**Summary of the models – Results**

1. **ARIMA**

|  |  |  |
| --- | --- | --- |
| **Error Metrics** | **Train** | **Test** |
| **RMSE** | 13.42 | 14.86 |
| **MAE** | 9.52 | 9.96 |

Obrázok, na ktorom je text, diagram, snímka obrazovky, vývoj

Automaticky generovaný popisFollowing plot is comparing actual and forecasted gold price for last 200 observations. While it is not possible to determine whether model is over or underpredicting actual values, mean error was computed, what resulted in positive value (0.4). Since optimal ARIMA configuration was computed (1,1,0), the model simply assume that predicted value is just 1-day lag of actual value.

Obrázok, na ktorom je diagram, text, rad, snímka obrazovky

Automaticky generovaný popisFollowing Figure illustrates distribution of errors. The mean is centered around zero, which suggest that model’s predictions are not systematically biased. The graph supports observation from previous section that mean error is close to 0. It is also possible to observe that most errors are 25 units off. There are few outliers, but it is considered usual when modeling financial time series data due to sudden spikes of price.

1. **Lasso Regression**

|  |  |  |  |
| --- | --- | --- | --- |
| Metric | Training | Validation | Testing |
| RMSE | 12.23 | 17.2 | 21.6 |
| MAE | 8.16 | 12.94 | 16.41 |
| ME | -7.8e-14 | -3.16 | -11.74 |

**Obrázok, na ktorom je text, vývoj, rad, snímka obrazovky

Automaticky generovaný popis**Following plot displays two curves – actual and predicted values. It can be observed that model is underpredicting the actual values. The model also seems to not react well to sudden spikes, underestimating actual price.

Histogram displayed in figure 11 in *LASSO regression Appendix* shows distribution of prediction errors. They are normally distributed around zero with left skew. As appears on the histogram, most errors fall within the range from -20 to 20, but presence of outliers can also be observed. The skewness on the left side supports earlier observation that model is underpredicting. The outliers as mentioned may represent price spikes or economic recession.

**Obrázok, na ktorom je vývoj, diagram, rad, text

Automaticky generovaný popis**

Following figure displays residuals over time. It is obvious, that errors fluctuate around zero, but tends to be negative at most predictions especially in second half of the plot. The magnitude of errors seems to increase towards the end of period, suggesting either more volatility in prices or decreasing model accuracy when it came across more recent data.

**Obrázok, na ktorom je text, vývoj, snímka obrazovky, purpurový

Automaticky generovaný popis**

1. **Random Forests**

|  |  |  |
| --- | --- | --- |
| Metric | Training | Testing |
| RMSE | 5.18 | 18.94 |
| MAE | 3.59 | 13.37 |

To see difference between actual and predicted value over time, with the help of R actual and predicted values were illustrated. The line plot clearly shows upward trend of gold price movement. Prediction line clearly follows increasing trend; however, it is obvious that the model is underpredicting, especially in the times of significant spike or drop in a gold price. It might be because the model is averaging noisy observations, which may make model less responsive to sharp spikes or drops.

Obrázok, na ktorom je text, snímka obrazovky, rad, vývoj

Automaticky generovaný popis

Histogram of residuals clearly shows that residuals are evenly distributed around 0, meaning that the model is not systematically biased. On the histogram it can also be observed that outliers are present, but this is considered common behavior in financial time-series data, which may occur in volatile, recession or uncertainty economic periods.

Obrázok, na ktorom je diagram, snímka obrazovky, vývoj, rad

Automaticky generovaný popis

Following illustration was obtained using *varImPlot()* function in R, which generates plot based on importance of variables within the trees. In other words the plot shows predictors that were most used in decision nodes in order to increase predictive performance e.g. lower error rate. This function relies on correlation between variables, so lagged predictors of gold were used most in splits, following with stock indexes, USD/INR exchange rate and silver. Other predictors seem to have negligible predictive performance according to this function. This provides clear reflection on which predictors were most important when conducting predictive analysis.

Obrázok, na ktorom je text, snímka obrazovky, písmo, číslo

Automaticky generovaný popis

1. **XGBoost**

|  |  |  |
| --- | --- | --- |
| Error Metric | Training | Testing |
| RMSE | 4.19 | 16.68 |
| MAE | 3.26 | 11.53 |

To clearly see the difference, I used last 200 observations. It is obvious, that the model tends to underpredict actual values most of the time. While model seems to capture overall upward trend, it seems like it does not predict well when it comes to sudden spikes in gold price. The model is obviously also not very accurate when it comes to spikes downward, where it adjusts slowly. One of possible reason this is occurring, might be, that gold itself require more than just one- or two-day lags. Another logical way to improve model’s performance could be adding volatility index as one of the predictors, where model would be able to recognize more and less volatile periods and adjusts predictions accordingly.

Obrázok, na ktorom je text, vývoj, rad, diagram

Automaticky generovaný popis

Following histogram displays distribution of residuals. In general, they are centered around zero, which means that model is not making systematic errors, and that model is considered unbiased. However, histogram also displays that distribution of residuals is right-skewed, which supports underestimation fact drawn from previous plot. It is also possible to see couple of outliers on both sides, but that could be the reason of sudden spikes in gold price, where model failed to predict spike, which is expected.

Obrázok, na ktorom je diagram, vývoj, snímka obrazovky, rad

Automaticky generovaný popis