# Studying Impact of different parameters on SAG Mill and using VAR and LSTM models for time series forecasting the power consumption of SAG mill

- Project by **Priyanshu Jha | 20JE0725**
- Under the guidance of **Dr Siddhartha Agarwal**
- **SRIP INTERN 2023 ( 21/05/2023-21/07/2023)**

# APPLICATIONS OF TIME SERIES IN MINING OPERATIONS

- **Predictive Maintenance**: Use time series analysis to monitor equipment health and schedule maintenance before equipment failure occurs, minimizing downtime and repair costs.
- **Environmental Monitoring**: Use time series analysis to monitor environmental variables and ensure compliance with regulations.
- **Energy Consumption Optimization**: Use time series analysis to optimize energy consumption and reduce costs.
- **Safety Analysis**: Analyze safety monitoring data to predict hazards and enhance mine safety.

# WHAT IS TIME SERIES?

- Time series data is a sequence of observations collected from a process with equally spaced periods of time
- Time Series Analysis is used in different fields for time-based predictions – like Weather Forecasting models, Stock market predictions, Signal processing
- There is only one assumption in time series analysis which is **“Stationarity”**

# STATIONARITY

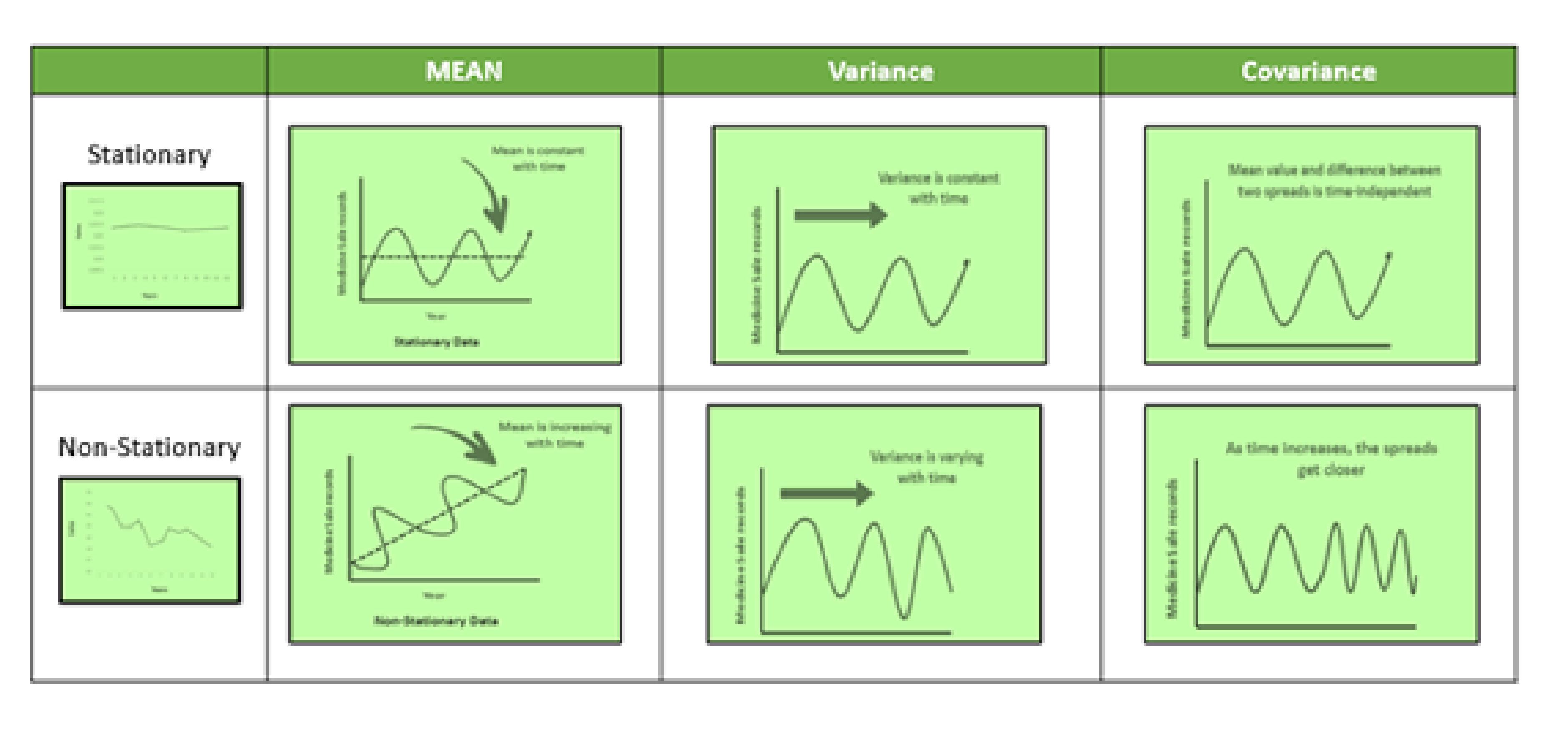
- Stationarity is a key concept in time series analysis.
- A stationary time series is one where the:

  1. mean,

  2. variance, and

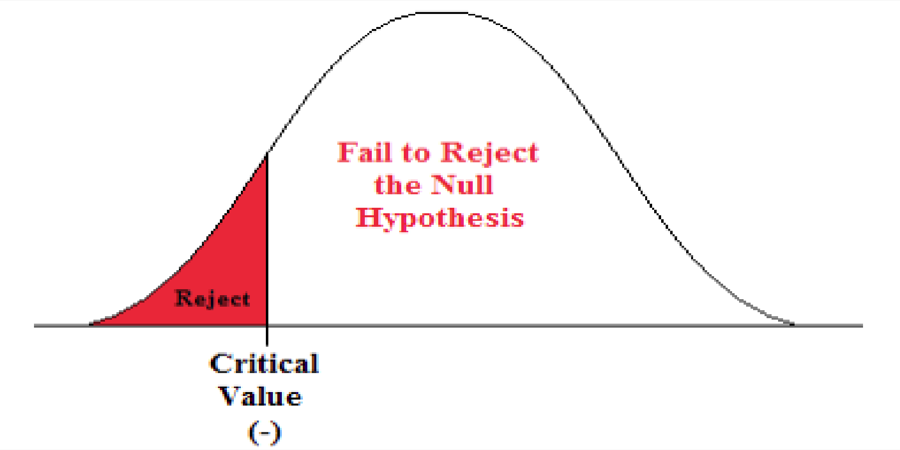
  3. autocorrelation are constant over time.
- A non-stationary time series can be difficult to analyze and can be converted to a stationary time series using differencing

# STATIONARY VS NON STATIONARY

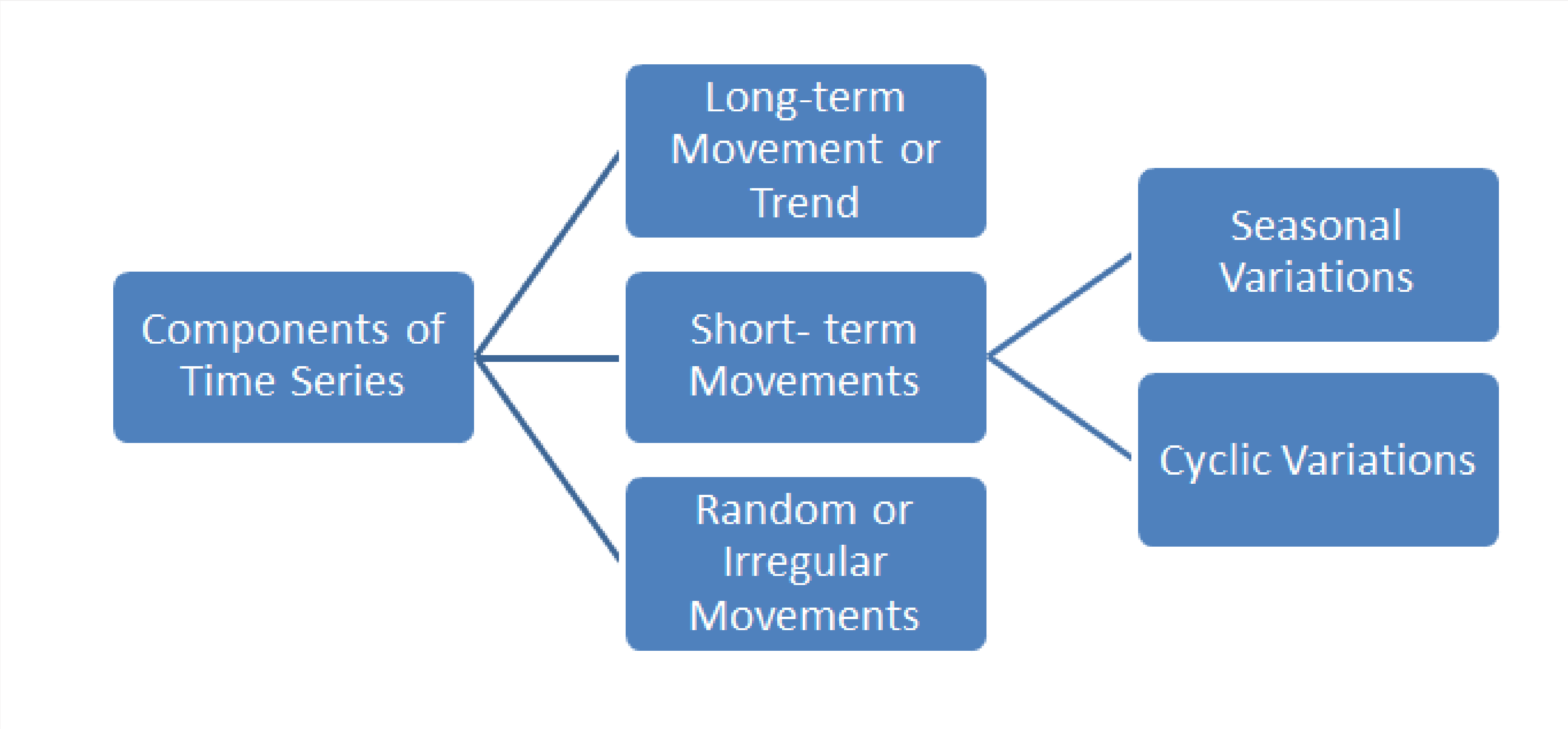


# TEST OF STATIONARITY: AUGMENTED DICKEY FULLER TEST

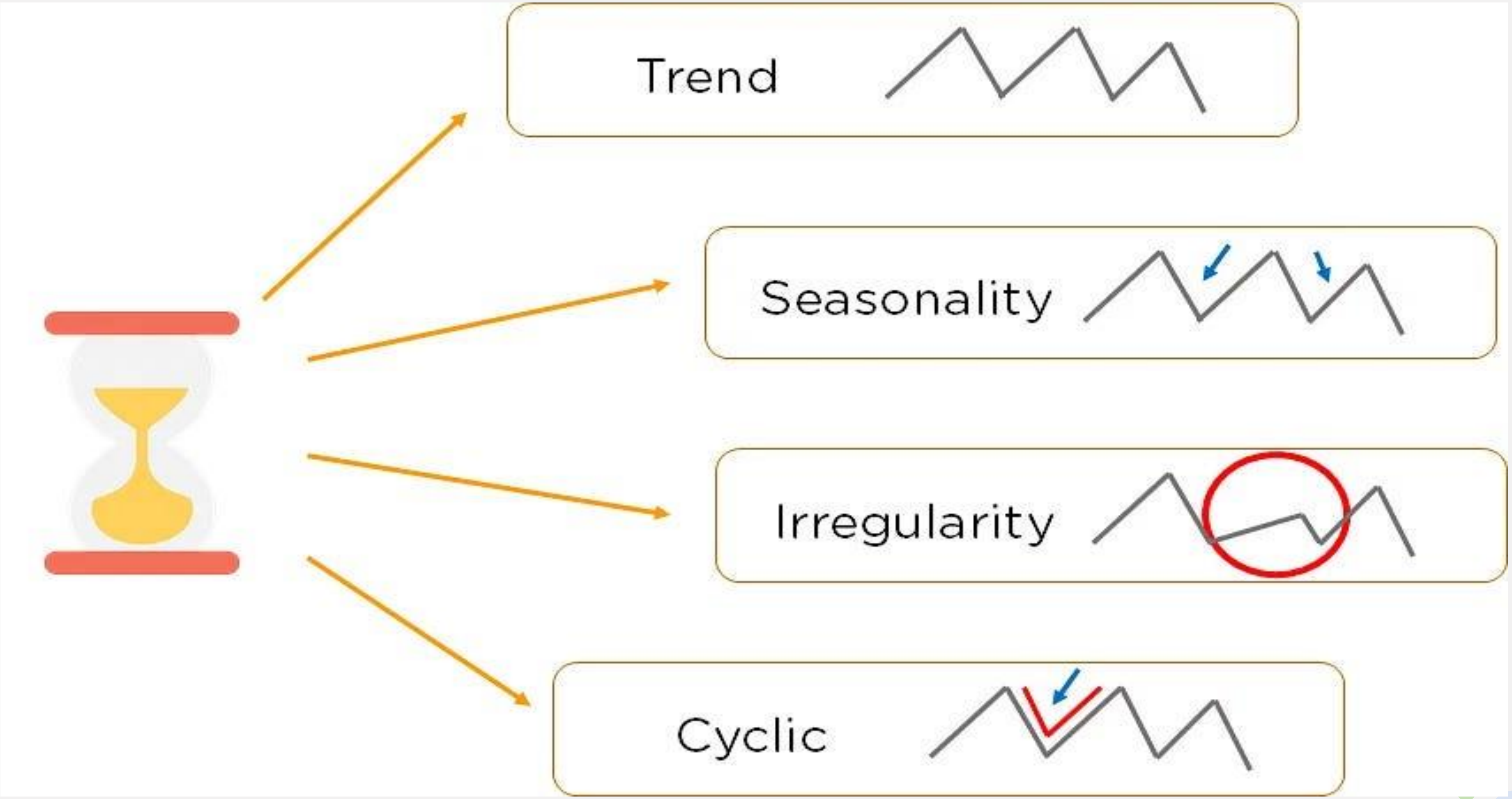
- **Augmented Dickey Fuller Test** gives an idea regarding stationarity of time series
- Here “null hypothesis (H0) is considered as data is non-stationary” and “alternative hypothesis (H1) is considered as data is stationary ”
- If Test statistic < Critical Value and p-value < 0.05 – **Reject Null Hypothesis(HO)**



# Components of Time Series Analysis



# Components of Time Series Analysis



# TYPES OF TIME SERIES MODELS

## MOVING AVERAGE

- used to analyze and predict data points based on the average of past observations
- type of linear regression model that utilizes a series of past error terms (residuals) to make predictions
- Error is the difference between the actual observed value and the predicted value at a particular time point.

$$y_t = c + \varepsilon_t + \theta_1 * \varepsilon_{(t-1)} + \theta_2 * \varepsilon_{(t-2)} + \ldots + \theta_q * \varepsilon_{(t-q)}$$

# EXPONENTIAL SMOOTHING

- Exponential smoothing is a time series method for forecasting univariate time series data
- Exponential Smoothing time series method works by assigning exponentially decreasing weights for past observations
- It is called so because the weight assigned to each demand observation is exponentially decreased.
- The three types of exponential smoothing are:

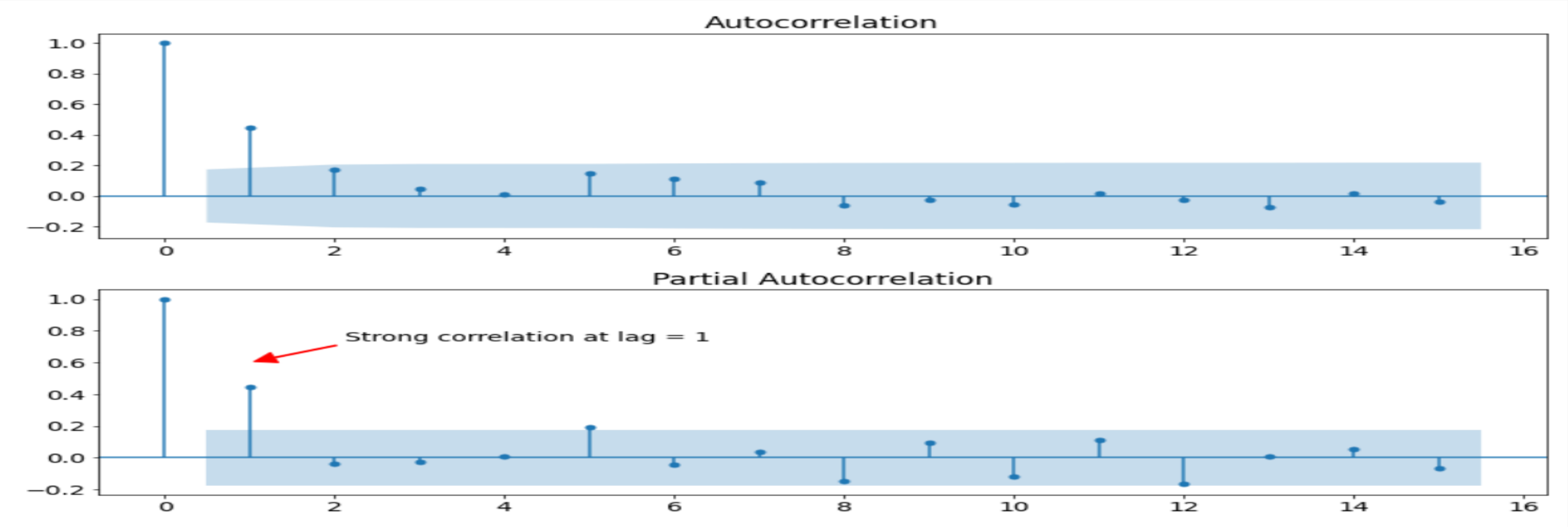
  Simple or single, Double and Triple exponential smoothing

# ARIMA MODEL

- ARIMA**(Auto Regressive Integrated Moving Average)** is a powerful tool for forecasting time series data.
- ARIMA has three components: **Autoregression (AR), Integrated (I), and Moving Average (MA).**
- (AR) refers to the use of past values of the time series to predict future values. (I) refers to the differencing of the time series to make it stationary. (MA) refers to the use of past errors to predict future values.
- The order of ARIMA is denoted by (p, d, q), where **p is the order of AR, d is the degree of differencing, and q is the order of MA**.

# ACF AND PACF PLOT

- In an ACF and PACF plot, each bar represents the correlation coefficient and partial correlation coefficient respectively for a specific lag.
- ACF plot helps to identify the order of q(MA) and PACF plot helps to identify the order of p(AR)

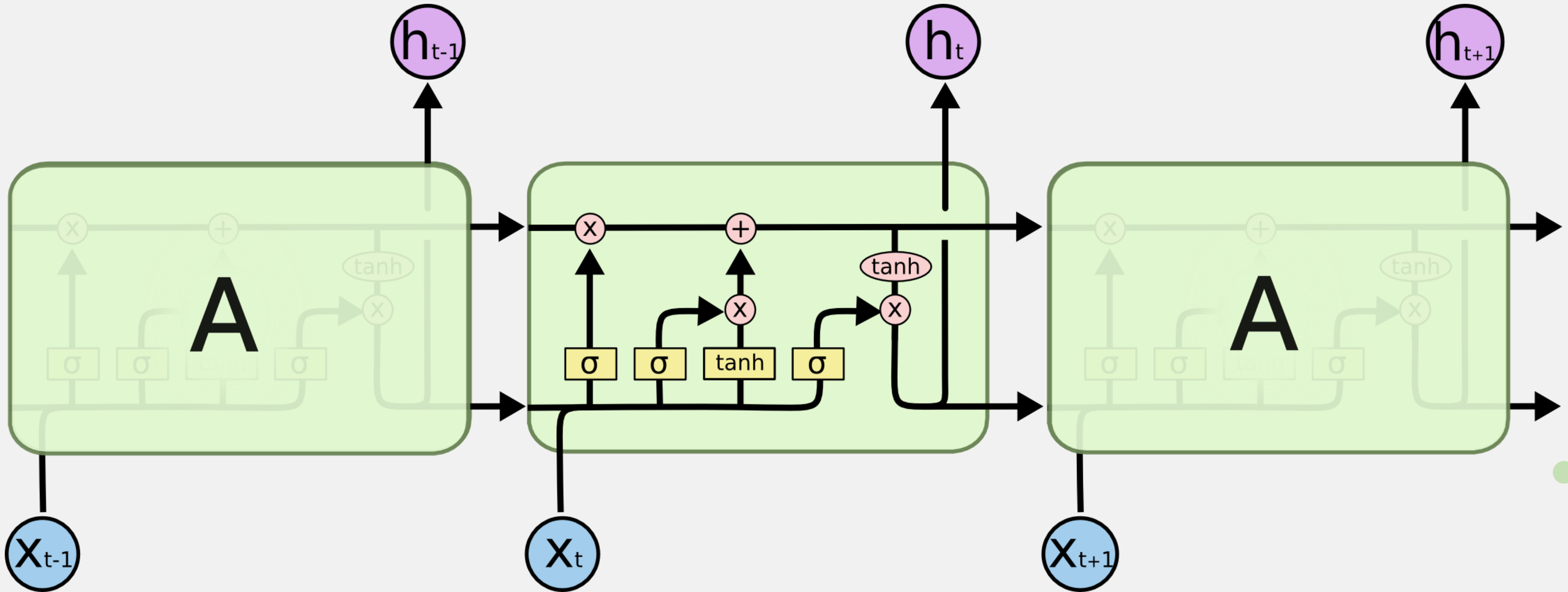


# AI BASED MODEL:
# LSTM (LONG SHORT TERM MEMORY)

- LSTM is a type of Recurrent Neural Network(RNN) that is designed to overcome the vanishing gradient problem.
- It is very difficult to train Simple RNN to retain information over many time steps and due to vanishing gradient problem it is not able to retain long term dependencies
- Here when LSTM comes into picture as it is able to retain long term dependencies unlike the Simple RNN

# LSTM NETWORK ARCHITECTURE

- LSTM has an internal memory cell, input gate, output gate and forget gate.
- These gates help in deciding what information to keep and what to discard. The LSTM architecture makes it particularly useful for tasks that involve long-term dependencies, such as speech recognition and natural language processing.
- LSTM works well when we have large amount of data as it type of neural network



# VAR (VECTOR AUTOREGRESSION)

- Vector Autoregression (VAR) is a statistical method used to model the relationship between multiple time series variables. .
- In VAR each variable is modeled as linear combination of past values of all the variables including itself. So the equation can be formulated as below,.

$$y_{t,i} = \alpha_i + \sum_{k=1}^{p} \sum_{j=1}^{n} \phi_{i,j} * y_{t-k,j} + \varepsilon_{t,i}$$

Where,

“i” ranges from 1 to number of variables used to prediction. Left hand side variable is the one which we will predict for time step “t” and “p” represents the lag value used and “n” represent the number of variables used for prediction of a particular variable and “epsilon” represent error term for forecast at time step t for variable i.

# GRANGER'S CAUSALITY TEST

- **Granger's Causality Test** gives an idea regarding how a particular variables effects the forecast of other variables.
- null hypothesis (H0) of this test is a statement that assumes there is no causal relationship between the two time series being analyzed
- if the null hypothesis is not rejected, it suggests that there is no evidence of Granger causality between the two time series
- For rejecting null hypothesis pvalue< critical value(which is usually 0.05 or 0.10)

# AKAIKE INFORMATION CRITERION(AIC)

- AIC lets you test how well your model fits the data set without over-fitting it.
- The AIC score rewards models that achieve a high goodness-of-fit score and penalizes them if they become overly complex.
- **Model with least AIC is the best model**

Number of model parameters

$$AIC = 2k - 2\ln(\mathcal{L})$$

$\mathcal{L} = \mathcal{L}(\hat{\theta})$ = maximum value of the likelihood function of the model

# PROBLEM STATEMENT OF MY INTERNSHIP

- Milling is the process of breaking up the material into fine particles for size reduction.
- **Semi-Autogenous Grinding mill (SAG mill)** is used in initial stages for grinding large size for material to smaller size and this material is sent to ball mill for further precise reduction
- According to the studies from FLSmidth organization they observed that **correct RPM according to the density of feed can save up to 6% of power consumption** and increase the life of liners present inside due to reducing the critical impacts.
- SAG mill working depends on previous feed input this is considered as Time-series data so LSTM can be used to optimize the power consumption and reduce the costs

# SEMI-AUTOGENOUS GRINDING(SAG)

- Semi-autogenous grinding mills known as SAG mills.

- The mills are
  used for primary stage
  grinding

- The purpose of this SAG mills is reduce the size of ore particles.

# DATA INPUTS

- Features used for forecasting SAG mill power consumption:
  - **Bearing Pressure**: Force exerted by grinding media on mill lining
  - **Density of steel balls used**: Mass of balls per unit volume
  - **Noise**: Sound generated during grinding
  - **Recycle Amount**: Proportion of output sent back for regrinding
  - **Feed Input**: Material fed into the mill for grinding
  - **RPM**: Rotational speed of the mill
- **Data output is energy consumption (kWh)**
- Data size is **18,651 samples** of one minute intervals were given. In that first **10,000** data points are used for training**, next 5000** data points are used for validation and **remaining** are used for final testing of the model.

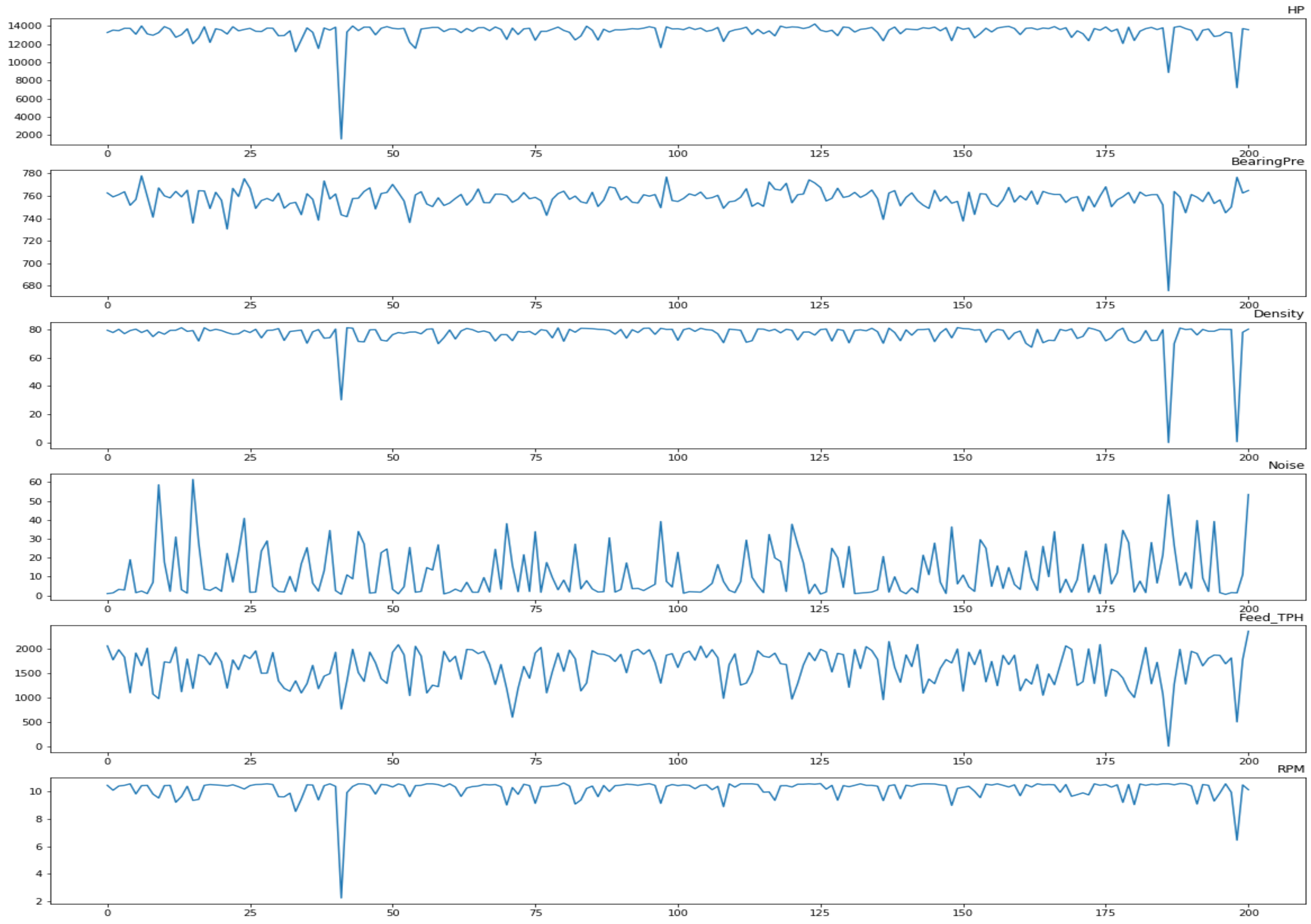
# METHODOLOGY

- Forecasting can divided into different categories:
  - **Short Term Forecasting** : Forecasting based on minute intervals
  - **Medium Term Forecasting** : Forecasting based on Hourly intervals or 8 hr. intervals.
  - **Long Term Forecasting** : Forecasting based on Daily intervals or weekly intervals.

# METHODOLOGY

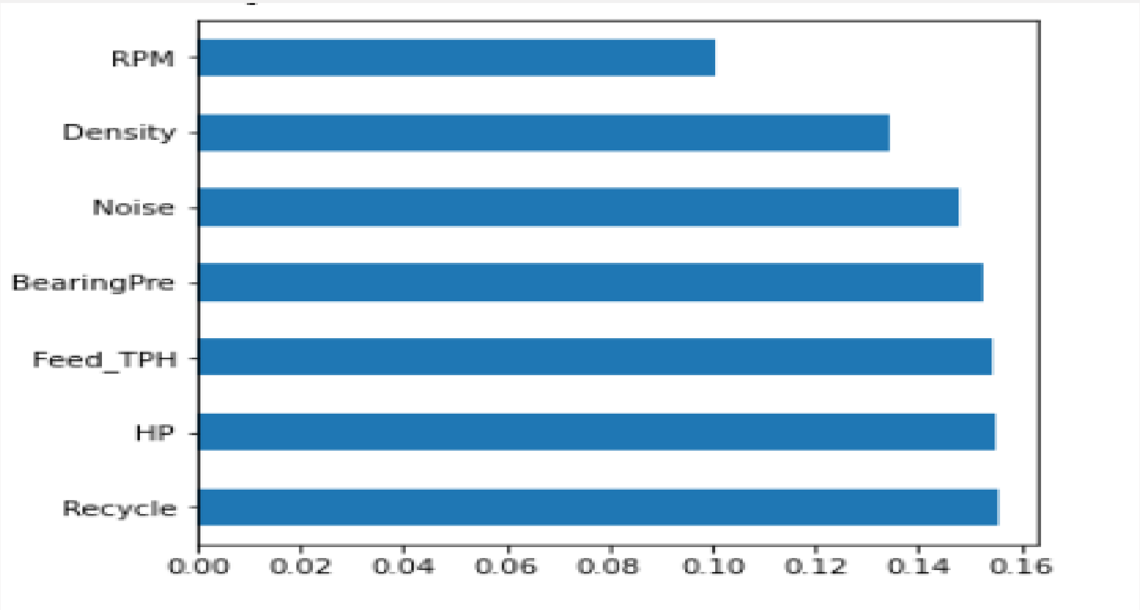
- Models used for forecasting:
    - **Long Short Term Memory (LSTM) Model**
        - LSTM is type of recurrent neural network which have internal architecture having main feature to handle input sequence and learn temporal dependency features.
    - **Vector Auto Regression (VAR) Model**
        - VAR is multivariate forecasting algorithm used multiple time series data simultaneously for forecasting the output variable

# DATA REPRESENTATION OF ALL VARIABLES FOR FIRST 200 MINUTES

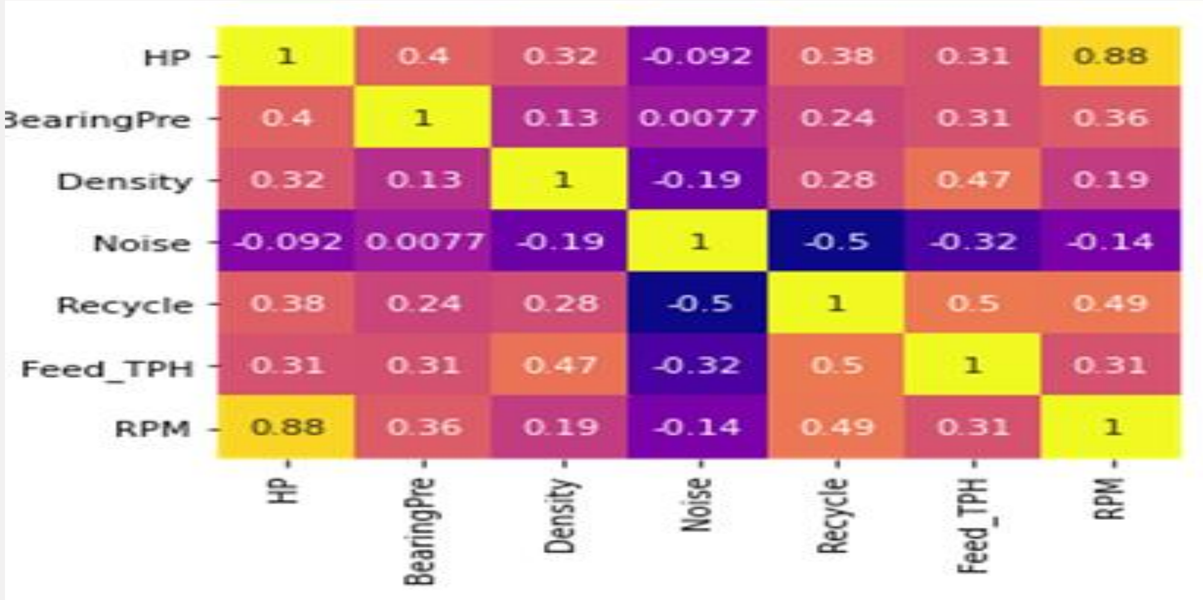


# DATA PREPROCESSING

- Preparing the data for the model
- Obtaining Feature Importance
- Obtaining Correlation between Variables
- Using **Granger's causality test** for finding out variables effecting the output variable using VAR

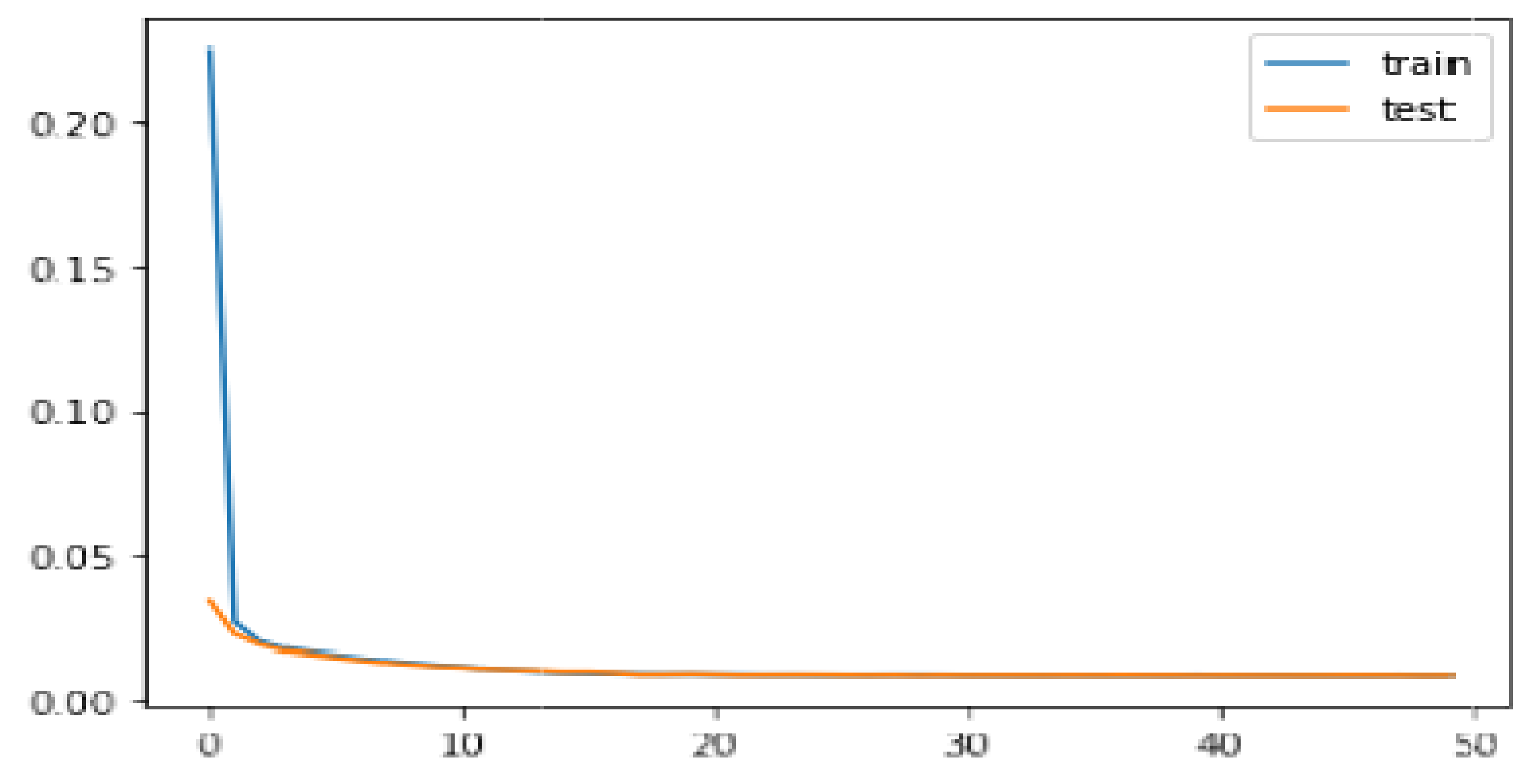



**Feature Importance of different variables of previous lag**




**correlation between variables**

# PLOT OF TRAINING AND VALIDATION LOSS




**AS THE VALIDATION LOSS IS ALMOST SIMILAR TO TRAINING LOSS SO WE CAN CONCLUDE THAT THE MODEL HAS NOT OVERFITTED**

# MODEL DEVELOPMENT

- Defining different models ( In our case LSTM and VAR)

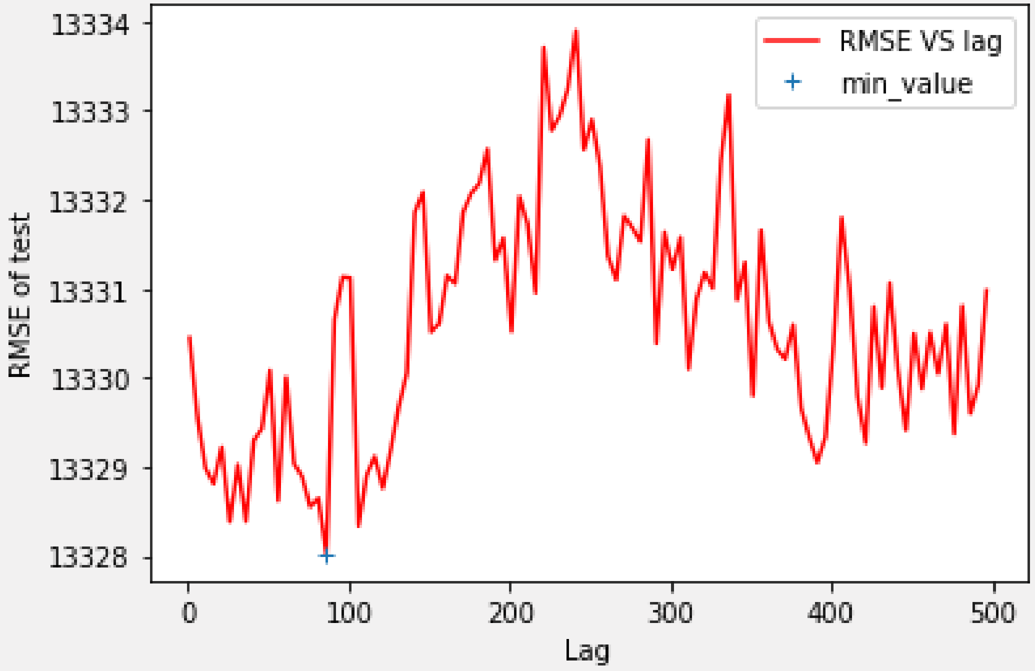
- Fit the model to the train data using random numbers for the parameters of neural network (like number of layers, number of epochs, optimizers etc.)

- Hyper parameter Tuning on number of layers, number of lags, number of neurons in a layer, number of epochs used in LSTM model on validation data set(using validation data)

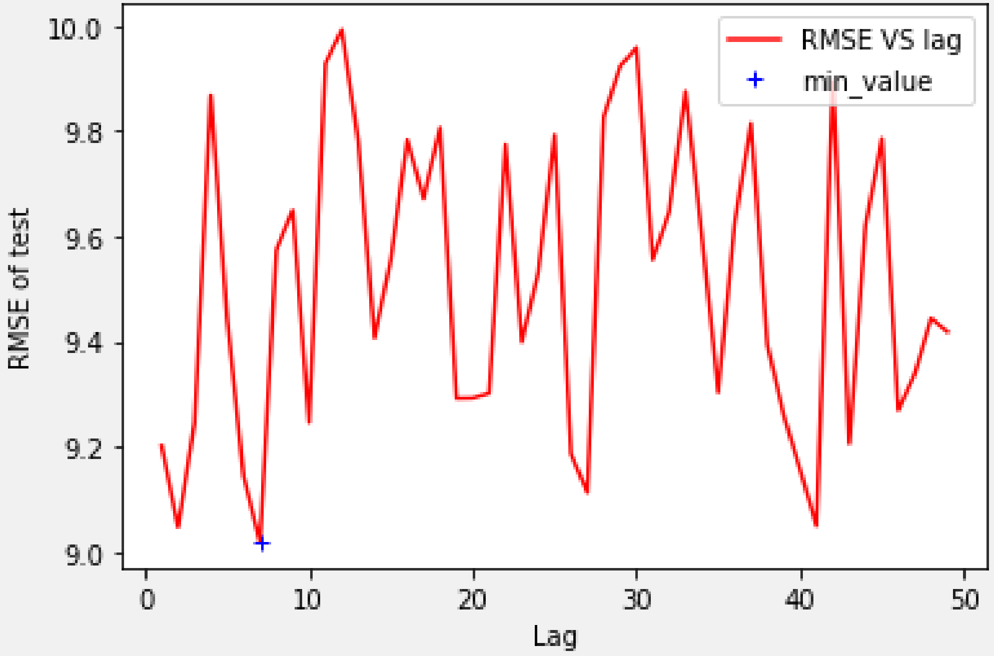
# TUNING THE HYPERPARAMETERS TO FIND THE OPTIMUM LAG OF TIME SERIES USING RMSE

# TUNING ON NUMBER OF LAGS

## Tuning on number of lags for short term prediction

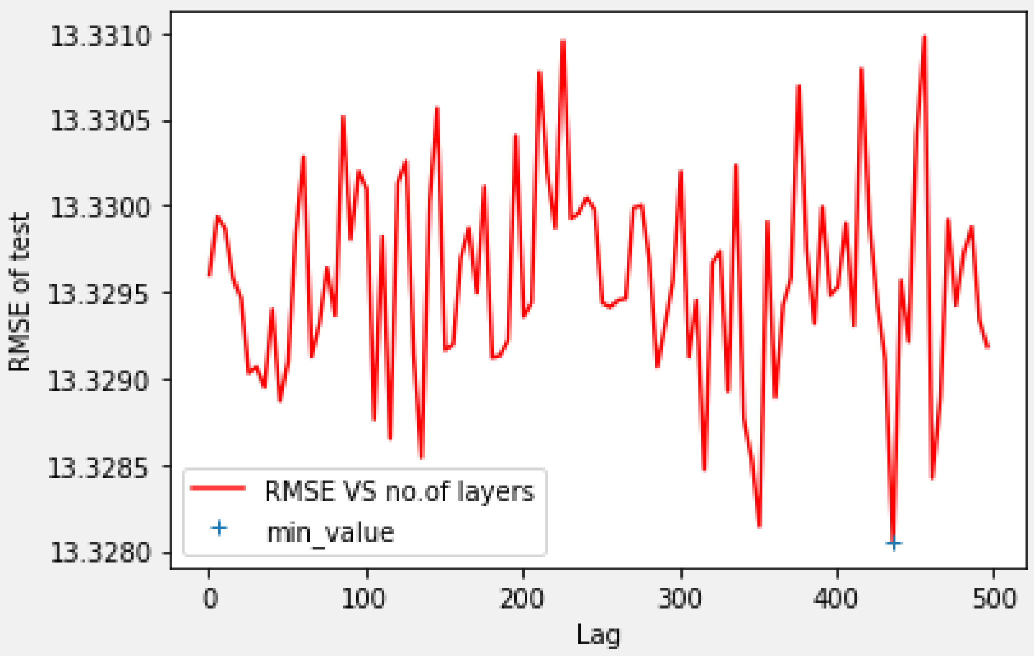



## Tuning on number of lags for medium term prediction

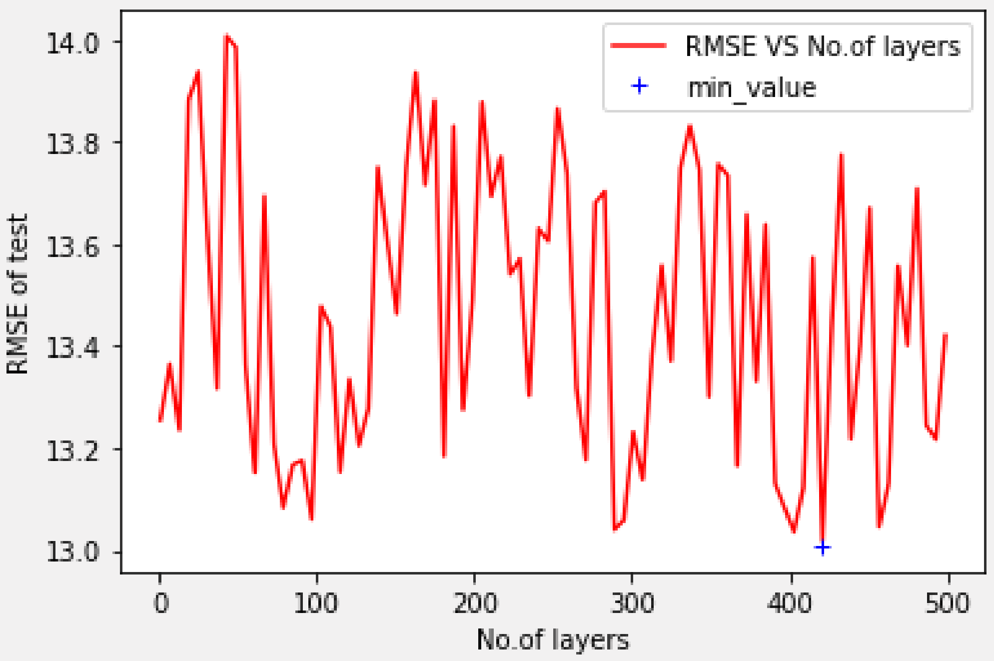


# TUNING ON NUMBER OF LAYERS

## Tuning on number of layers for short term prediction

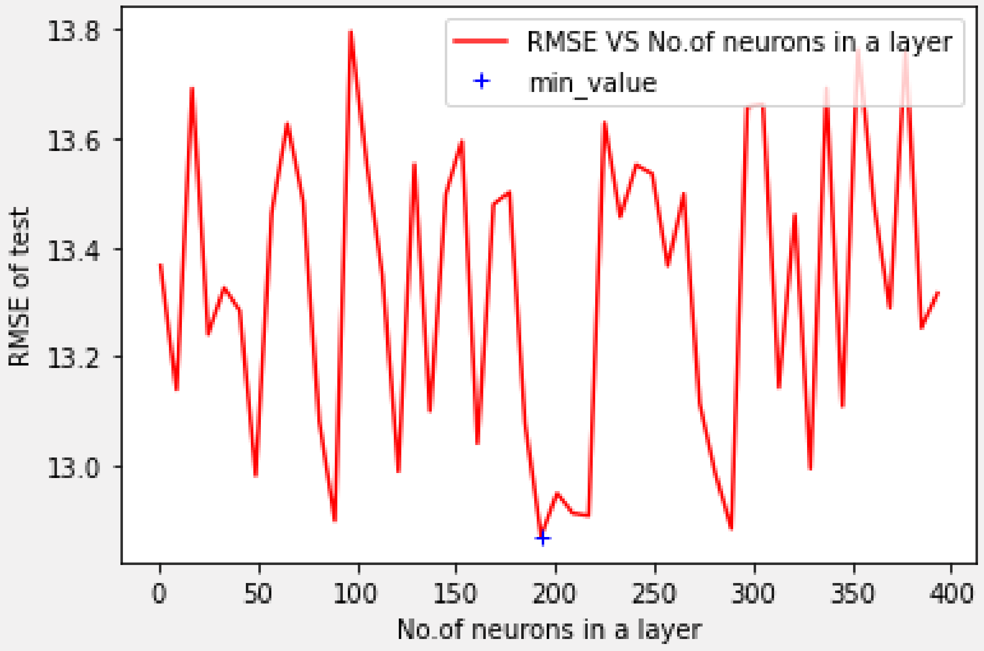



## Tuning on number of layers for medium term prediction

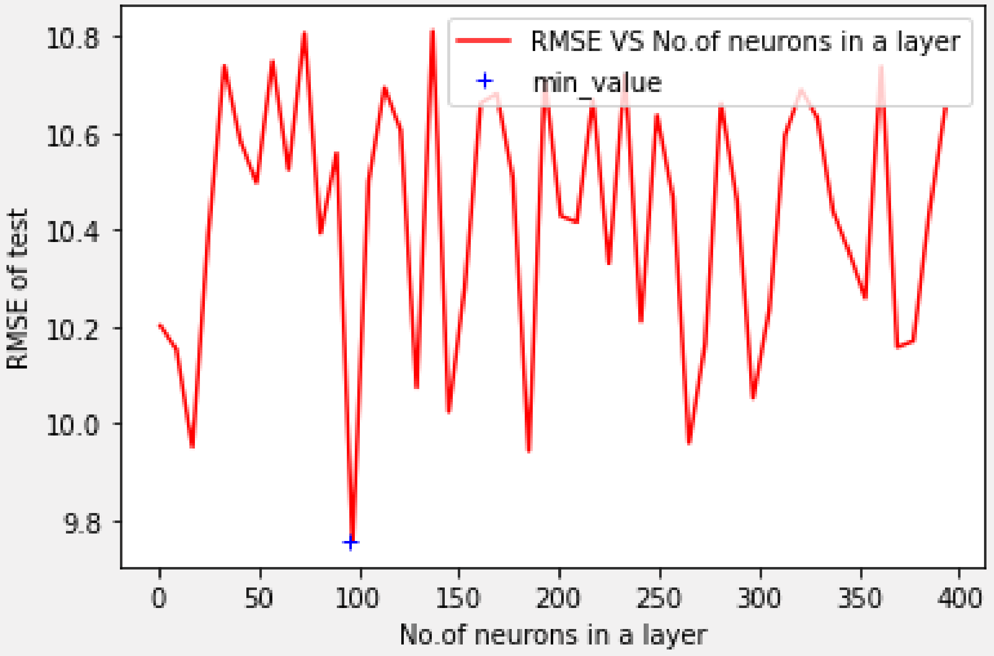


# TUNING ON NUMBER OF NEURONS IN A LAYER

## Tuning on number of neurons in a layer for short term prediction

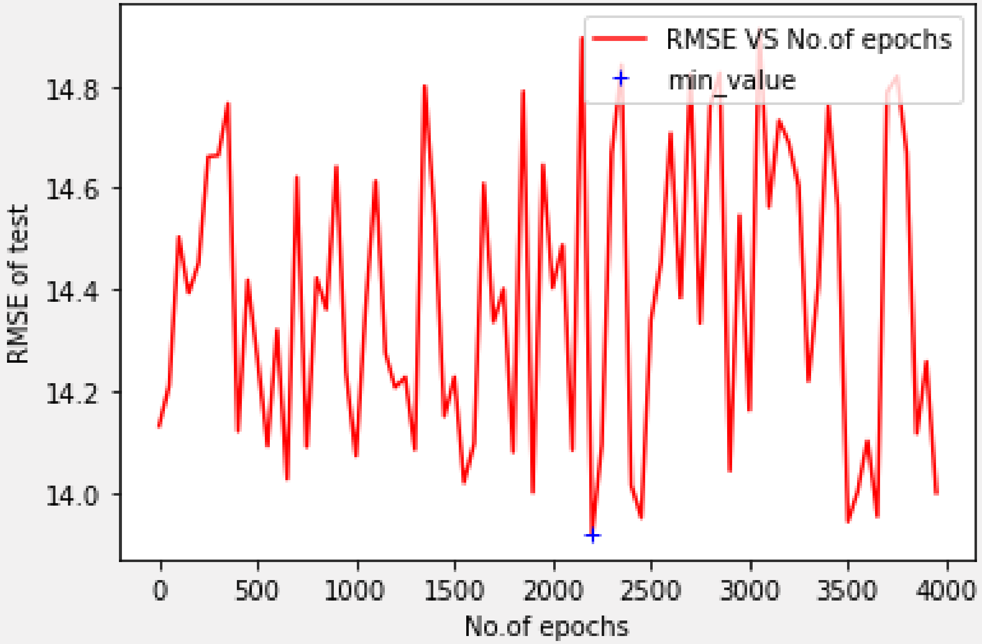



## Tuning on number of neurons in a layer for medium term prediction

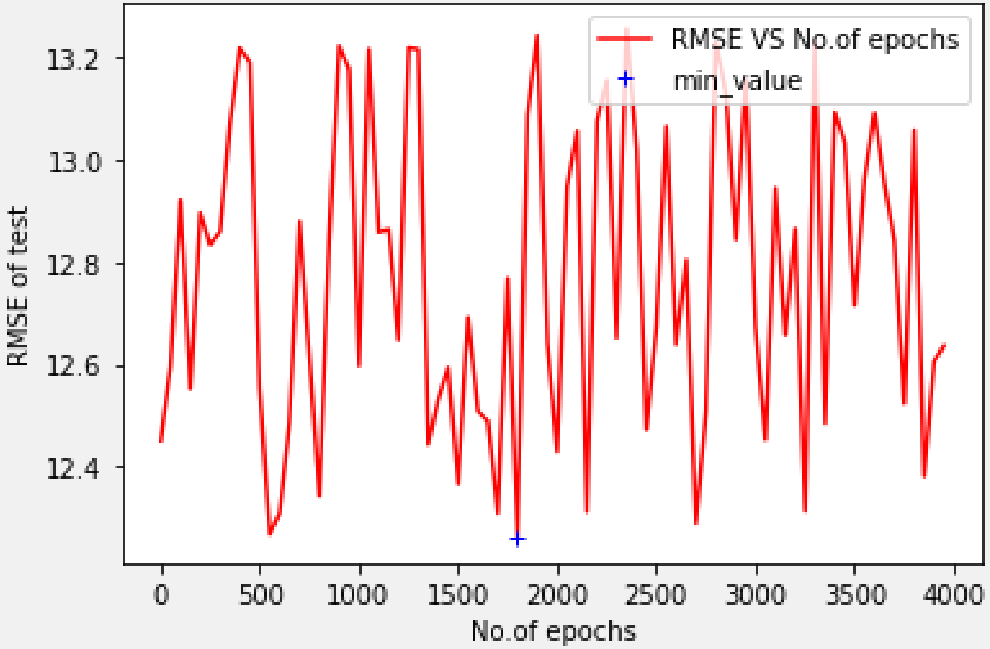


# TUNING ON NUMBER OF EPOCHS

## Tuning on number of epochs used for short term prediction




## Tuning on number of epochs used for medium term prediction



# GRANGER'S CAUSALITY TEST IN OUR DATA

| | HP_x | BearingPre_x | Density_x | Noise_x | Recycle_x | Feed_TPH_x | RPM_x |
|---|---|---|---|---|---|---|---|
| HP_y | 1.0000 | 0.1369 | 0.0290 | 0.1671 | 0.0822 | 0.3395 | 0.0113 |
| BearingPre_y | 0.1547 | 1.0000 | 0.1712 | 0.1818 | 0.5428 | 0.0539 | 0.0466 |
| Density_y | 0.6410 | 0.4642 | 1.0000 | 0.1390 | 0.0865 | 0.0510 | 0.5739 |
| Noise_y | 0.3484 | 0.0017 | 0.0760 | 1.0000 | 0.6737 | 0.1635 | 0.0635 |
| Recycle_y | 0.6896 | 0.3272 | 0.5832 | 0.0768 | 1.0000 | 0.6872 | 0.4087 |
| Feed_TPH_y | 0.0819 | 0.2681 | 0.0007 | 0.2844 | 0.1201 | 1.0000 | 0.2237 |
| RPM_y | 0.0146 | 0.2073 | 0.0970 | 0.1841 | 0.3311 | 0.1479 | 1.0000 |

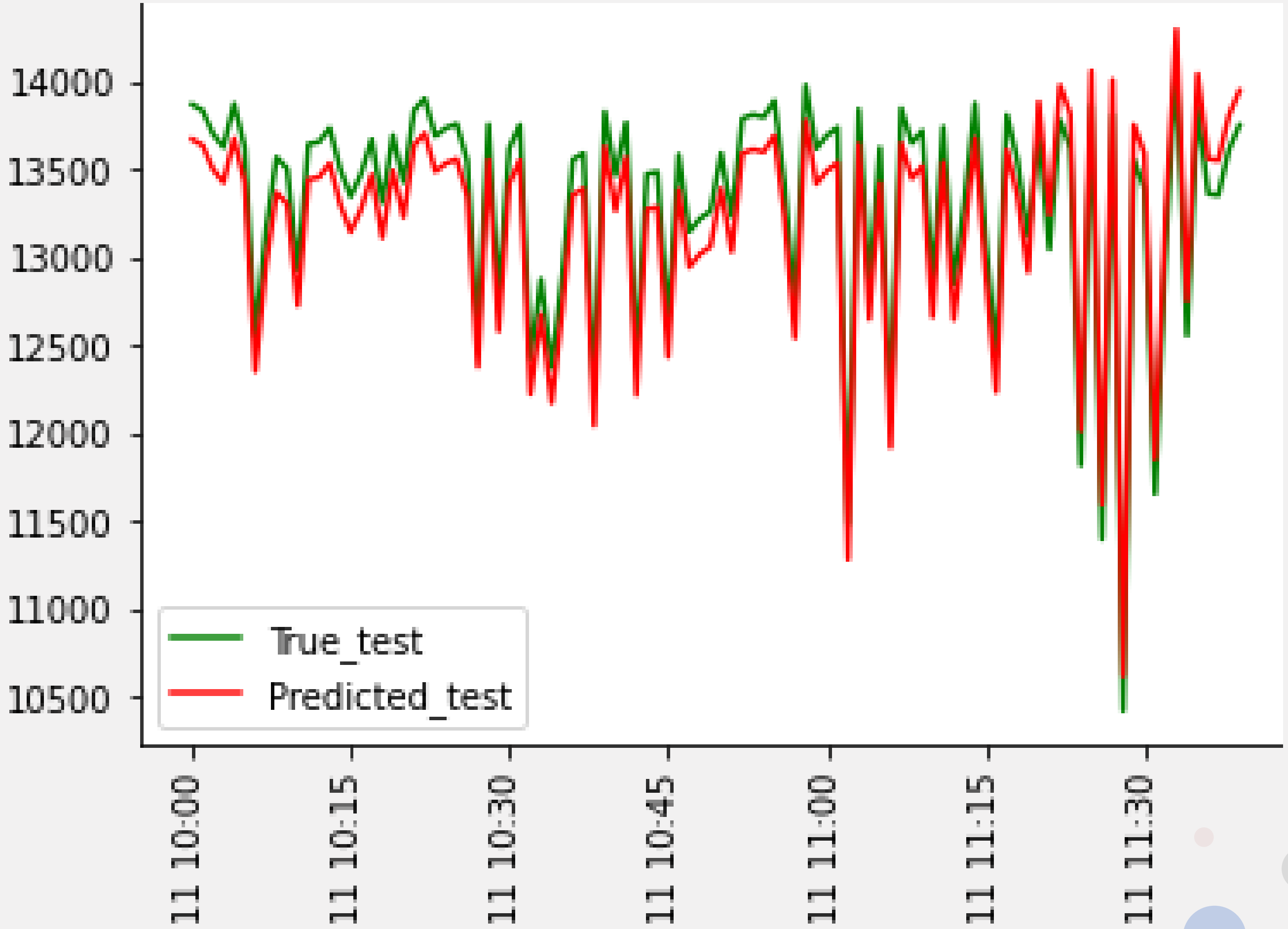
# MODEL EVALUATION

- Using Augmented Dickey Fuller test for stationarity of the data
- Based on Granger's causality test for VAR model, the power consumption depends only on density, RPM, recycle amount (critical value= 0.10)
- Based on AIC values obtained at different lags, we have selected the lag which has least AIC value as optimal lag for both short term and medium term predictions.
- Using Root Mean Squared Error(RMSE) as metric for comparing the LSTM and VAR model on test data.

# RESULTS FOR SHORT TERM PREDICTION

- For Short Term Prediction:

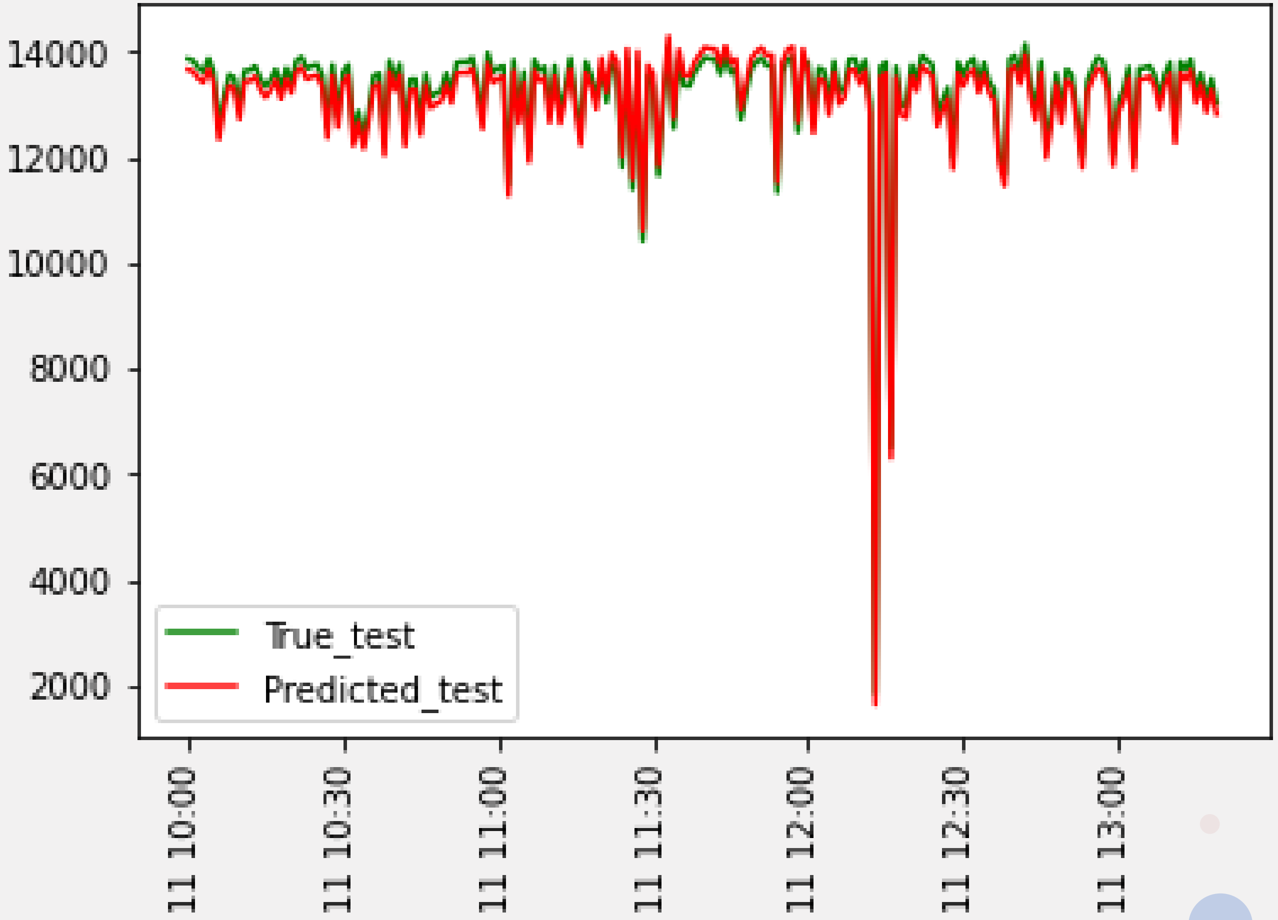
| Model used: | RMSE : | Correlation : |
|---|---|---|
| LSTM | 13.2567 | 0.921 |
| VAR | 49.1627 | 0.873 |



# RESULTS FOR MEDIUM TERM PREDICTION

- For Medium Term Prediction:

| Model used: | RMSE : | Correlation : |
|---|---|---|
| LSTM | 9.32 | 0.949 |
| VAR | 43.546 | 0.889 |



# CONCLUSION

- LSTM model predicted better than VAR model for both short term and medium term predictions based on RMSE metric and correlation between predicted and true values.

- The main reason LSTM performed better is that it can learn and handle the temporal dependencies for larger lag values as well as , LSTM architecture which we have used after hyper parameter tuning is a complex architecture so, it learn more complex dependencies where as VAR model is just a  linear model dependent on lags.

# Thank You