Multivariate Time Series Forecasting of Crude Palm Oil Price Using Data Science Approach

Markson Ofuoku
Data Decision Labs
November 30, 2021

Overview

Introduction

1. The primary purpose of this study is to investigate the correlation between monthly crude price and other vegetable oils (soybean oil, coconut oil, rapeseed oil, sunflower oil, etc.), crude oil, and exchange rates

2. The investigation also applies machine learning techniques to predict monthly CPO prices using other commodity prices

   - Classical methods: Triple Exponential Smoothing(TES) aka Holt-Winters technique

   - Deep learning approaches:

     1. Support Vector Regression (SVR)

     2. Multi-layer Perceptron (MLP)

Data Set

- We used nine time series data in this study. Monthly closing prices for all variables such as CPO price, sunflower oil price, olive oil price, rapeseed oil price, coconut oil price, peanut oil price, soybean oil price, West Texas Intermediate (WTI) crude oil spot price were taken from http: //www.indexmundi.com.

- The exchange rate from US dollars to Malaysian ringgit was taken from http://www.tradingeconomics.com. The frequency of data used is the monthly price from January 2002 to March 2021.

Visualization of Target Variable

- The target variable is monthly crude palm oil price from 2002 - 2021. This is shown on the plot on the right
- The plot below show the average crude palm oil price resampled over month, quarter, and year
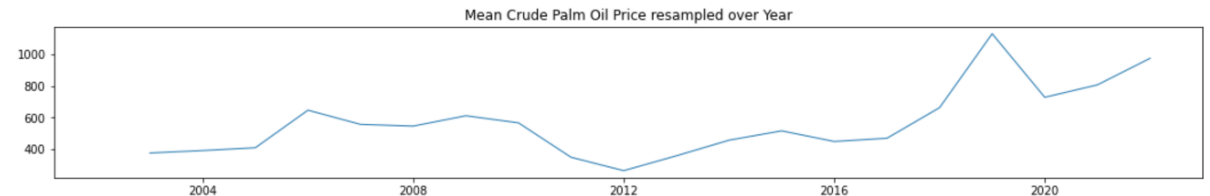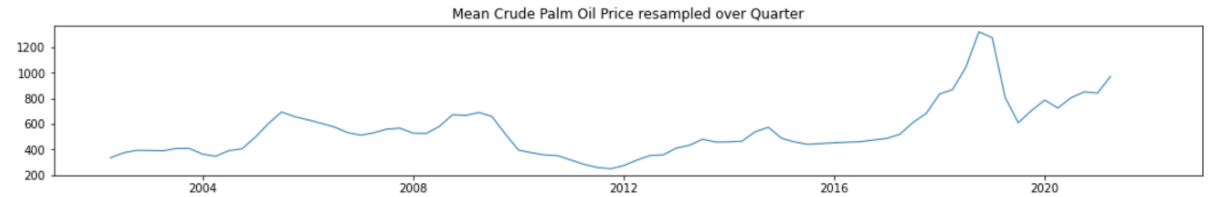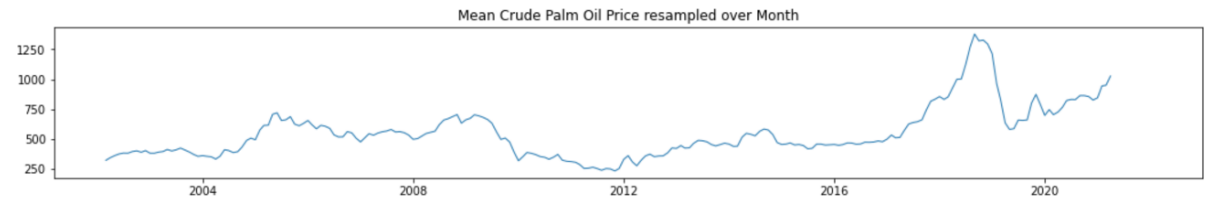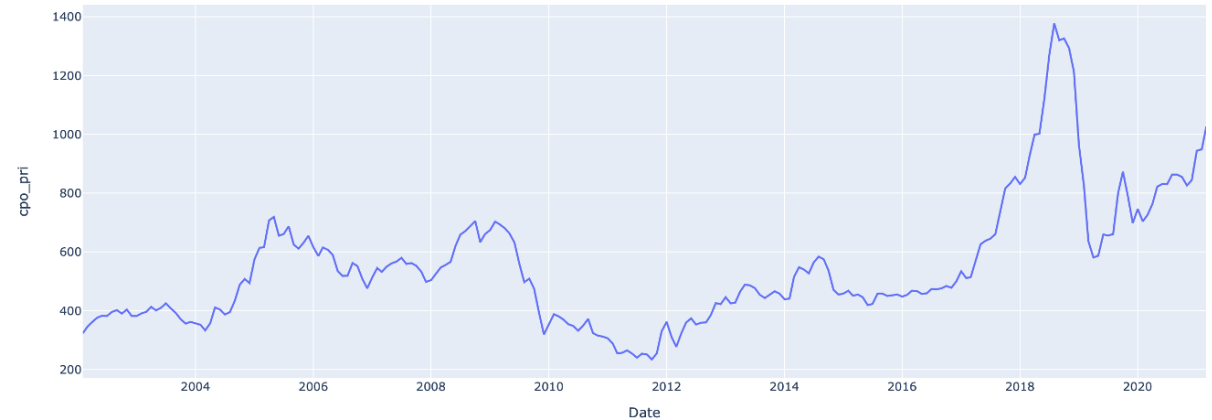
The plots of independent variables vs target variable

- The target variable, crude palm oil price and the independent variables: coconut oil price, soy bean oil price, peanut oil price, sunflower oil, rape seed oil price, west Texas intermediate oil spot price
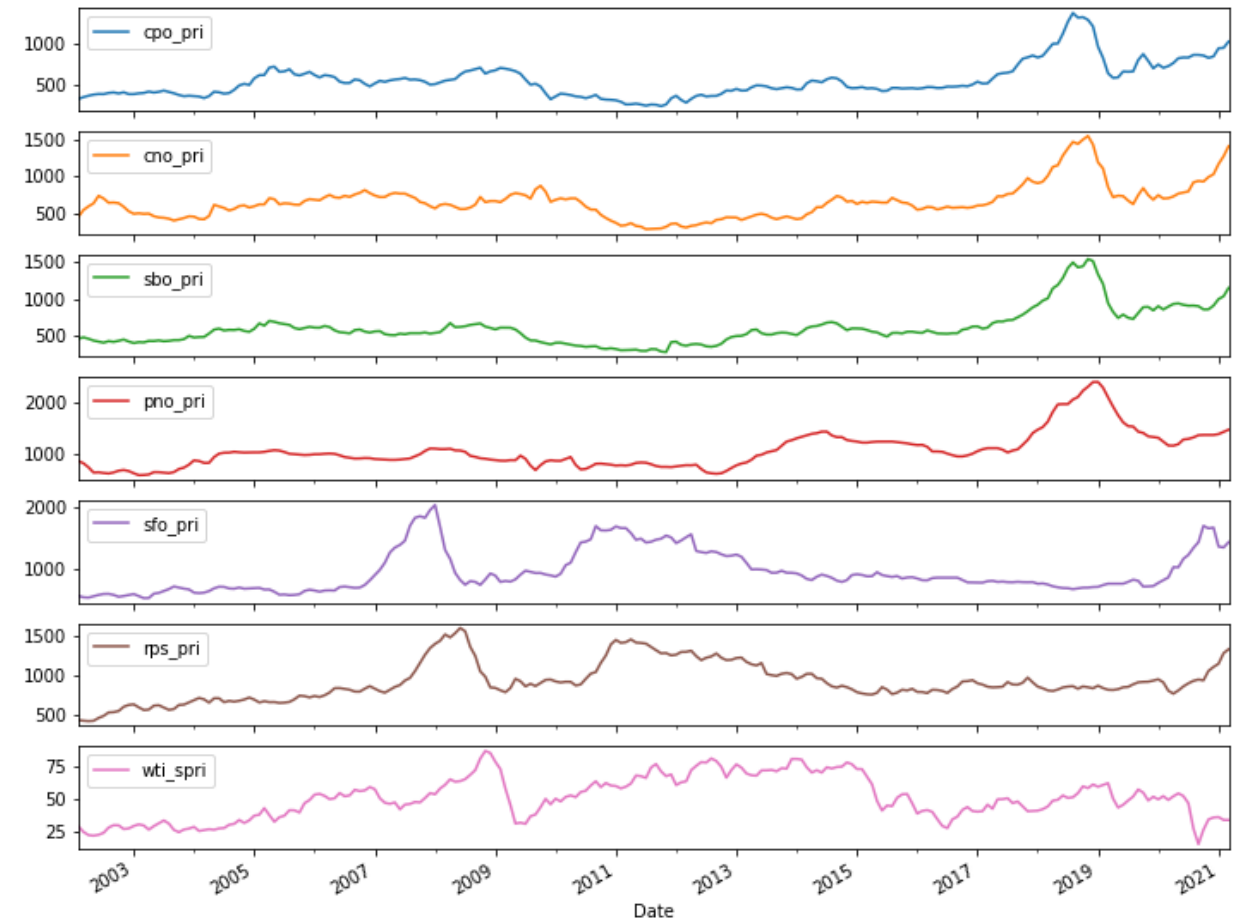

Figure: Target vs independent variables

Correlation Analysis

- Statistical correlation was used to determine the strength and association between crude palm oil and the other vegetables oils shown in the correlogram on the right

- The selected variables are: Soy bean price, coconut oil price, rapeseed oil price, sunflower oil price, Peanut oil price, and
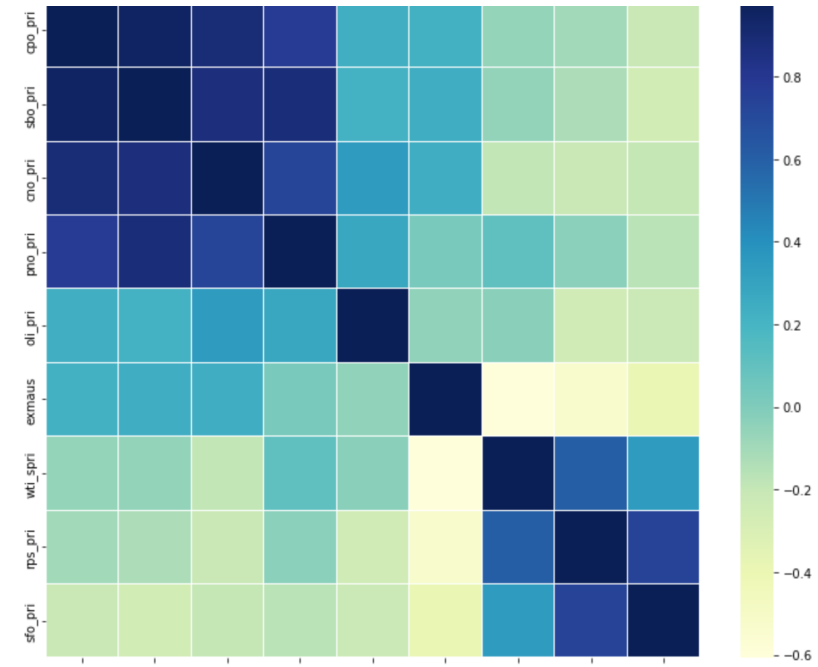

Figure: Correlogram of the data

	cno_pri	cpo_pri	exmaus	oli_pri	rps_pri	pno_pri	sbo_pri	sfo_pri	wti_spri
cno_pri	1.000000	0.891723	0.243321	0.349833	-0.208911	0.735385	0.868407	-0.198463	-0.190697
cpo_pri	0.891723	1.000000	0.226669	0.236296	-0.094080	0.776916	0.953051	-0.210459	-0.055629
exmaus	0.243321	0.226669	1.000000	-0.042444	-0.519051	0.023160	0.245384	-0.389682	-0.606916
oli_pri	0.349833	0.236296	-0.042444	1.000000	-0.246997	0.276659	0.220800	-0.220377	-0.027588
rps_pri	-0.208911	-0.094080	-0.519051	-0.246997	1.000000	-0.034386	-0.125801	0.741891	0.608432
pno_pri	0.735385	0.776916	0.023160	0.276659	-0.034386	1.000000	0.883398	-0.166050	0.110538
sbo_pri	0.868407	0.953051	0.245384	0.220800	-0.125801	0.883398	1.000000	-0.255978	-0.051980
sfo_pri	-0.198463	-0.210459	-0.389682	-0.220377	0.741891	-0.166050	-0.255978	1.000000	0.341807
wti_spri	-0.190697	-0.055629	-0.606916	-0.027588	0.608432	0.110538	-0.051980	0.341807	1.000000

Figure: Correlation table

Holt-Winters Method

- Plot between Actual and Predicted of the Holt-Winters method using the test data is shown on the right
- it did not produce reasonable forecast of the monthly crude palm oil price
- When we zoomed in over the period of 2017 through 20217 we can observe that predicted value is simply a extremely poor flat forecast
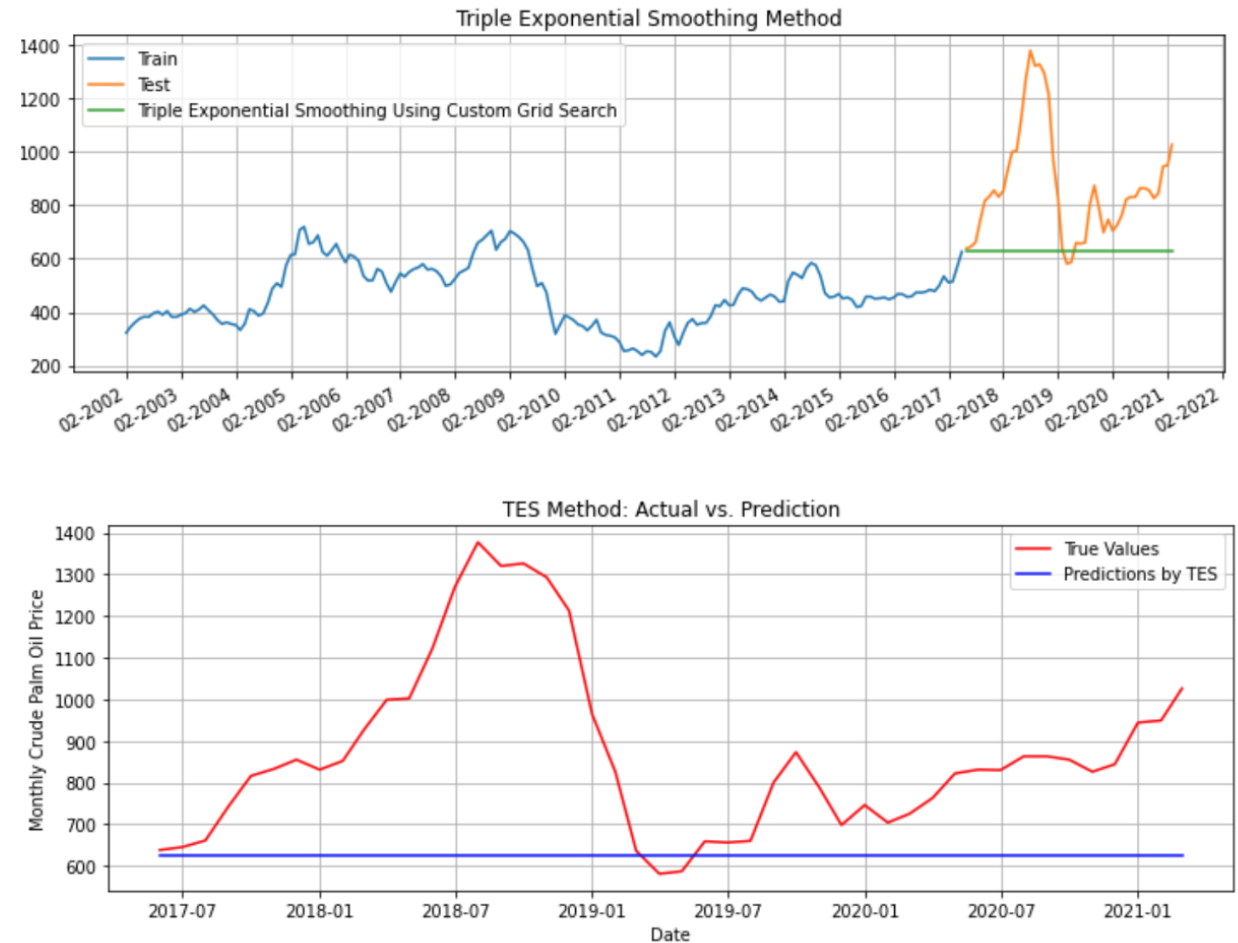

Figure: Holt-Winters

SVR Method

- Plot between Actual and Predicted of the Support Vector Regression using the test data is shown on the right
- The model predicted very well
- When we zoom into the lines, we can observe that predicted values mimic the actual values
- When the actual value changes direction, predicted value follows
- One typical snapshot of the prediction compared with the truth
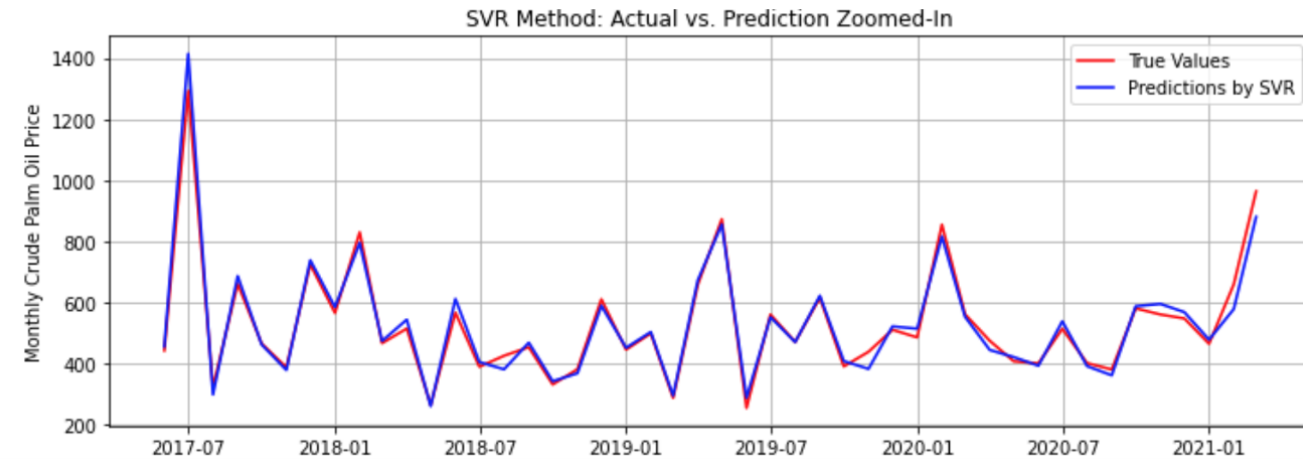
MLP Method

- Multilayer perceptron (MLP) is a basic example of a deep neural network.
- The MLP architecture consists of several hidden layers to capture the more complex relationships in the training dataset.
- The more layers hidden in the neural network, the deeper the network.
- Deep networks can learn more complex representations of inputs.
- The concept of hierarchical representation is that each layer learns a set of functions that describe the input and relays that information hierarchically through hidden layers
- First, the hidden layer near the input layer learns a simple set of features.
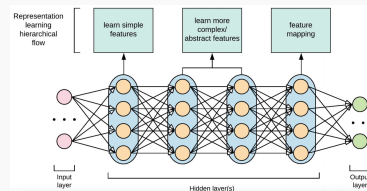
MLP Method Predictions

- The first plot on the right show the actual values versus predict for the MLP model
- The model predicted very well
- When we zoom into the lines (the plot below on the right), we can observe that predicted values mimic the actual values.
- It is observed from both plots that MLP follow the actual data closely.
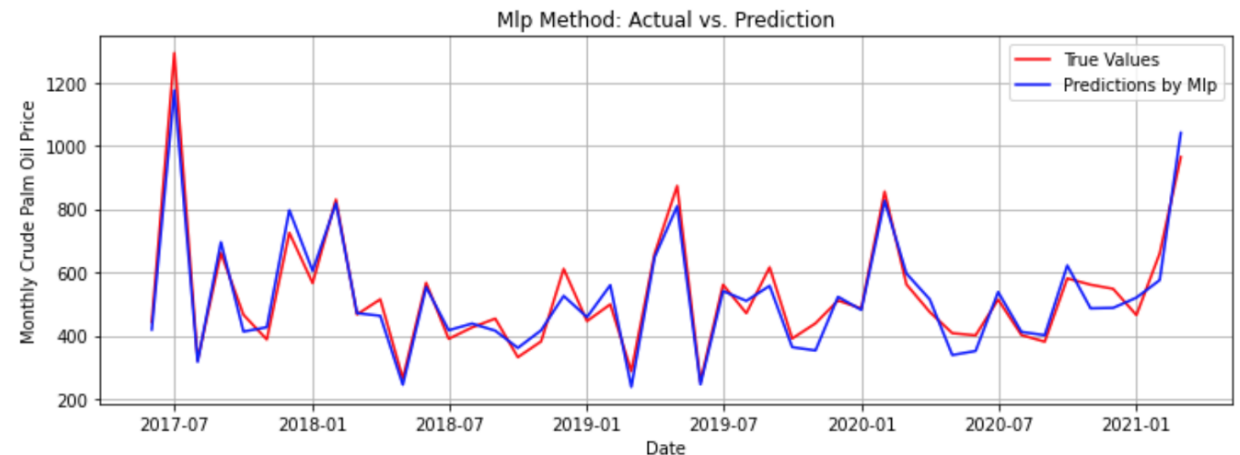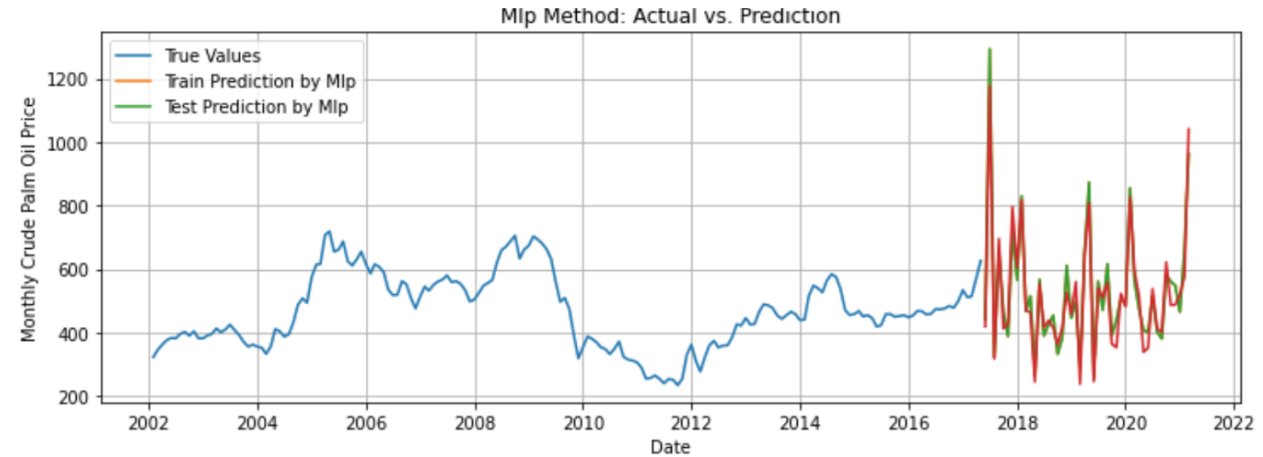- The figure top show the MLP predictions where the validated (test) is green line

Figure: Actual and predicted using MLP during testing and zoomed-in.

Results

- Plot between Actual and Predicted Trend of LSTM The above graph figure 3
- By comparing the two forecasting plots, we can see that the ARIMA model has predicted the closing prices very lower to the actual prices....HOLT-WiNTERS
- This large variation in prediction can be seen at the majority of the places across the plot.
- But in the case of the LSTM model, the same prediction of closing prices can be seen higher than the actual value.
- So we can conclude that, in the task of stock prediction, the LSTM model has outperformed the ARIMA model.
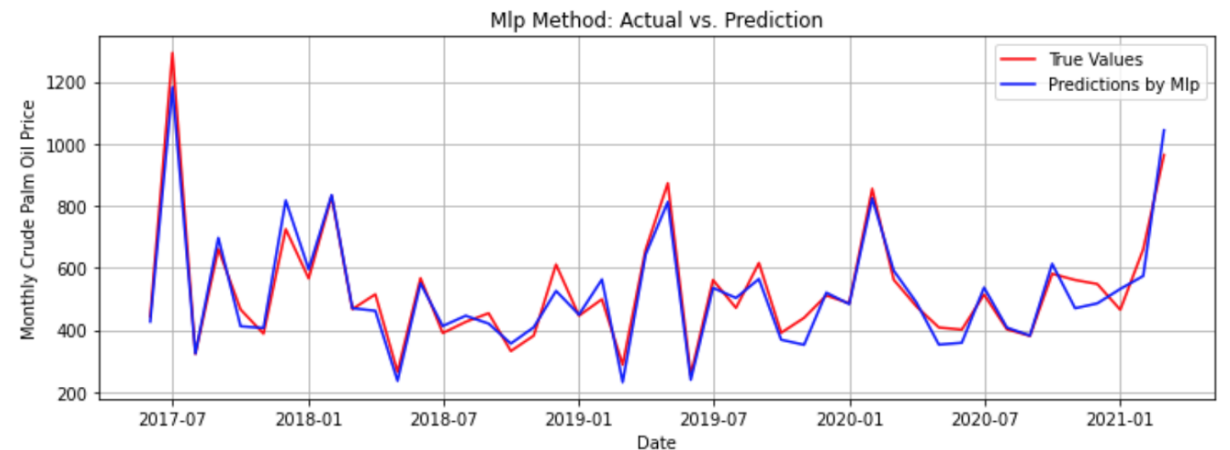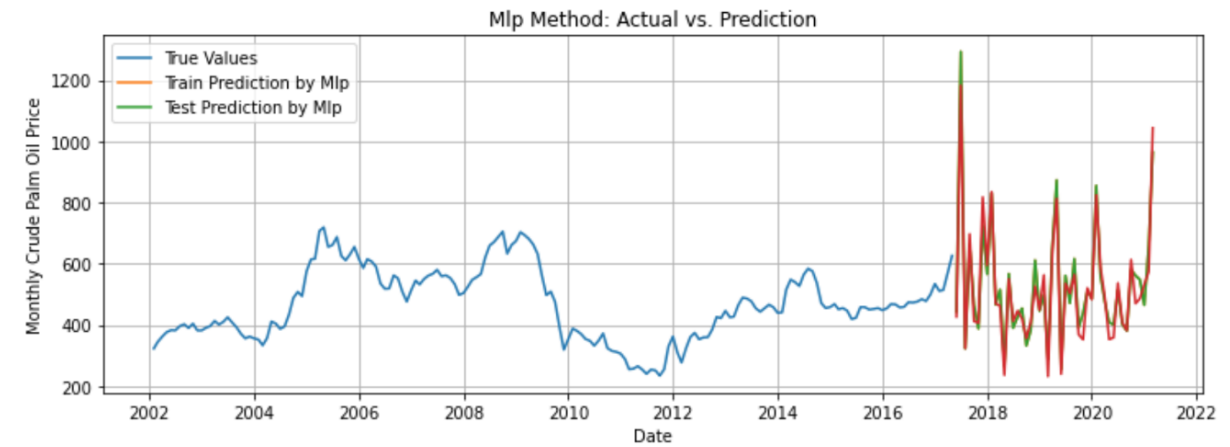
Support Vector Regression (SVR) Method

Support Vector Regression

The function itself has two types of parameters:

- Kernel-dependent parameters. For radial basis functions, these parameters are the epsilon and gamma set prior to training.

- And the so-called support vector determined during learning. Support vectors are just parameters for decision-making functions.
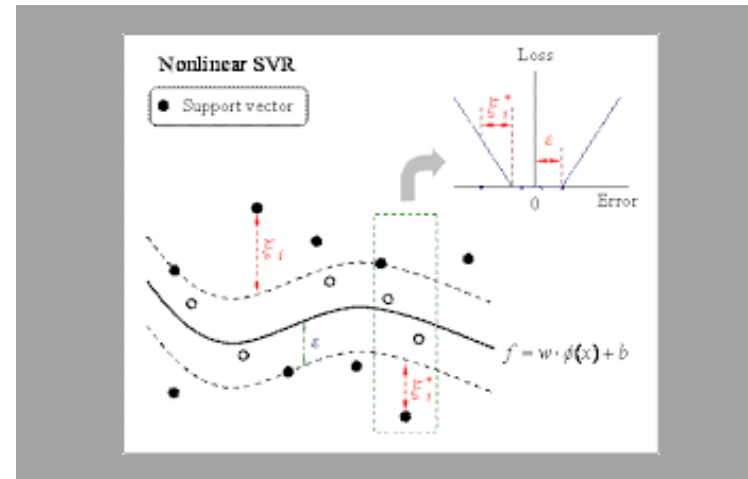

Figure: SVR Diagram

Results & Comparison

- The performance of models was evaluated using various statistical indicators

- The results for the three different forecasting techniques are summarized in the table on the right.

- Looking at the table, it is clear that SVR is the best model for predicting monthly crude palm oil price accuracy results

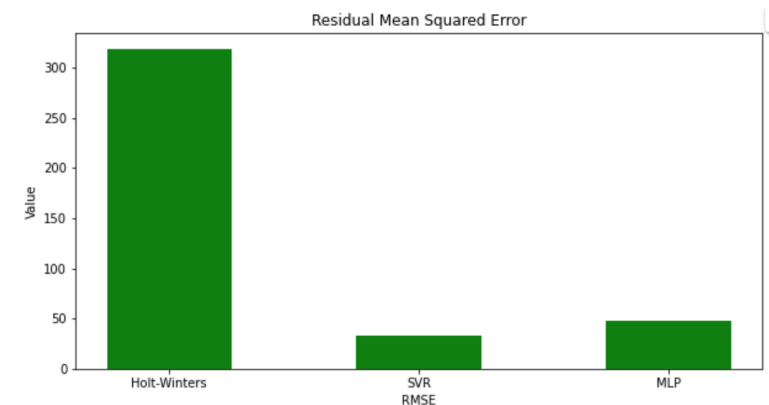| METHOD	MAE	MDA	MAPE	RMSE	MSE
Holt-Winters	248.1	0.0	25.11	38.55	101476.0
SVR	23.34	0.95	4.25	32.79	1075.64
MLP	37.7	0.95	7.13	47.69	2274.0

Figure: Performance metric table

Figure: RMSE bar plot

Results & Comparison

- The figure shows the forecasting results of three models on the test data

- As we cab see from the plot the SVR and MLP models predict very well. The SVR produce predicted values that agree with the observed values on average. The worst performing is the Holt-Winters
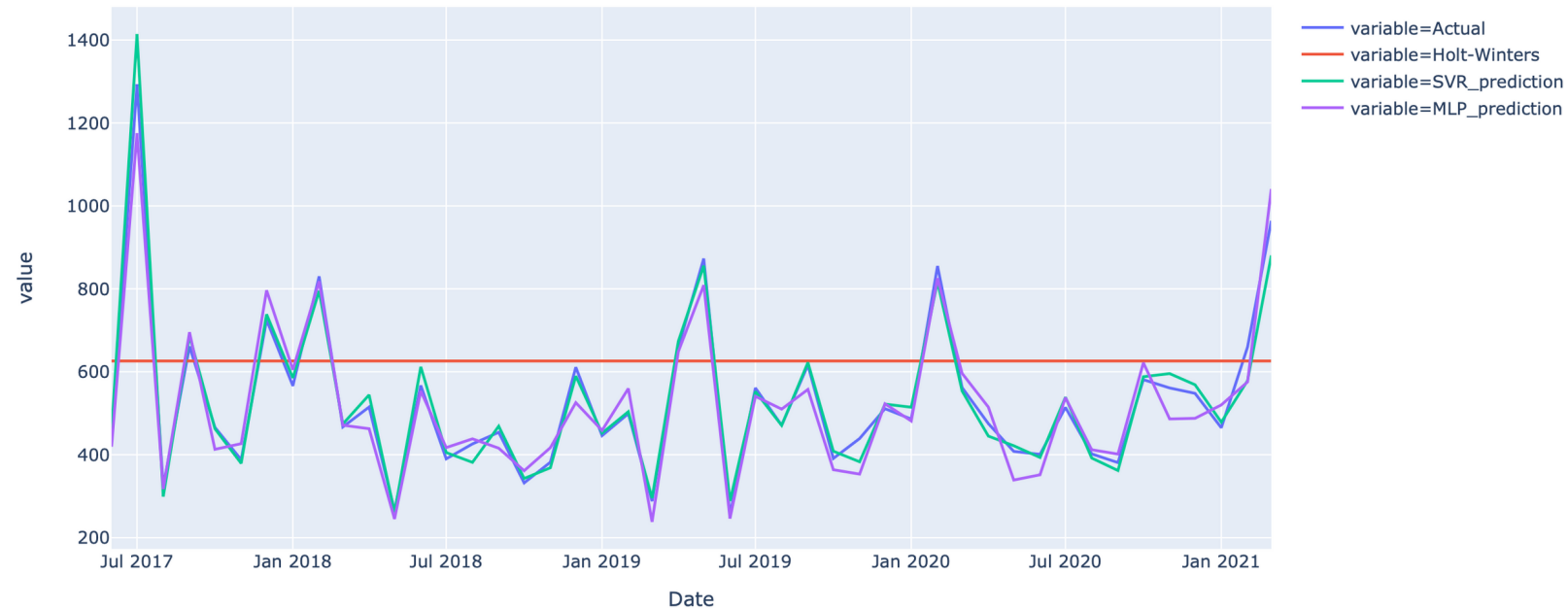
  Tlhe SVR is the best model at predicting monthly

Figure: Forecasting results of the three models

References

1. Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners - Ekaba Bisong

2. Multivariate Time Series Forecasting of Crude Palm Oil Price Using Machine Learning Techniques VL - 226 DO - 10.1088/1757-899X/226/1/012117 JO - IOP Conference Series: Materials Science and Engineering ER