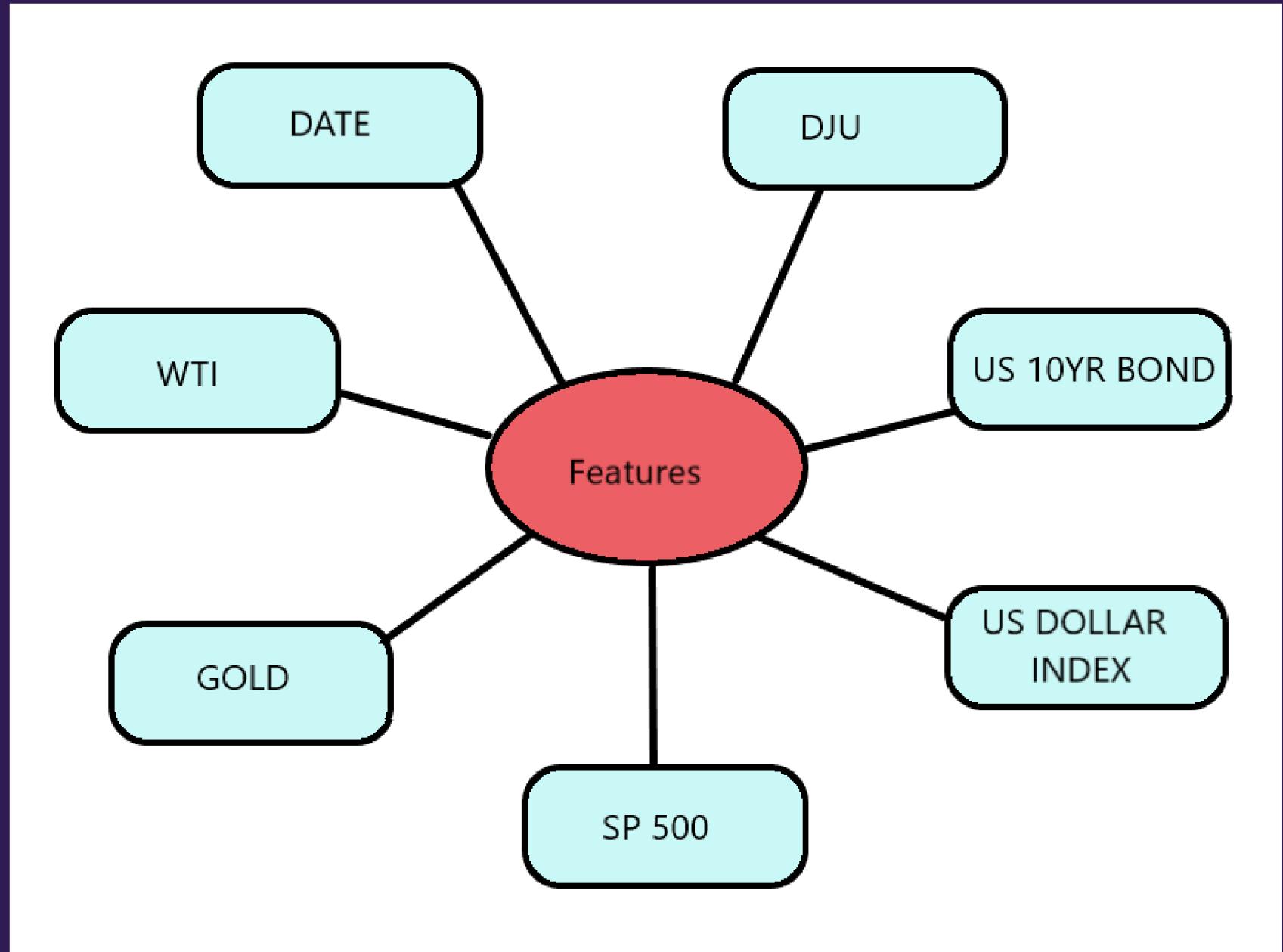
Finding the best
performing Model

Found the best model
based on the it's
performance on
validation set.



We have considered six variables -

1. DATE - from 4th January 2000 to 10th Jaunuary 2019
2. WTI - west texas intermediate
3. GOLD - the price of Gold
4. SP 500 - top 500 companies' overall market stability can be ascertained by S&P 500.
5. US DOLLAR INDEX - As we know that US DOLLAR is a common currency to trade in the international market so foreign exchange market also gets effected, so it is a good metric.
6. US 10YR BONDS - Represents US 10-Year Bond yeild . Is a certificate of loan with federal government that matures in 10 years. Great indicator of investor sentiment about the economy.
7. DJU - Dow Jonas Utility -> 15 important utility stocks,these being sensitive to changes in the price of commodities such as natural gas or coal.



Data Pre-processing

Stationarity -> ADF test

Null values

The dataset contains null values around -
GOLD - 6
SP 500 - 1
US DOLLAR INDEX - 30
US 10YR BOND - 27
DJU - 88

Total attributes - 4947

Removing NULL values -

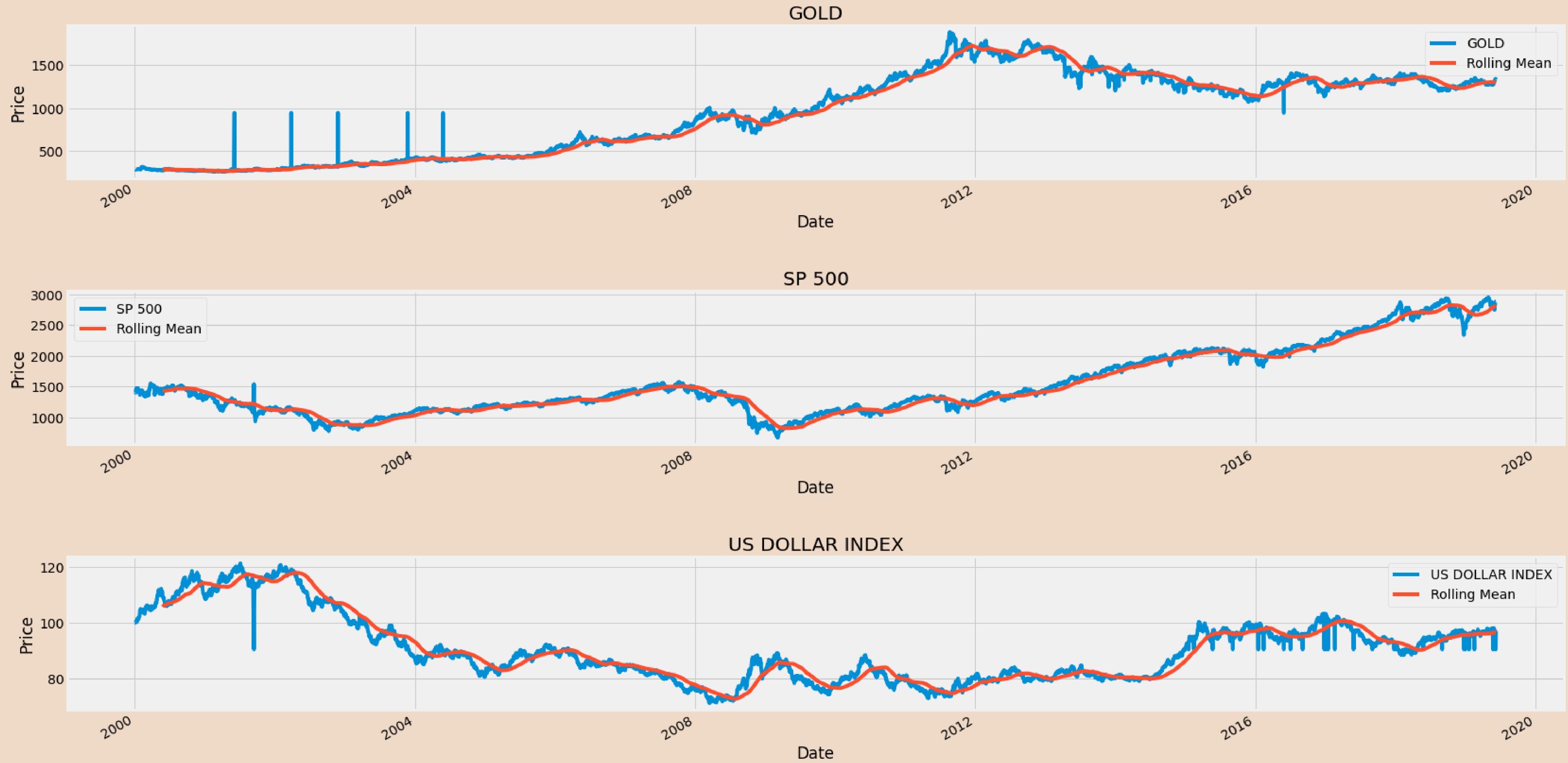
1. using MEAN

In LSTM, we removed NULL values using mean values.

2. removal

In VAR, we removed all null values rows.

EDA



US 10YR BOND

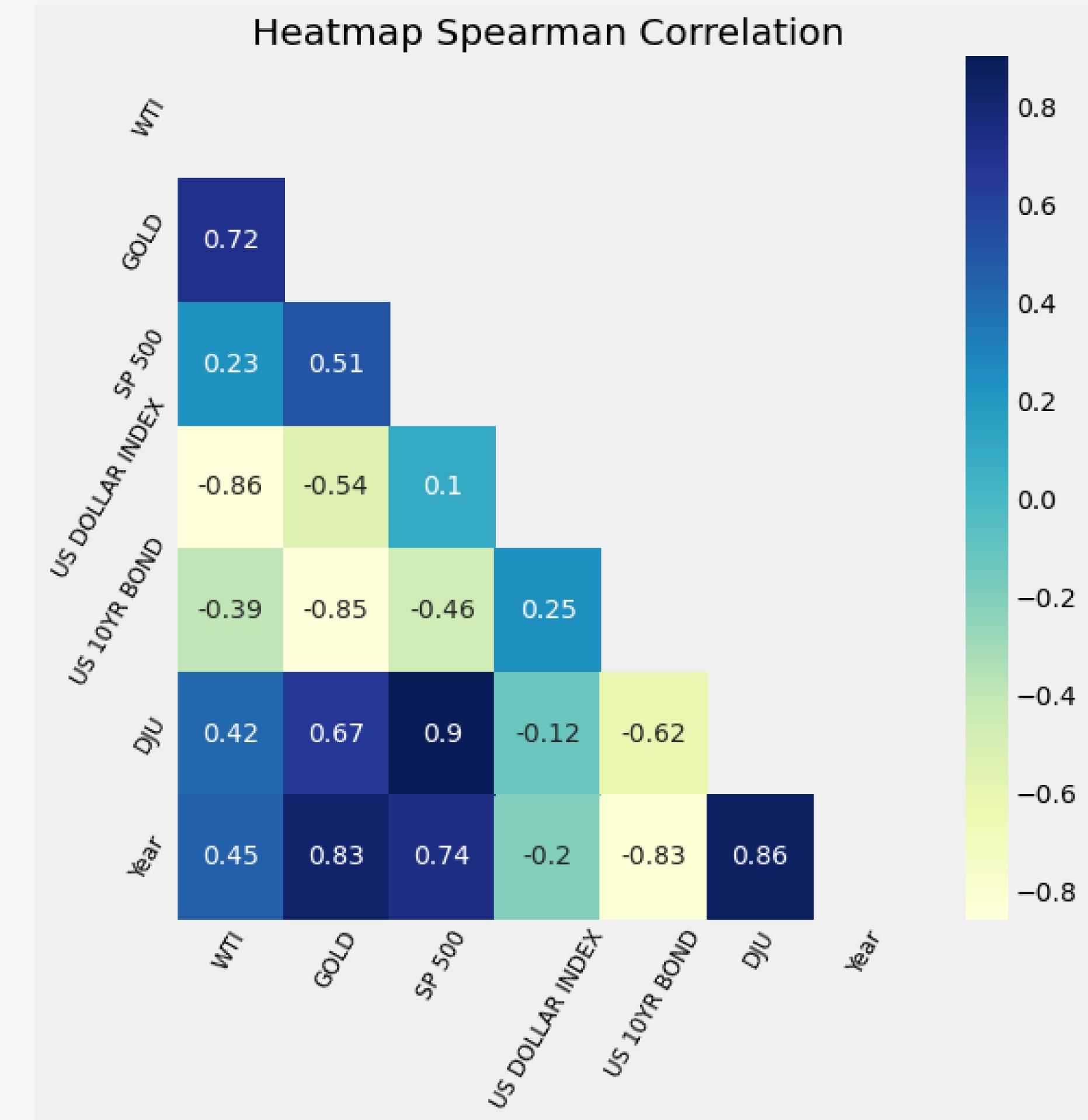


DJU



Correlation graph

shows the correlation between all the attributes present in the dataset. We did feature selection based on this and selected WTI, GOLD, and US DOLLAR INDEX.



OUTLIER REMOVAL TECHNIQUES

Financial crisis outlier removal

The US faced a financial crisis around 2007-2009. There was a contraction of liquidity in the US market.

--> 482 outliers

z-score outlier removal

It tells about the standard deviation of each point from the mean.

$z\text{-score} > 2.5$
--> 37 outliers

Mahalanobis distance outlier removal

find the distance of points in multivariate space.
 $\text{mahalanobis_dis} > 8$
--> 145 outliers



LSTM

Compared all three outlier techniques
using the LSTM model.

1

Deep neural network

2.

3 LSTM layers + 1
hidden layers

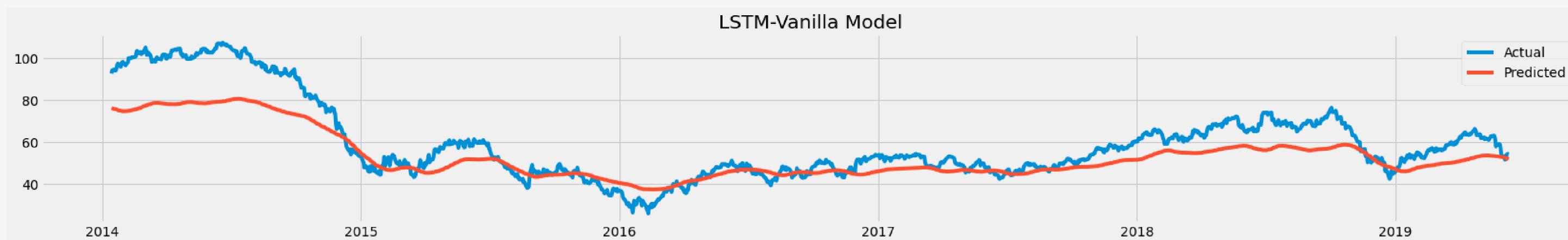
3.

Hyper-parameter :
Early stopping

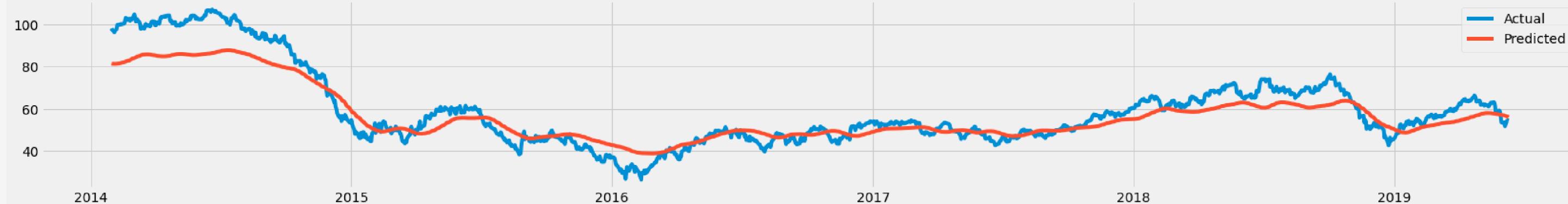
4.

Optimisation - Adam

LSTM-Vanilla Model



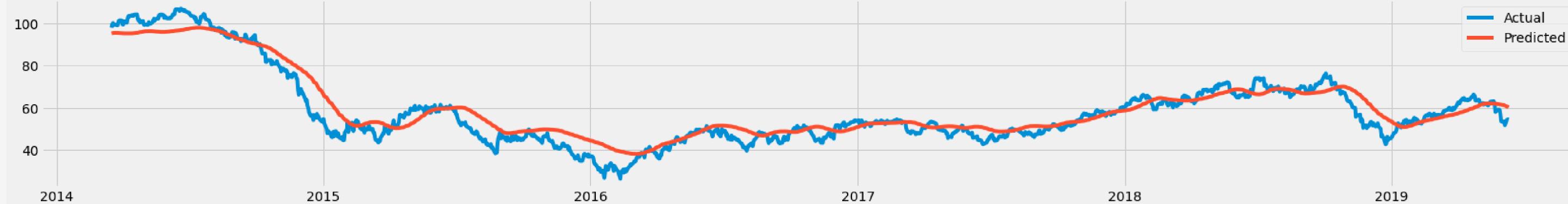
LSTM-ZScore Outlier Removal Technique



LSTM-Financial Outlier Removal Technique



LSTM-Forecasting-Mahalanobis Outlier Removal Technique



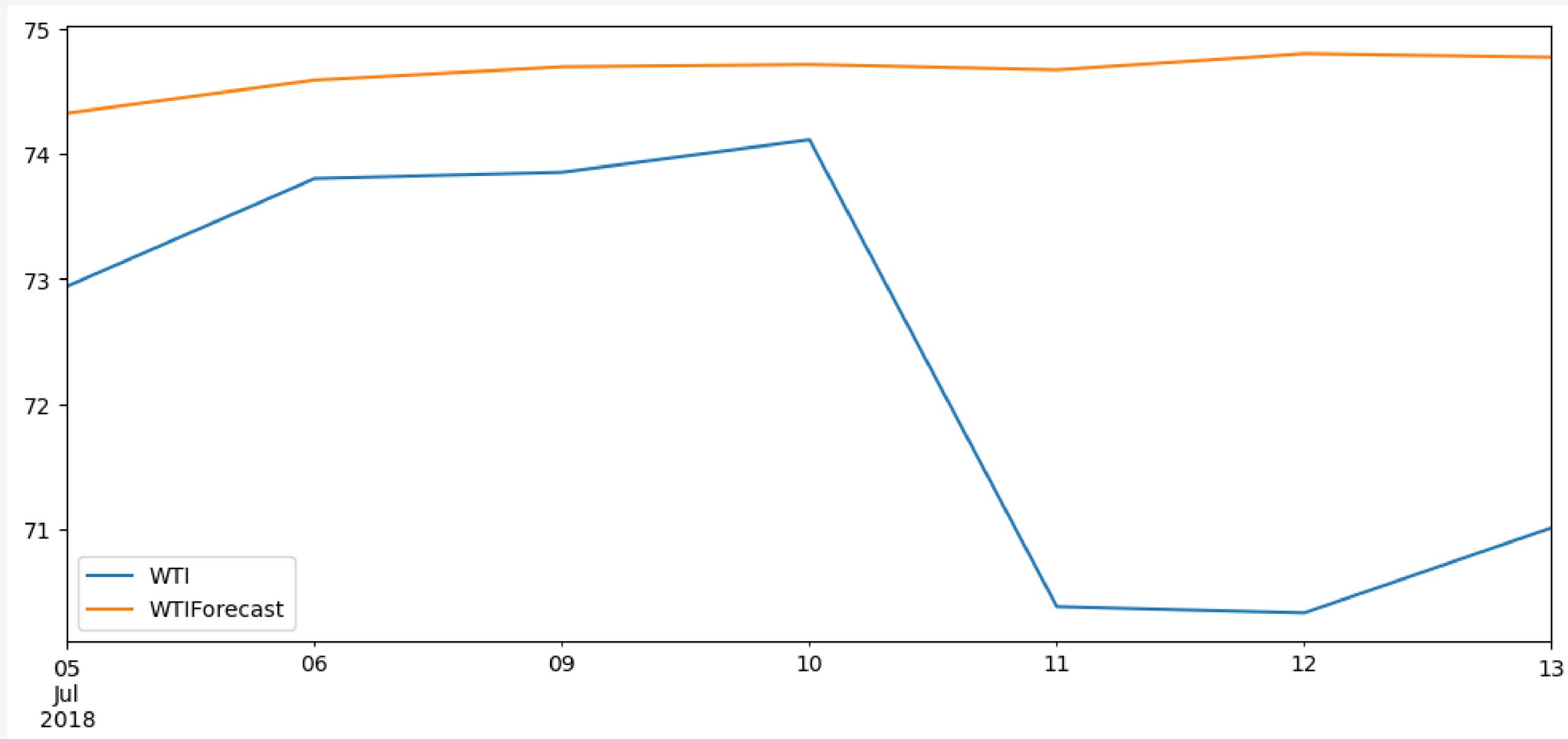
Mahalanobis > Financial > Z-Score Outlier Removal > No Outlier Removal

VAR

Developed the same model using selected features as well as all the features.

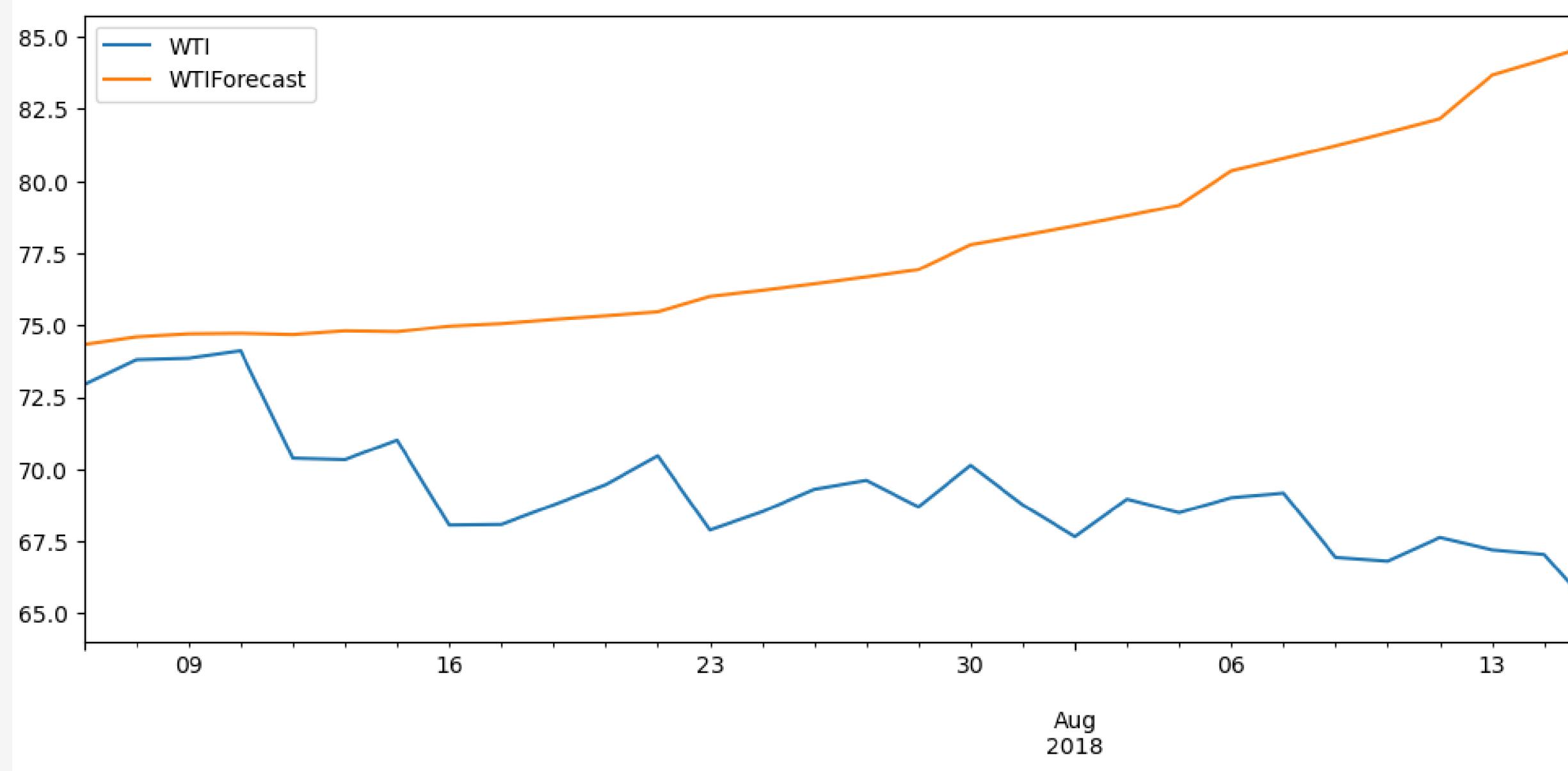
- 1 Checked stationarity of all the features
- 2 Made all the features stationary
- 3 Applied VAR of order 11 on obtained features
- 4 Forecasted the validation values
- 5 Compared the forecast values with original
- 6 Plotted the graphs and calculated the MAPE

VAR using all the features



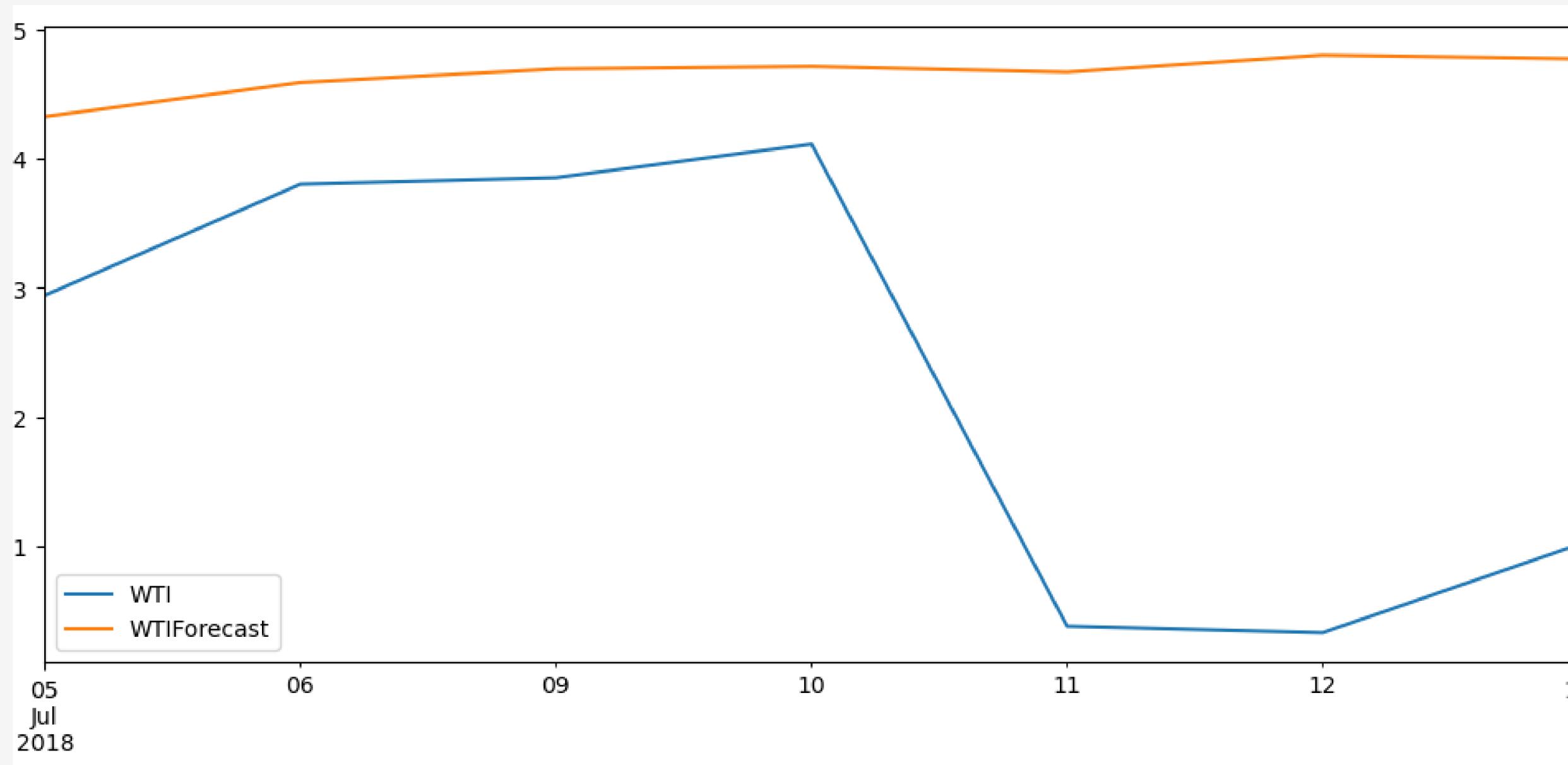
Forecasting for a week
MAPE=0.0323

VAR using all the features



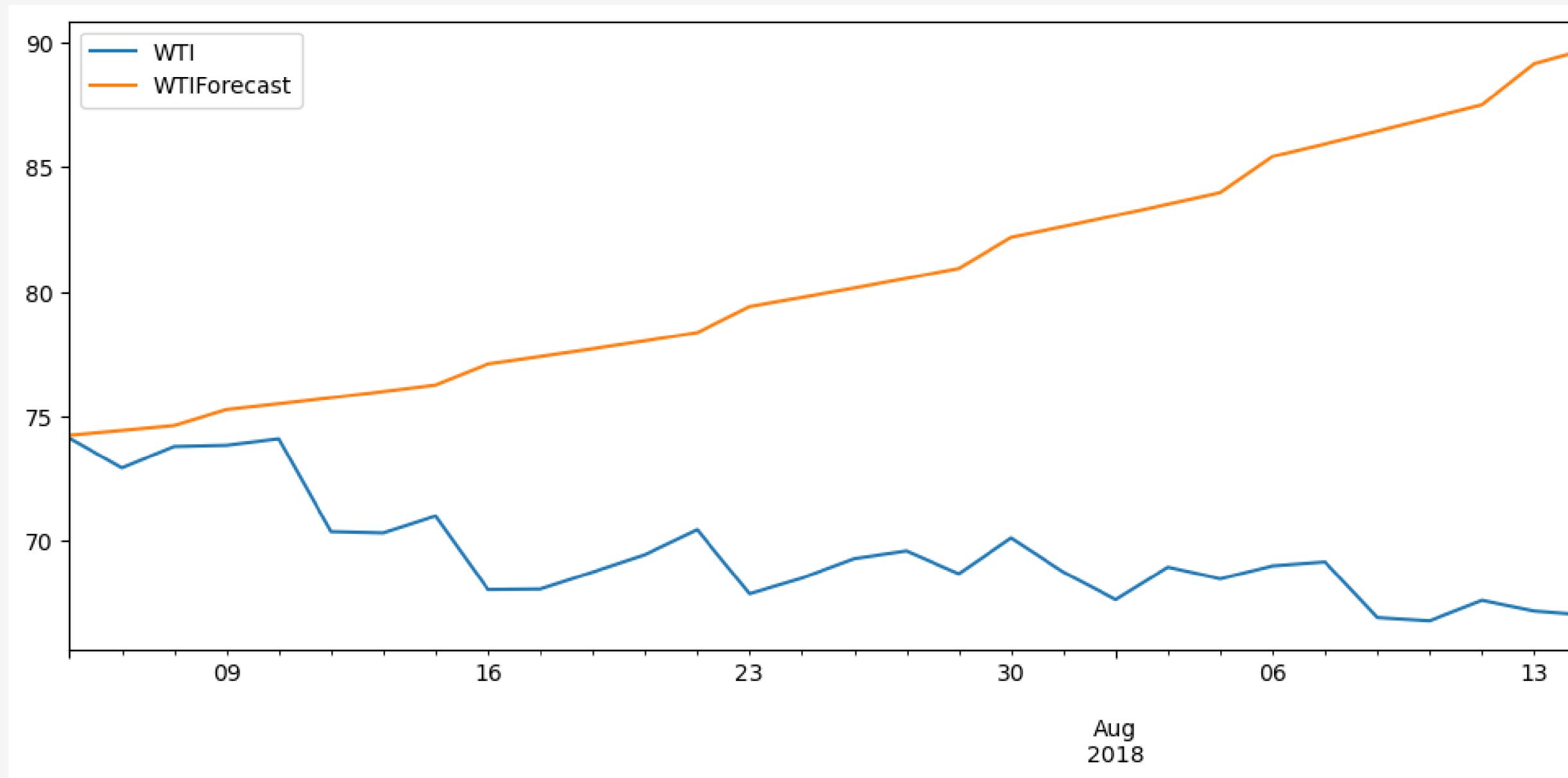
Forecasting for a month
MAPE=0.12

VAR using selected features



Forecasting for a week
MAPE=0.0326

VAR using selected features



Forecasting for a month
MAPE=0.16

Conclusion

0
Selecting features based on correlation is making the prediction less accurate.

0
VAR model is giving more accurate results in short term

0
LSTM-based Deep learning model is giving more accurate results in long term

