Analysis and Prediction of Market Clearing Price (MCP) Values in Power Markets

S. Abhinay Teja
AE, IIT Kanpur
sabhinayt23@iitk.ac.in

JLS Laharii
EE, IIT Kanpur
lakshmi23@iitk.ac.in

Abstract—In the deregulated and highly dynamic Indian electricity market, accurate price forecasting is critical for stakeholders aiming to optimize bidding strategies, plan operations, and manage financial risks. This work presents a comprehensive study of Market Clearing Price (MCP) forecasting through both statistical and machine learning approaches. We begin with an in-depth statistical analysis of MCP data from January to June 2023, uncovering strong periodic and seasonal patterns through decomposition and autocorrelation techniques. Classical models such as ARMA, harmonic regression, and SARIMAX using weather-based exogenous variables (temperature, wind, and solar irradiance) were evaluated but found inadequate due to high variance and non-linearity. Clustering further validated the dependency between MCP regimes and weather patterns. Advanced neural models LSTM and NARX were deployed for day-ahead MCP forecasting. The best LSTM configuration (hidden = 32, layers = 1, learning rate = 0.0005, batch size = 16) achieved a MAPE of 7.82%, accurately capturing MCP trends and peaks.

I. INTRODUCTION

The electricity sector in India has witnessed rapid transformation, driven by the rise of power exchanges that facilitate transparent and competitive energy trading. Among these, the Indian Energy Exchange (IEX) plays a dominant role, serving as the leading platform for day-ahead electricity transactions. The Market Clearing Price (MCP), determined through bidding on IEX, reflects the equilibrium between supply and demand in the short-term electricity market. In this dynamic environment, MCP is influenced by a variety of factors including fluctuating demand, fuel costs, renewable energy penetration, and weather variability. These factors contribute to sharp intraday and seasonal fluctuations, making price prediction increasingly complex. Accurate MCP forecasting is vital for power producers, utilities, and consumers to plan generation schedules, manage procurement costs, and optimize bidding strategies. Classical statistical models such as ARMA and SARIMAX provide a foundational understanding but often fall short in capturing the nonlinear and volatile nature of MCP data. To explore hidden patterns, clustering techniques are applied to group similar price-weather regimes, offering valuable insight into underlying structure. In parallel, advanced neural models like Long Short-Term Memory (LSTM) and Nonlinear AutoRegressive models with eXogenous inputs (NARX) are employed to capture complex temporal dependencies in MCP time series. This report examines these forecasting approaches using historical IEX data and evaluates their performance in terms of accuracy, interpretability, and responsiveness to volatility.

II. DATA DESCRIPTION

- **Period:** Data used for training spans from 1 January 2023 to 30 June 2023, while testing is conducted on data from 1 July 2023.
- **Frequency:** Hourly MCP (Rs/MWh).
- **Exogenous variables:** Temperature, Wind Speed, Solar Irradiance, Day of the Week, Week of the Day, Holiday.

III. STATISTICAL ANALYSIS

A. Descriptive Statistics

Table 1: Monthly MCP statistics (Jan-Jun 2023)

Month	Mean (Rs)	Variance($\times 10^5$)
Jan	6,178.96	110.56
Feb	6,680.51	93.61
Mar	5,426.49	46.29
Apr	5,417.24	68.13
May	4,737.77	58.70
Jun	5,369.77	76.33
Overall	5,620.46	79.21

The MCP exhibits large volatility and daily as well as seasonal effects.

B. Trend, Seasonality, and Stationarity

Autocorrelation analysis of the raw Market Clearing Price (MCP) series as shown in Figure 1, reveals strong autocorrelation at lag 1 and recurring peaks at multiples of lag 24, indicating the presence of both trend and daily seasonal components. The slow decay of the ACF further confirms that the series is not white noise and exhibits non-stationary behavior.

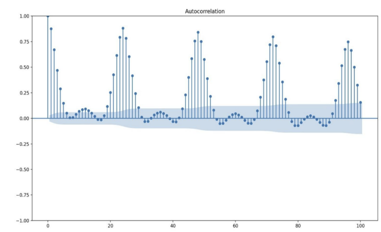

Figure 1: ACF Plot of Raw Data.

To transform the series into a stationary form suitable for modeling, classical decomposition is first employed to analyze and separate the seasonal and trend components as shown in Figure 2. This is followed by a two-step differencing process.

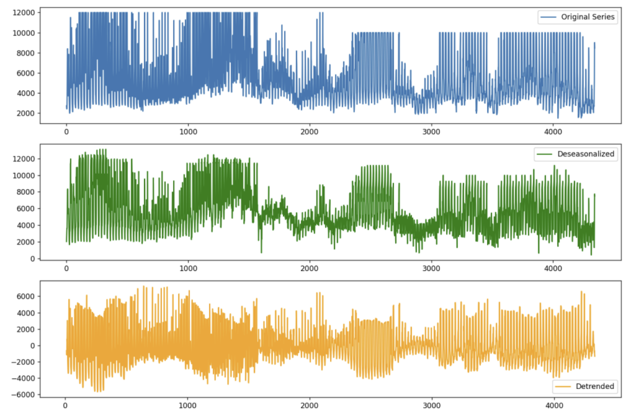

Figure 2: Decomposition of original series into deseasonalized and detrended components for analysis.

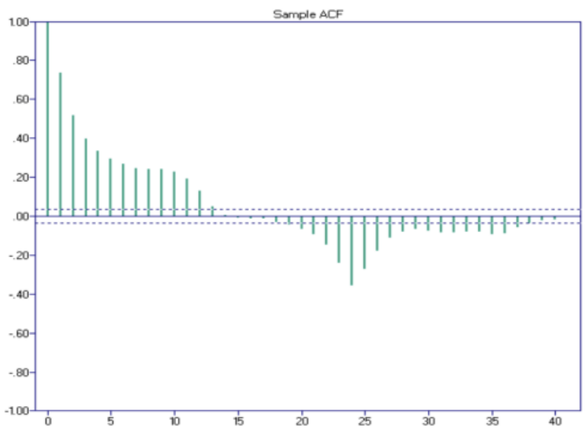

Figure 3: ACF Plot of first differenced data.

Seasonal differencing at lag 24 is initially applied to remove the dominant 24-hour cycle. However, the resulting series still exhibits a declining trend, as observed in the ACF of the differenced data in Figure 3. To address this, a second differencing at lag 1 is performed, effectively removing the trend and yielding a stationary series.The final transformed series exhibits constant mean and variance with significantly reduced autocorrelation, confirming its stationarity and making it suitable for classical time series models like ARIMA and SARIMAX.

IV. CCF ANALYSIS

We aim to estimate how the past values of weather variables such as temperature, wind speed, and solar irradiance impact the current Market Clearing Price (MCP). This helps us understand the delayed effect of weather conditions on electricity pricing.

A. *Plot of CCF for Temperature vs MCP*

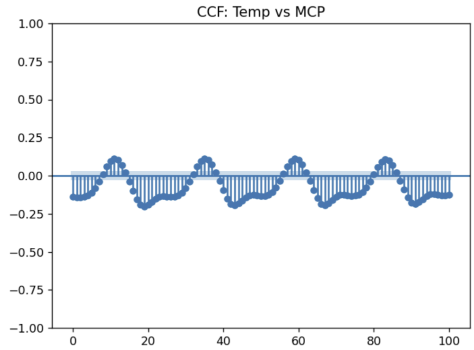

Figure 4: Cross-Correlation Function (CCF) for Temperature vs MCP

A regular pattern with peaks at lags 10, 20, ... indicates that MCP is influenced by temperature values from $(10 \times n)$ hours earlier. The predominantly negative correlations suggest an inverse relationship.

B. *Plot of CCF for Wind Speed vs MCP (Weak Correlation)*

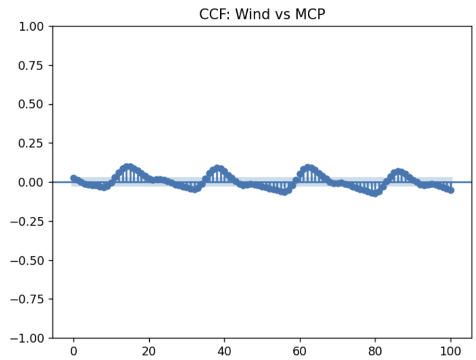

Figure 5: Cross-Correlation Function (CCF) for Wind Speed vs MCP

A weak cyclic pattern with peaks at lags 20, 40,.... suggests minimal influence of wind speed on MCP due to low correlation magnitudes.

C. *Plot of CCF for Solar Irradiance vs MCP (Stronger Relation)*

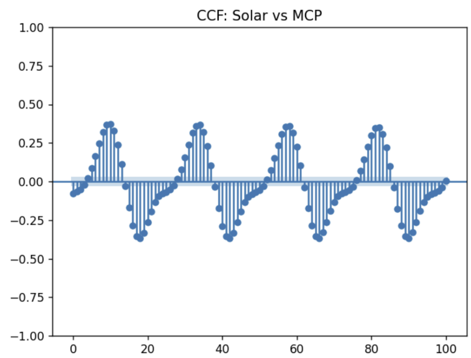

Figure 6: Cross-Correlation Function (CCF) for Solar Irradiance vs MCP

A sine-like pattern with a 24-hour period shows that solar irradiance from $(10–12) \times n$ hours ago is directly, and from $(20–24) \times n$ hours ago is inversely, related to current MCP.

D. Summary Table of CCF Observations

Variable	Dominant Lag(s)	CCF Pattern	Relation to MCP
Temperature	10, 20, ...	Periodic	Mostly inverse
Wind Speed	20, 40, ...	Weak periodic	Weak influence
Solar Irradiance	10–12, 20–24	Sine-like	Direct and inverse

Table 2: Summary of Weather Variable Impact on MCP based on CCF Analysis

V. CLASSICAL AND REGRESSION-BASED MODELING

A. ARMA(p,q)

To forecast day-ahead MCP values, ARMA modeling was applied following a thorough stationarization procedure. Initial model order selection was based on the behavior of the autocorrelation (ACF) and partial autocorrelation (PACF) plots of the differenced series from Figure 7. The ACF indicated a significant spike at lag 1 with rapid decay, suggesting a moving average component of order 3 (q=3), while the PACF gradually decayed up to lag 7, indicating an autoregressive component of order 7 (p=7). Thus, an ARMA(7,3) model was specified.

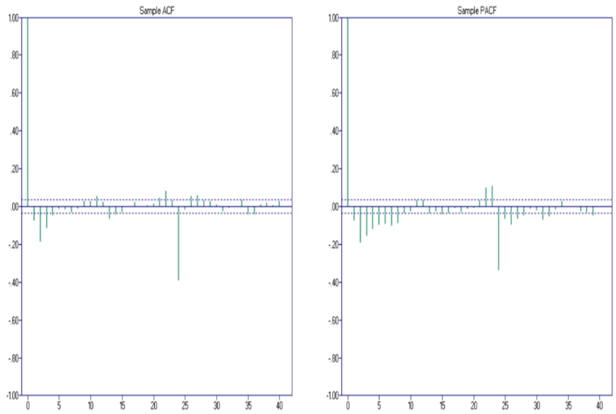

Figure 7: Plots of ACF and PACF for differenced series.

The model was estimated using maximum likelihood. However, all ARMA coefficients were found to be near zero, indicating that the differenced series closely resembles white noise. This suggests that, after removing trend and seasonality, no significant linear relationships remain in the data, thereby limiting the effectiveness of classical ARMA-based forecasting.

B. SARIMAX

After analyzing the cross correlation between Market Clearing Price (MCP) and exogenous variables (Temperature, Solar Irradiance, and Wind Speed), and a strong seasonal component as observed earlier in the time series of MCP data, Seasonal AutoRegressive Integrated Moving Average with eXogenous variables (SARIMAX) model, specifically SARIMA(1,0,2)$\times$(1,1,1)[24], was chosen for forecasting the day-ahead MCP values. The following plot in Figure 8 illustrates the day-ahead MCP predictions produced by the SARIMAX model. The red line displays the forecasted values over the selected horizon.

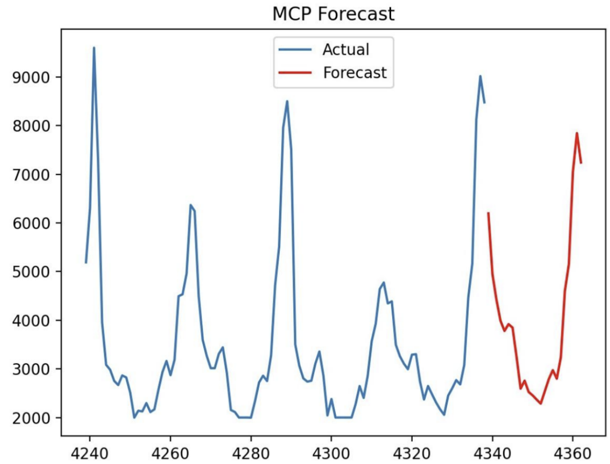

Figure 8: SARIMAX Predictions.

The table below represents the hour-wise forecasted MCP values:

Table 3: Sample SARIMAX Day-ahead MCP Forecast

Time	MCP	Time	MCP
00:00	6,726	12:00	2,453
01:00	5,424	13:00	2,367
02:00	4,772	14:00	2,666
03:00	4,185	15:00	2,929
04:00	3,812	16:00	3,486
05:00	4,020	17:00	4,212
06:00	4,830	18:00	5,372
07:00	5,609	19:00	6,916
08:00	5,362	20:00	8,818
09:00	4,189	21:00	10,952
10:00	3,352	22:00	12,459
11:00	2,698	23:00	7,908

The mean of the predicted MCP values is 4215.80, with the variance of 3.34×10^6.

These results indicate that while the SARIMAX model leverages relevant exogenous variables and seasonal components, it was not able to effectively capture the inherent nonlinearity present in the MCP series. As a result, the forecasting performance was limited, with the model yielding a relatively high Mean Absolute Percentage Error (MAPE) of 41.9%. This suggests that SARIMAX may not be fully adequate for modeling the complex dynamics of MCP, particularly during periods of rapid price fluctuations and nonlinear behavior.

C. Harmonic Regression

To effectively model the trend and pronounced seasonal patterns in the original data, harmonic regression was employed. After extensive experimentation, the best fit was achieved using a 10th-order polynomial combined with Fourier harmonics at indices that are multiples of 144 (with a data window up to 3445, as plotted by ITSM2000). This approach accurately captures both the underlying trend and the repeating seasonal fluctuations, providing a strong fit to the observed data. However, despite these results, harmonic regression struggles with nonlinear patterns found in real-world scenarios, making it unsuitable for practical applications where data complexity and unpredictability are high.

VI. CLUSTERING ANALYSIS

A clustering analysis was performed on market clearing price (MCP), temperature, wind speed, and solar irradiance using the KMeans algorithm. This revealed four distinct clusters, each representing specific combinations of weather conditions and associated MCP values. The results underscore the significant influence of meteorological variables on electricity market prices.

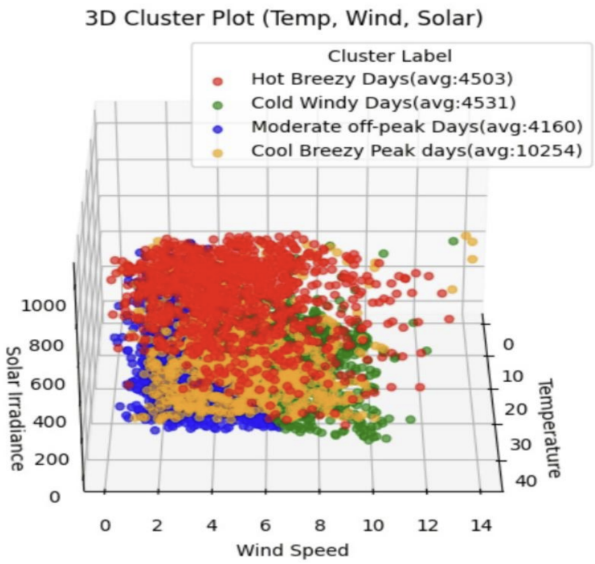

Figure 9: Clusters based on MCP and Weather data

Table below represents the centroids of MCP-weather clusters obtained from unsupervised clustering. Each cluster represents a distinct weather regime associated with a typical Market Clearing Price (MCP). The ”Red” and ”Green” clusters represent high-temperature and low-solar conditions respectively, while the ”Yellow” cluster shows the highest MCP under moderate weather conditions, possibly reflecting peak demand scenarios.

Table 4: Cluster centroids for MCP-weather regimes

Cluster	MCP (Rs)	Temp (°C)	Wind (m/s)	Solar (W/m^2)
Red	4,503.97	34.31	4.87	690.05
Green	4,531.75	17.47	7.00	32.84
Blue	4,160.34	23.65	3.34	116.54
Yellow	10,254.3	22.59	5.02	150.99

VII. NARX-BASED FORECASTING

Initial experiments using classical models such as ARMA and SARIMAX revealed their limitations in fully capturing the non-linear and exogenous dependencies in Market Clearing Price (MCP) behavior. These models, while effective for linear time series, struggle with sudden price spikes, weather-induced fluctuations, and complex temporal patterns common in energy markets. To overcome these limitations, a neural-network-based approach was adopted.

The NARX (Nonlinear AutoRegressive with eXogenous inputs) model was selected due to its ability to incorporate:

- Nonlinear relationships in the target series,
- Temporal dependencies via lagged MCP inputs,
- Contextual influence from weather and time-based exogenous features.

Data Preparation: Six months of hourly MCP and weather data were used for training, and a separate July 1st dataset was reserved for forecasting. The MCP values and weather variables were normalized using Min-Max scaling. Temporal features such as normalized hour-of-day, day-of-week, and a binary holiday indicator (with Sundays treated as holidays) were added to capture seasonal and weekly demand cycles.

Input Construction: The model inputs for each time step consist of:

- Lagged MCP values at 24, 48, and 168 hours (representing daily, bi-daily, and weekly cycles),
- Current exogenous variables (scaled weather features + time features).

This hybrid input allows the model to learn from both past trends and present contextual cues.

Model Architecture: The NARX neural network consists of a multi-layer perceptron (MLP) with:

- Two fully connected hidden layers with ReLU activations,
- Dropout for regularization,
- An output layer predicting the MCP for the current time step.

Training and Forecasting: The model was trained using the Adam optimizer and mean squared error (MSE) loss for 150 epochs.

Evaluation: The model's performance was evaluated using Mean Absolute Percentage Error (MAPE), achieving a final score of 13.44% on the July 1st data. This reflects a significant improvement in forecasting accuracy over classical models. The comparative forecast curve is shown in Figure 10.

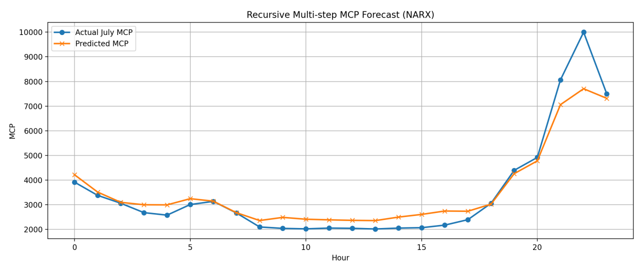

Figure 10: MCP Day Ahead Forecast using NARX Model

Table 5: Hourly Actual vs Predicted MCP using NARX Model

Hour	Actual	Predicted	Hour	Actual	Predicted
01	4701.53	4386.37	13	2763.19	3035.34
02	3518.37	3806.71	14	3266.04	2976.17
03	3746.95	3874.61	15	3194.85	3180.38
04	3469.67	3612.92	16	3298.08	3069.47
05	3974.05	3748.08	17	3710.16	2943.64
06	4596.42	3669.18	18	3859.03	3009.84
07	5116.65	3349.20	19	4387.26	3581.77
08	3905.00	3114.53	20	6477.71	5309.98
09	2778.61	2872.97	21	6101.57	5329.89
10	2722.38	3115.77	22	9306.60	8707.35
11	2557.08	2975.59	23	10000.00	9271.86
12	3251.89	3026.97	24	10000.00	8691.07

VIII. LSTM-BASED FORECASTING

While feedforward models like NARX can capture nonlinear dependencies using lagged inputs and exogenous features, they do not maintain an internal temporal state. This limits their capacity to model subtle sequential dependencies in the Market Clearing Price (MCP) series. To address this, a recurrent neural network (RNN) architecture based on Long Short-Term Memory (LSTM) units was adopted. LSTMs are well-suited for time series prediction as they can retain information over longer sequences through memory cells, effectively modeling sequential dependencies that span multiple time steps.

Data Preparation: Six months of hourly MCP and weather data were used for training. MCP values were scaled using Min-Max normalization, and weather features were augmented with temporal context: normalized hour of day, day of week, and a binary holiday flag (marking Sundays as holidays). These exogenous features were crucial for capturing periodic market behavior influenced by human activity and environmental conditions.

Input Construction: Each input vector to the LSTM model comprised:

- Lagged MCP values at 24, 48, and 168 hours,
- Corresponding weather and time-based features at the current hour.

Inputs were reshaped into three-dimensional tensors of shape $(\text{batch\_size}, 1, \text{input\_size})$ to conform to LSTM input requirements, where each "sequence" consisted of a single timestep with rich contextual input.

Model Architecture: The LSTM network consisted of:

- One LSTM layer with hidden size 32,
- A fully connected output layer to produce the final MCP value.

Hidden and cell states were initialized at each forward pass. The model was trained using the Adam optimizer and MSE loss over 75 epochs.

Recursive Forecasting: A recursive multi-step forecasting strategy was used to predict MCP values across the July horizon. For each future hour, the model was fed with lagged MCP values — initially from ground truth, and progressively from its own predictions once the forecast horizon advanced beyond available historical values. This recursive design reflects real-world deployment where only past predictions are accessible during live inference.

Evaluation: The LSTM model achieved a final **Mean Absolute Percentage Error (MAPE) of 7.82%** on July 1st data. The best configuration was found with a hidden size of 32, 1 stacked LSTM layer, learning rate of 0.0005, and batch size of 16. A comparative plot of actual vs. predicted MCP values is shown in Figure 11.

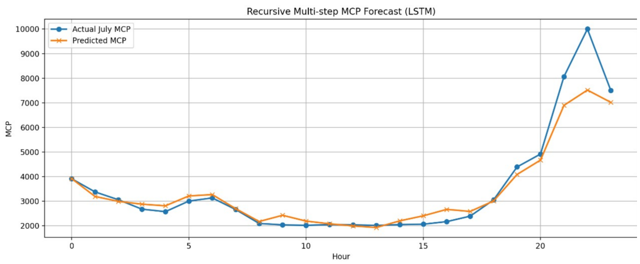

Figure 11: Recursive MCP Forecast for July Using LSTM Model

IX. MODEL COMPARISON

Table 6: July MCP Forecast Model Comparison

Model	MAPE (%)
SARIMAX	41.9
NARX	13.44
LSTM	**7.82**

The SARIMAX model showed limited accuracy with a MAPE of 41.9%. NARX reduced the error to 13.44% by modeling nonlinear patterns, while LSTM achieved the best performance with 7.82% MAPE due to its strength in capturing temporal dependencies.

X. CONCLUSION

This report presented a comprehensive approach to analyzing and forecasting Market Clearing Price (MCP) in the Indian electricity market using a blend of statistical and machine learning techniques. Initial exploration through autocorrelation, decomposition, and cross-correlation confirmed the strong presence of seasonality, trends, and exogenous weather-driven influences in the MCP series. Classical models such as ARMA and SARIMAX were employed after appropriate stationarization and inclusion of relevant exogenous variables; however, they struggled to handle nonlinearity and yielded high prediction errors.

Clustering analysis revealed distinct MCP-weather regimes, underscoring the contextual variability of electricity prices and motivating a shift toward more adaptive models. The NARX neural model, incorporating lagged

MCP values and weather/time-based features, significantly improved forecasting accuracy, reducing MAPE to 13.44%. However, it lacked the ability to capture sequential dependencies across time.

To address this, a Long Short-Term Memory (LSTM) model was implemented using recursive multi-step prediction. By leveraging both historical MCP lags and exogenous features while preserving temporal memory, the LSTM achieved superior forecasting performance, with a MAPE of 7.82%. The results affirm the advantage of deep learning architectures in modeling the dynamic, nonlinear, and temporal nature of power market prices.

Overall, the LSTM model proved to be the most effective solution, offering accurate and robust day-ahead MCP forecasts, which are critical for informed market participation, grid stability, and energy trading decisions in the evolving Indian power sector.

XI. ACKNOWLEDGEMENTS

We are extremely grateful to **Assistant Professor Dr. Swathi Battula(IIT Kanpur)** for providing us with the opportunity to undertake this project and for her constant encouragement and support throughout. We would also like to express our heartfelt thanks to our mentor **Mr. Naveen Kumar Thokala** for their invaluable guidance from building foundational understanding to sharing resources and providing continuous support at every stage of the project.

XII. REFERENCES

[1] DeepLearningAI. *Deep Learning Specialization*, YouTube Playlist. Available: YouTube Link

[2] Indian Energy Exchange. *Day Ahead Market Snapshot*. Available: iexindia.com

[3] P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer, 2nd Edition.

[4] M. P. Mohite. *Price Forecasting for Indian Electricity Markets*, M.Tech Thesis, Dept. of Sustainable Energy Engineering, IIT Kanpur, Dec. 2024.

[5] N. K. Thokala, A. Bapna, and M. G. Chandra. *A Deployable Electrical Load Forecasting Solution for Commercial Buildings*. ResearchGate, 2018. Available: ResearchGate Link

[6] C. Olah. *Understanding LSTM Networks*, colah's blog, Aug. 2015. Available: colah.github.io

[7] Greg Hogg. *LSTM Time Series Forecasting Tutorial in Python*, YouTube, 2023. Available: YouTube Video