**Report:**

1. LSTM Model to predict stock ‘Close’ prices at Recession (2008-2009) and during the Covid pandemic (2020-2021) using the features: [Open, High, Low, and Volume]:

NOTE: All the experiments are performed at sequence length= 30; batch size=32

At **Recession**:

Shape of data: **(2579, 7)**

Shape of Train\_x: **(2033, 30, 4)**

Shape of Test\_x: **(486, 30, 4)**

RMSE at Sequence length = 30, Epochs=10 and batch size = 32 is **11.25**

A graph with red and blue lines

Description automatically generated

At **Covid Pandemic**:

Shape of data: **(5472,7)**

Shape of Train\_x: **(4347, 30, 4)**

Shape of Test\_x: **(1065, 30, 4)**

RMSE at Sequence length = 30, Epochs=10 and batch size = 32 is **17.19**

A graph showing a line graph

Description automatically generated with medium confidence

1. Time Series are combination of mainly three decomposition components. Decomposition helps us to better analyze the data. So, I’m considering this decomposition components as features
2. **Trend**: This is overall motion of the series
3. **Seasonality**: Any regular seasonal pattern in the series
4. **Residuals:**  This is left over after we consider the trend and seasonality (Statistical Noise).

And along with these Decomposition Components, I have also included technical operators like ‘Simple Moving Average’ and ‘Exponential Moving Average’, ‘Relative Strength Index’ and a Economic Data i.e. Interest rates in United States from 1999 to 2009.

Features: [trend, seasonality, residual, Simple Moving Average, Exponential Moving Average, Relative Strength Index]

* Plotted a correlation matrix for data from 1999 to 2009 (Recession Period) using all the features.

A screenshot of a graph

Description automatically generated

* Highly Correlated features are: [Open, High, Close, Trend, SMA, EMA]
* Not Highly Correlated Features are: [Interest rates, Seasonal, Residual, RSI]

1. A LSTM model to predict stock ‘Close’ prices at Recession (2008-2009) using All these Highly Correlated features,

|  |  |
| --- | --- |
| **Epochs:** | **RMSE:** |
| 10 | 13.05 |
| 50 | 9.75 |
| 75 | 9.55 |
| 100 | 10.53 |
| 150 | 9.37 |
| 200 | 10.03 |

1. A LSTM model to predict stock ‘Close’ prices at Recession (2008-2009) using All these Not Highly Correlated features,

|  |  |
| --- | --- |
| **Epochs:** | **RMSE:** |
| 10 | 115.06 |
| 50 | 123.15 |
| 75 | 130.81 |
| 100 | 129.52 |
| 150 | 129.34 |
| 200 | 130.78 |

**Key findings:**

* The LSTM trained on highly correlated features vastly outperformed the one on non-highly correlated features in all tested epochs, as reflected in RMSE differences.
* The model's performance deteriorates consistently with increasing epochs, suggesting these features may not contain the required information for accurate prediction.
* The RMSE stabilizes between 9-10 for the correlated feature model, with the most optimal performance observed at 150 epochs.
* Correlation matrix for data from 1999 to 2021 (Covid Period) using all the features.

A screenshot of a graph

Description automatically generated

1. A LSTM model to predict stock ‘Close’ prices at Covid period using All these Highly Correlated features,

|  |  |
| --- | --- |
| **Epochs:** | **RMSE:** |
| 10 | 16.58 |
| 50 | 10.48 |
| 75 | 19.97 |
| 100 | 17.14 |
| 150 | 17.11 |
| 200 | 17.76 |

1. A LSTM model to predict stock ‘Close’ prices at Covid period using All these Not Correlated features,

|  |  |
| --- | --- |
| **Epochs:** | **RMSE:** |
| 10 | 161.93 |
| 50 | 145.10 |
| 75 | 143.10 |
| 100 | 146.86 |
| 150 | 158.41 |
| 200 | 146.60 |

**Key Findings**:

* For the model with highly correlated features, training for around 50 epochs seems to provide the best results.
* The LSTM model trained with highly correlated features significantly outperforms the model trained with non-highly correlated features across all epochs.
* The RMSE values for the model trained with correlated features are much lower than those for the non-correlated features, indicating better predictions.

🡺 The model trained on highly correlated features has consistently outperformed the model trained on non-highly correlated features across all epochs in terms of RMSE. This suggests that the correlated features carry significant predictive power for the target variable. So I’m looking forward take these highly correlated features for both recession and Covid period.