INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

Modeling and Simulation of 'Univariate and Multivariate analytics' by applying 'Deep Learning and Machine Learning' Application of Support Vector Regression, Random Forest, K-Nearest Neighbors, Long Short-Term Memory and Gated Recurrent Units Algorithms for Time Series forecasting in the Neural Network Model.

Ashraf Shahriar

Department of Electrical and Computer Engineering (ECE), North South University, Bangladesh.

DOI : https://doi.org/10.51583/IJLTEMAS.2024.131202

Received: 05 December 2024; Accepted: 16 December 2024; Published: 30 December 2024

Abstract: The research focuses on the modeling and simulation of univariate and multivariate time series analysis by applying machine learning (ML) and deep learning (DL) techniques for accurate forecasting. The study includes the individuals and combined application of algorithms such as Support Vector Regression (SVR), Random Forest (RF), K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) in the context of neural networks. These techniques are evaluated for their analytical performance in a number of time series estimation settings. ML approaches are known for their simplicity and efficiency in small data sets. On the other hand, DL approaches require larger data sets and more computational resources, but offer higher accuracy and flexibility in demanding setups. The study concludes by highlighting the importance of choosing the right algorithm based on the nature of the data and the forecast object. The results provide valuable insights for applications in finance, energy, healthcare, and other fields where time series forecasting plays an important role. Future research may consider hybrid and interpretability-enhanced approaches to develop applications of these models in real-world settings. In this study, for the forecasting specification especially Univariate analysis, the demonstrations show the actual values for the past 15 days and the forecasts for each model for the next 10 days.

Key words: Univariate, Multivariate, SVR, RF, KNN, LSTM, GRU, Bangladesh, Time Series, Sylhet, Time Series, Neural networking.

I. Introduction

This study investigates and compares the effectiveness of SVR, RF, KNN, LSTM, and GRU algorithms in forecasting univariate and multivariate time series. By integrating these models in a neural network context, the study attempts to provide insight into their performance and classify the best methods for different forecasting scenarios. Furthermore, the study investigates the trade-off between ML and DL techniques in terms of forecast accuracy, computational power, and model complexity. In this framework, univariate time series studies are concerned with forecasting a single time-dependent variable, while multivariate time series forecasting includes multiple variables that may be interrelated or affect each other. The challenge of multivariate forecasting lies in modeling the complex relationships between these sets of variables. Therefore, it is important to apply algorithms that are suitable to include such dynamics. The results of this study are expected to benefit practitioners and researchers in areas where time series research is important by providing guidance for selecting appropriate task prediction models. The study area is located in Sylhet district of Bangladesh, at longitude 92.16 and latitude 24.8392.

II. Methodology

A. This systematic method guarantees a complete assessment of ML and DL algorithms, presenting actionable insights into their suitability for univariate and multivariate time collection forecasting applications.

B. Data Collection and Preprocessing

- **Data Sources:** Data Sources: Appropriate statistics were gathered from NASA for Multivariate and BWDB for univariate time collection evaluation.
- **Data Cleaning**: Data Cleaning: Missing values have been dealt with via way of means of imputation techniques and the outliers have been perceived and indifferent to make certain statistics excellence.
- **Normalization**: The statistics became standardized the use of performances like Min-Max Scaling to enhance version schooling and convergence, specifically for DL algorithms.
- **Feature Engineering**: For multivariate analytics, correlation evaluation became directed to categorize the important things and systems that effect the goal variable. Lag systems, transferring averages, and different time-structured variables have been formed to seize temporal shapes.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

C. Model Selection

The studies carried out each ML and DL fashions to assess and prediction time collection statistics:

- **Machine Learning Models:**
  - Support Vector Regression (SVR)
  - Random Forest (RF)
  - K-Nearest Neighbors (KNN)
- **Deep Learning Models:**
  - Long Short-Term Memory (LSTM)
  - Gated Recurrent Units (GRU)
- **Model Training and Simulation**
  - **Univariate Analytics:**
  - Different time collections have been evaluated by the use of single-variable fashions to forecast destiny values primarily based totally on the ancient trends.
  - **Multivariate Analytics:**
  - Models merged a couple of interdependent variables to manipulate relationships among systems for advanced prediction exactness.
  - **Hyperparameter Tuning:**
    - Grid Search and Random Search were used to enhance model parameters
  - **Loss Functions:**
    - Mean Squared Error (MSE) and Mean Absolute Error (MAE) became carried out because the loss feature for schooling fashions.

D. Evaluation Metrics

- To check the overall performance of every version, the under-valuation metrics have been carried out:
- **Mean Absolute Error (MAE)**: For comparing the common absolute alteration among prophesied and the real values.
- **Root Mean Squared Error (RMSE)**: Evaluates the importance of estimating mistakes with higher emphasis on cumbersome deviations.

E. Mean Absolute Percentage Error (MAPE): Analyzes blunders as a percent of actual values, useful for evaluating athwart datasets.

F. Comparative Analysis

- **Machine Learning vs. Deep Learning:** The overall performance of ML fashions (SVR, RF, KNN) became related to DL fashions (LSTM, GRU) to categorize the strengths and weaknesses.
- **Univariate vs. Multivariate:** Consequences from univariate and multivariate analytics have been likened to demonstrate the assistances of leveraging inter-variable dependencies.

G. Simulation and Visualization

- **Model Forecasts:** Forecasts had been analyzed for short-time period and long-time period horizons.
- **Visualization:** Forecast vs. real values had been designed to visualize version overall performance and classify the styles or variances.
- **Error Analysis:** Lingering plots and numerical checks had been completed to evaluate version correctness and come across biases.

H. Hybrid Approach

- **Combination of ML and DL Models**: Hybrid fashions have been advanced with the aid of combining ML strategies with DL algorithms. This technique aimed to leverage the strengths of each paradigm to increase the predicting correctness.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

I. Computational Setup

- **Frameworks:** Python-primarily based totally libraries had been implemented for understanding ML and DL fashions.
- **Hardware:** Training changed into directed on GPU-improved systems to deal with the computational lines of the deep mastering fashions.

J. Optimization: Early preventing and dropout layers had been hired in DL fashions to keep away from overfitting and development simplification.

K. Validation and Deployment

- **Cross-Validation**: Time-primarily based totally cross-validation changed into carried out to verify the robustness of the version forecasts.
- **Deployment**: The best-acting fashions had been packaged into the scalable contexts for placement inside the predictive analytics pipelines.

L. ***Output Illustration:*** To deal with the outcome, changed into accompanied by an established approach. Here additionally covered the plotting of beyond facts and destiny predictions as intreated.

- **Preprocessing:** Handle lacking values and normalize or standardize the dataset. And for the multivariate analysis, choose applicable functions wherein different variables influencing the target.
- **The plot will show:** The plot will show Actual values for the ultimate 15 days, and the Predictions from every version for the subsequent 10 days.
    - **Visualization:** Plotted real facts designed for the ultimate 15 days and the Predicted facts for the subsequent 10 days. Separate traces designed for the predictions from every version for evaluation.

M. Split data for Training and Testing

The enter statistics is break up as education and testing (Figure:2): 65% for Training and 35% for Testing analysis.

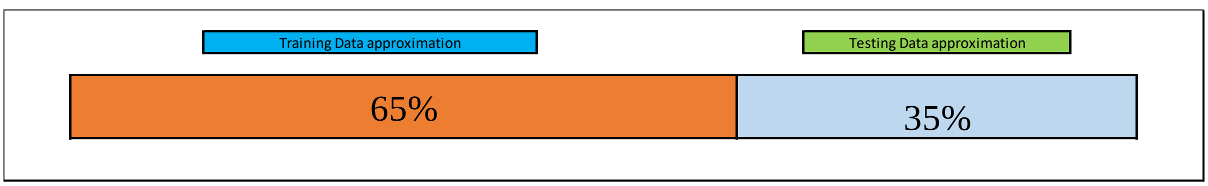


Figure 1: Split data for training and testing

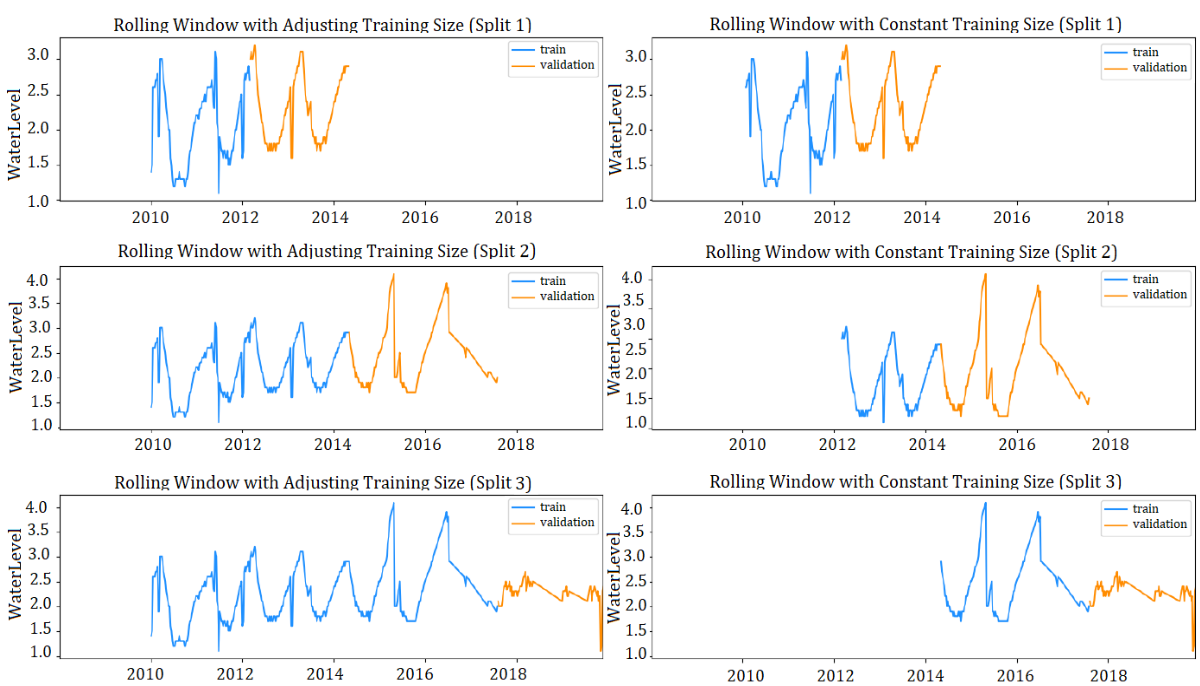


Figure 2: Split data for training and testing rolling window with periodical analysis, Sylhet

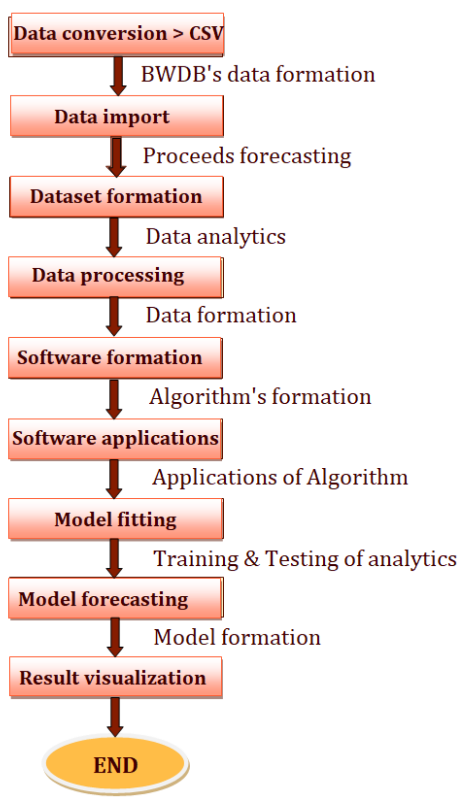
Observations: The graph titled (Figure 2) compares numerous rolling window procedures this is the changing schooling length vs. regular schooling length throughout 3 splits (Split 1, Split 2, and Split 3).

- **Rolling Window with Adjusting Training Size:** This method adapts the schooling length for in all likelihood upgrades withinside the version gaining knowledge of from incidental traits.
- **Rolling Window with Constant Training Size:** This approach keeps a regular schooling length to keep away from the ability danger of old statistics affecting version accuracy.
- **Split-wise Observations:** Across all 3 splits, the rolling window techniques display wonderful separations among schooling and validation statistics.
- **Split 1:** Both techniques cognizance on in advance years, and the validation set is smaller in comparison to later splits.
- **Split 2:** The statistics factors span extra current years in comparison to Split 1. The validation set shifts ahead with time.
- **Split 3:** It makes a specialty of the maximum current statistics. The validation set is longer, reflecting a bigger checking-out window.
- **Comparison: adjusting training size captures vs constant training size:** The adjusting schooling length captures extra incidental traits however dangers incorporating older statistics that won't constitute modern dynamics. The regular schooling length discards older statistics, making sure extra current traits are prioritized however doubtlessly lacking longer-time period patterns.

N. Applied Algorithms, Time Series Forecasting and Evaluation Criteria:

- ***Applied Algorithms:*** Support Vector Regression, Random Forest, K-Nearest Neighbor Model, LSTM, GRU, and LSTM+GRU models are used to forecast groundwater level.
- **Approaches of Time Series Forecasting:** My research focuses on Support Vector Regression, Random Forest, K-Nearest Neighbor, LSTM, GRU, and LSTM+GRU models for groundwater level forecasting.
- **Evaluation Criteria:** In this study, statistical formation was applied to evaluate the simulations for MAE, MSE, RMSD, P-value, and R2. Overall, the frequency is shown in Figure 2.

O. Research Design and formation of the process flow



Flowchart 1: Research and process flow

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

Time series forecasting to predict future groundwater levels taking into account factors such as groundwater depth, parapet height, and geographical orientation checking the Latitude and Longitude. (Table 1)

SL	DISTRICT	OLD ID	waterLevel	RL PARAPET (m)	PARAPET HEIGHT (m)	DEPTH (m)	LATITUDE	LONGITUDE
5	Sylhet	SY071	2.6	10.5	0.46	24.4	24.8392	92.16
9	Sylhet	SY071	1.9	10.5	0.46	24.4	24.8392	92.16
42	Sylhet	SY071	1.6	10.5	0.46	24.4	24.8392	92.16
3	Sylhet	SY071	2.6	10.5	0.46	24.4	24.8392	92.16
4	Sylhet	SY071	2.6	10.5	0.46	24.4	24.8392	92.16

Table 1: water table depth, parapet height, and geographical directs (main and original)

```
bist100['date'] = pd.to_datetime(bist100.date)
bist100.head()
```

	SL	DISTRICT	ID	waterLevel	RL PARAPET (m)	PARAPET HEIGHT (m)	DEPTH (m)	LATITUDE	LONGITUDE
0	1	Sylhet	SY071	1.4	10.5	0.46	24.4	24.8392	92.16
1	2	Sylhet	SY071	1.5	10.5	0.46	24.4	24.8392	92.16
2	3	Sylhet	SY071	2.6	10.5	0.46	24.4	24.8392	92.16
3	4	Sylhet	SY071	2.6	10.5	0.46	24.4	24.8392	92.16
4	5	Sylhet	SY071	2.6	10.5	0.46	24.4	24.8392	92.16

Table 2: water table depth, parapet height, and geographical directs (forecast).

Applications and overview of SVR, RF, KNN, LSTM, GRU, LSTM+GRU algorithms. This chart provides a visual comparison of the performance of different machine learning algorithms in predicting GWL on Sylhet (Table: 3).

Algorithms Name	Train RMSE	Test RMSE	Train MSE	Test MSE	Train MAE	Test MAE	Train VRS	Test VRS	Train R2 Score	Test R2 Score	Train MGD	Test MGD	Train MPD	Test MPD
Support Vector Regression	0.347433	0.359753	0.120710	0.129423	0.289585	0.217165	0.610634	0.063568	0.609046	0.023765	0.026110	0.023138	0.055008	0.053643
Random Forest	0.186454	0.321839	0.034765	0.103580	0.142269	0.201723	0.887759	0.218889	0.887403	0.218693	0.007370	0.018354	0.015421	0.042650
K Nearest neigbor	0.471409	0.357265	0.222226	0.127638	0.383947	0.222874	0.288293	0.051502	0.280256	0.037225	0.044699	0.021909	0.097806	0.051932
LSTM	0.481184	0.290422	0.231538	0.084345	0.388651	0.188874	0.265600	0.366508	0.250097	0.363787	0.046214	0.015087	0.101519	0.034759
GRU	0.483063	0.354106	0.233350	0.125391	0.376583	0.267349	0.297243	0.378148	0.244229	0.054172	0.046742	0.023278	0.102631	0.053010
LSTM+GRU	0.484155	0.288242	0.234406	0.083084	0.394018	0.183692	0.260870	0.379342	0.240808	0.373300	0.047150	0.014845	0.103209	0.034220

Table 3: visual comparison of the performance of different machine learning algorithms (loss function)

Algorithms Name	Main Tuning/Hyper parameters	Train RMSE	Test RMSE	Train MSE	Test MSE	Train MAE	Test MAE	Train VRS	Test VRS	Train R2 Score	Test R2 Score	Train MGD	Test MGD	Train MPD	Test MPD
Support Vector Regression	kernel= rbf, C= 1e2, gamma= 0.1,epsilon=0.1	0.347433	0.359753	0.120710	0.129423	0.289585	0.217165	0.610634	0.063568	0.609046	0.023765	0.026110	0.023138	0.055008	0.053643
Random Forest	n_estimators=100 ,random_state=1	0.186454	0.321839	0.034765	0.103580	0.142269	0.201723	0.887759	0.218889	0.887403	0.218693	0.007370	0.018354	0.015421	0.042650
K Nearest neigbor	n_neighbors=15,metric=minkowski	0.471409	0.357265	0.222226	0.127638	0.383947	0.222874	0.288293	0.051502	0.280256	0.037225	0.044699	0.021909	0.097806	0.051932
LSTM	loss=mse,optimizer=adam,3 lstm layers with 32 ...	0.481184	0.290422	0.231538	0.084345	0.388651	0.188874	0.265600	0.366508	0.250097	0.363787	0.046214	0.015087	0.101519	0.034759
GRU	loss=mse,optimizer=adam,4 gru layers with 32 n...	0.483063	0.354106	0.233350	0.125391	0.376583	0.267349	0.297243	0.378148	0.244229	0.054172	0.046742	0.023278	0.102631	0.053010
LSTM+GRU	loss=mse,optimizer=adam,2 gru and 2 lstm layer...	0.484155	0.288242	0.234406	0.083084	0.394018	0.183692	0.260870	0.379342	0.240808	0.373300	0.047150	0.014845	0.103209	0.034220

Table 4: visual comparison of the performance of different machine learning algorithms (Summary Chart)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

III. Modeling And Simulation

A. Univariate Time Series Forecasting for Groundwater Level (GWL):

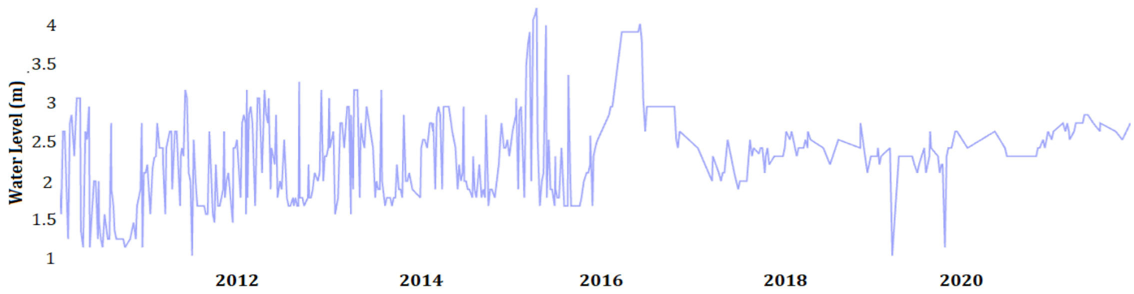


Figure 3: GWL chart, Water Level, Sylhet zone

Observations: Continuously monitor water levels for early warning systems in case of floods or droughts. Correlate this data with precipitation, temperature, or human activity to better understand trends.

SL	DISTRICT	UPAZILA	WELL ID	OLD ID	date	waterLevel	RL PARAPET (m)	PARAPET HEIGHT (m)	DEPTH (m)	LATITUDE	LONGITUDE
5	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-02	2.6	10.5	0.46	24.4	24.8392	92.16
9	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-03	1.9	10.5	0.46	24.4	24.8392	92.16
42	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-11	1.6	10.5	0.46	24.4	24.8392	92.16
3	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-18	2.6	10.5	0.46	24.4	24.8392	92.16
4	Sylhet	Sylhet Sadar	GT9162024	SY071	2010-01-25	2.6	10.5	0.46	24.4	24.8392	92.16

Table 5: Sorted Dataset by Date, Sylhet zone

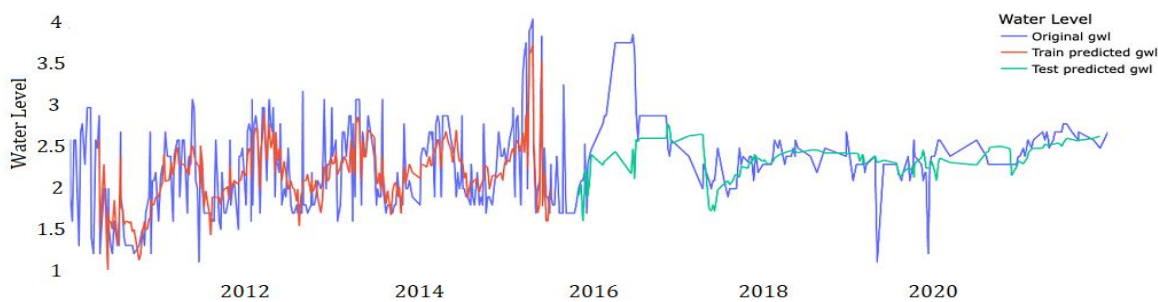


Figure 4: Original GWL vs predicted GWL, Sylhet zone by SVR

Observations: Trends are generally (Figure 4) consistent, but the model appears to smooth out some of the fluctuations in the original data. While the model captures trends well in both the training and test datasets, it can have problems capturing extreme values and high frequency fluctuations.

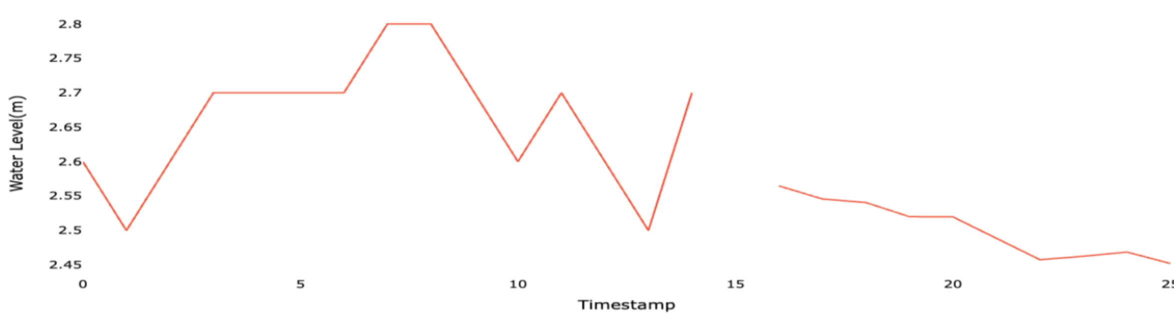


Figure 4a: Plotting last 15 days and next predicted 10 days by SVR of Sylhet

Observations: The decrease after (Figure 4a) the peak may indicate a change in conditions that affect the water level, such as: B. Seasonal changes, mining activity, or a decrease in replenishment rates.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

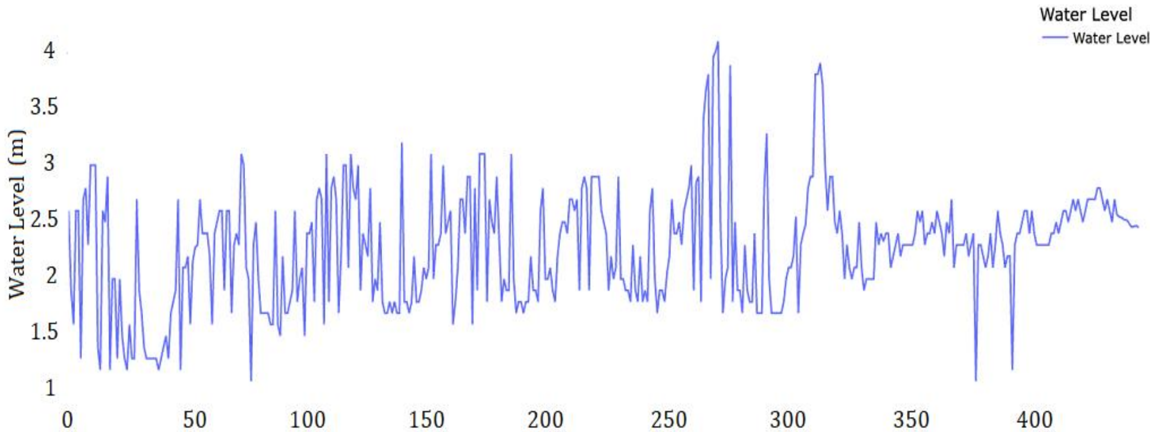


Figure 5: Plotting whole GWL with next 10 days prediction, Sylhet zone by SVR

Observations: The spikes and fluctuations (Figure 5) suggest that the data may be influenced by external factors such as seasonal recharge, pumping activity, and other environmental processes.

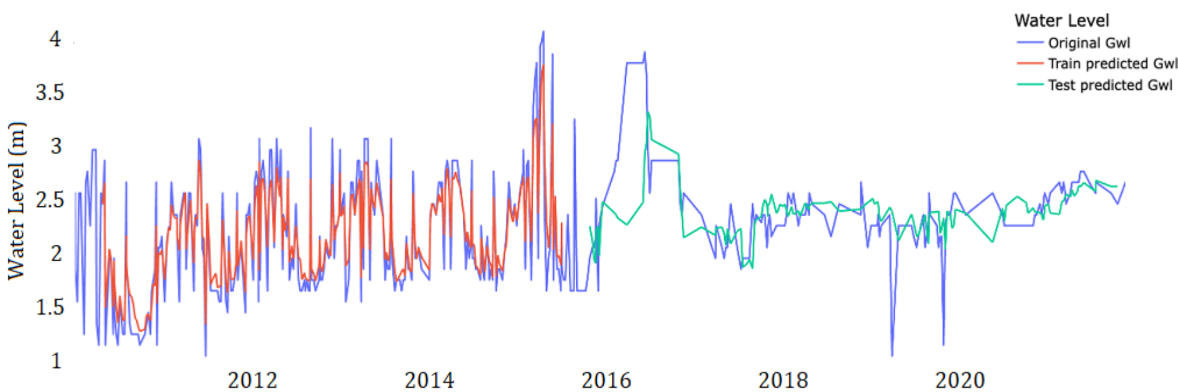


Figure 6: Comparison between original GWL vs Predicted GWL with chart of Sylhet zone by RF

Observations: The model seems (Figure 6) to work well during the training phase. The predictions during the testing phase are generally accurate, but there are some gaps and deviations from the actual data, especially in the case of sudden changes.

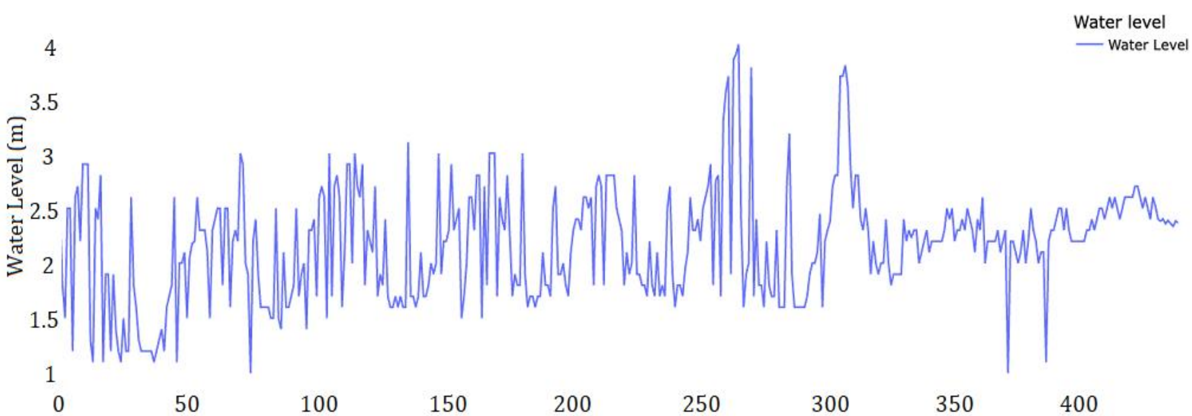


Figure 7: Plotting whole GWL with the next 10 days prediction, Sylhet zone by RF

Observations: The sample shifts (Figure 7) from exceptionally variable to fairly steady, probably indicating a greater balanced hydrological situation withinside the latter period. There appears to be no clean long-time period trend, however localized spikes and drops dominate the chart.

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

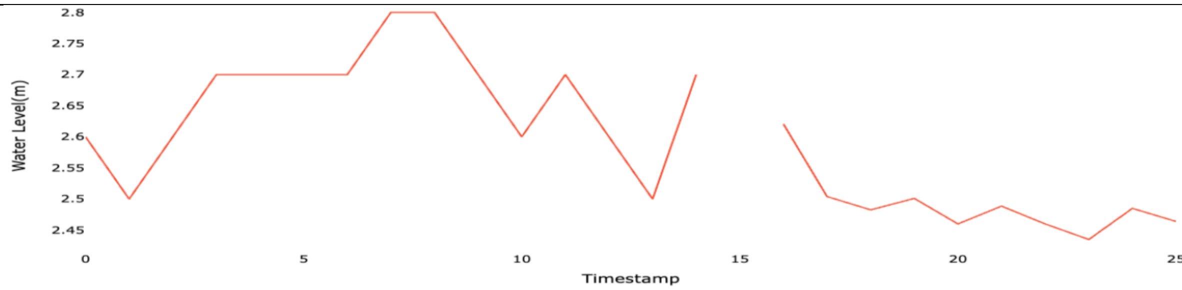


Figure 7b: Plotting last 15 days and next predicted 10 days of Sylhet by RF

Observations: The peak is observed (Figure 7b) around hour 6, with the water level reaching a maximum at about 2.8 meters. Between hours 7 and 12, the water level shows fluctuations with rapid rises and falls.

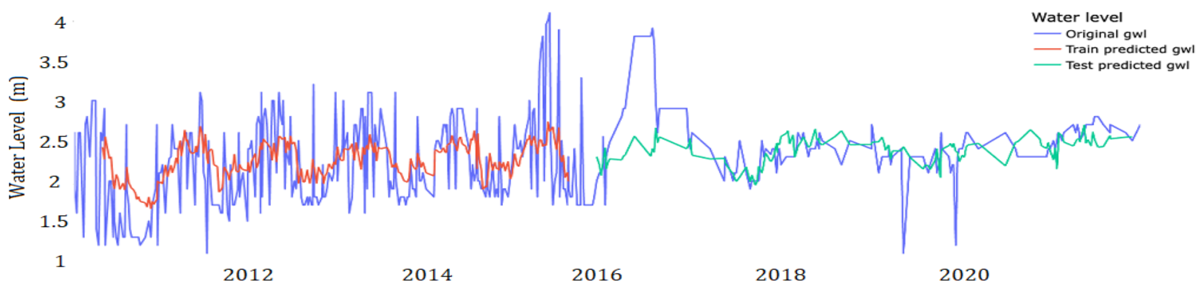


Figure 8: Comparison between original GWL vs predicted GWL with chart of Sylhet zone by KNN

Observations: It follows (Figure 8) the overall fashion of the authentic statistics however seems barely smoother and much less volatile. After 2016, each the real and anticipated values display decreased variability, suggesting a greater solid groundwater stage fashion. The anticipated groundwater ranges fairly align with the authentic ranges, indicating an awesome in shape of the prediction model.

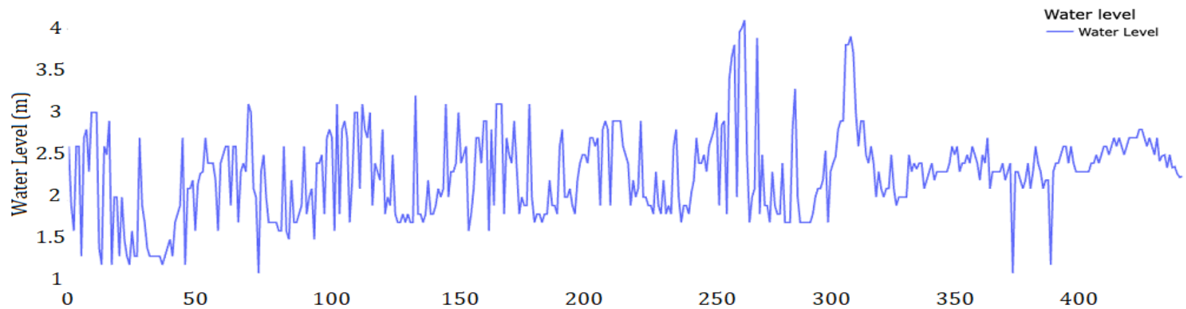


Figure 9: Plotting whole GWL with the next 10 days prediction, Sylhet zone by KNN

Observations: Water levels fluctuate (Figure 9) greatly over time, with no consistent trend of increasing or decreasing. Certain periods show less fluctuation in water levels than other periods and are relatively stable. Between 200 and 300 degrees, water levels show large fluctuations, including rapid rises and falls, with several peaks of over 3.5 meters. Above 400 degrees, water levels seem to have a slight tendency to decrease.

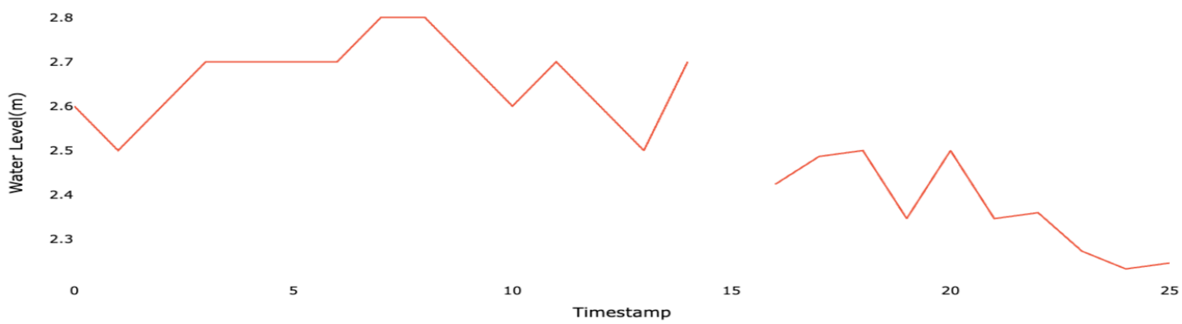


Figure 9a: Plotting last 15 days and next predicted 10 days of Sylhet by KNN

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

Observations: The first section (0 to 11) shows (Figure 9a) frequent short-term changes (rapid rises and falls), while the second section (15 to 25) shows the rapid changes. The water level drops more gradually.

LSTM Application

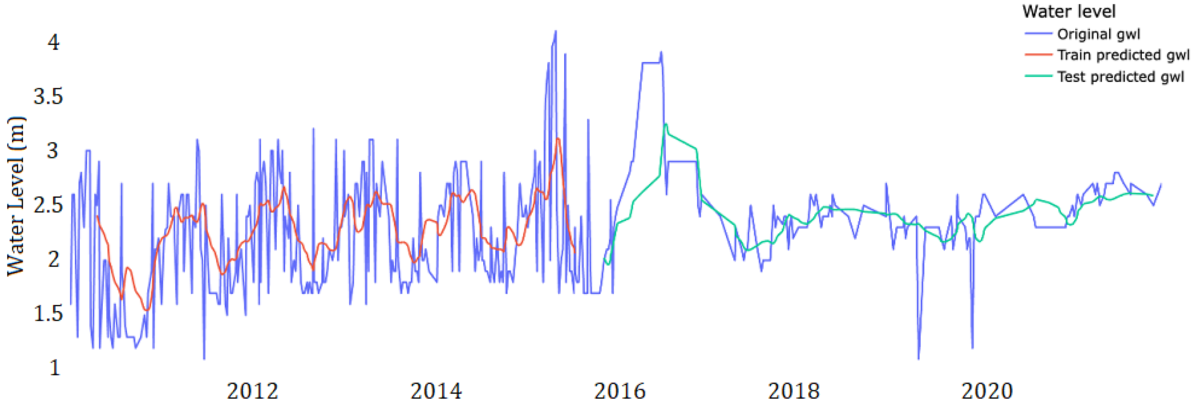


Figure 10: Comparison between original GWL vs predicted GWL with chart, Sylhet zone by LSTM

Observations: The model appears (Figure 10) reliable in capturing long-term trends in groundwater levels. Further analysis could focus on improving predictions during abrupt transitions. Examining seasonal variations in the original data can help understand recurring patterns and improve forecast accuracy.

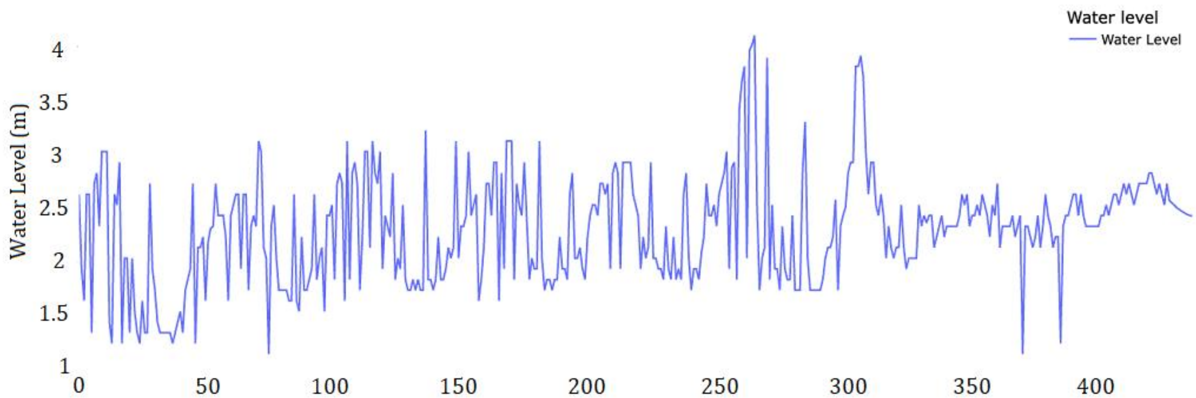


Figure 11: Plotting whole GWL with next 10 days prediction, Sylhet zone by LSTM

Observations: Constant fluctuations (Figure 11) in water levels suggest possible seasonal or environmental influences that could be investigated further; significant increases in water levels could indicate external factors such as heavy rainfall, flooding, or operational changes in the monitored system.

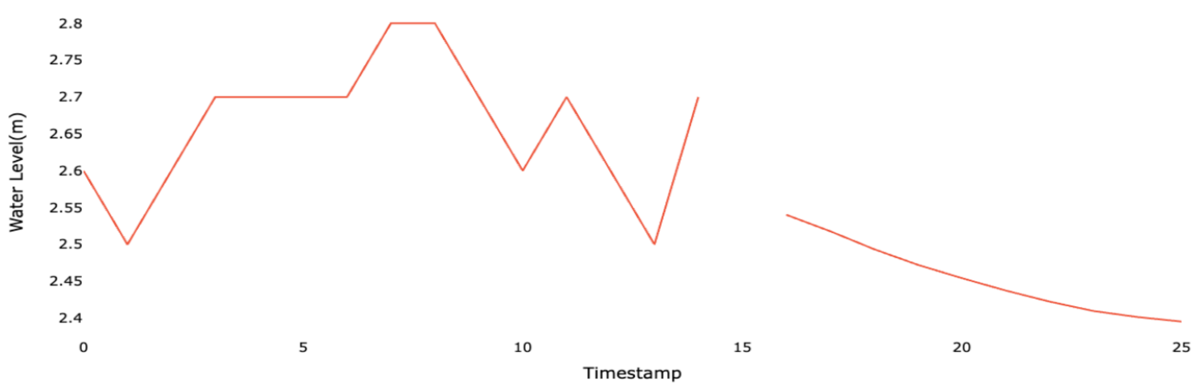


Figure 11a: Plotting last 15 days and next predicted 10 days of Sylhet by LSTM

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

Observations: Between timestamps (Figure 11a) three and 10, the water degree indicates sizeable fluctuations with peaks and valleys. After timestamp 10, the water degree progressively declines, forming a regular downward fashion till timestamp 25. The information is non-stop besides for the space among timestamps 15 and 20, wherein the trend maintains seamlessly.

GRU Applications:

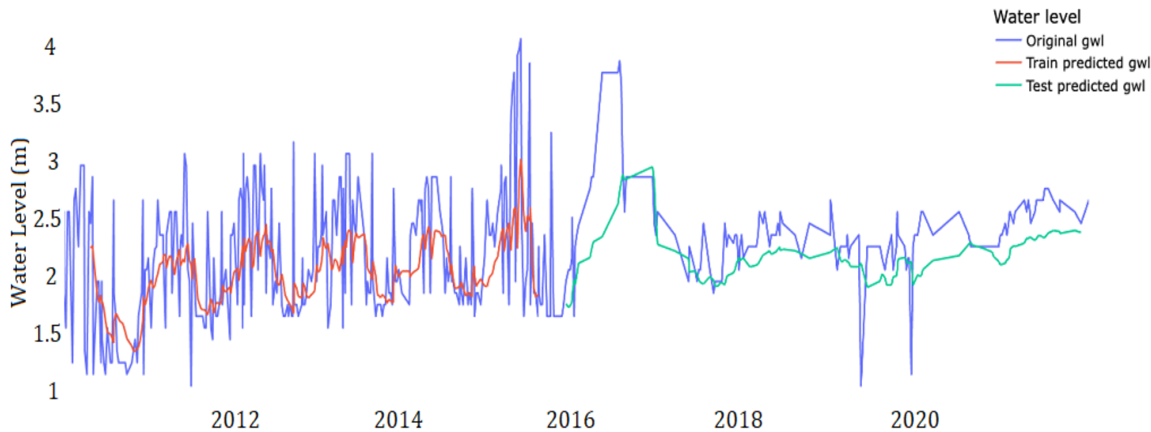


Figure 12: Comparison between original GWL vs predicted GWL with chart, Sylhet zone by GRU

Observations: The training predictions carefully observe (Figure 12) the unique records throughout the education period, efficaciously taking pictures of the general developments and fluctuations. The inexperienced line suggests a smoother pattern, efficaciously taking pictures of the overall developments inside the unique records, even though it lacks the acute fluctuations visible inside the real values.

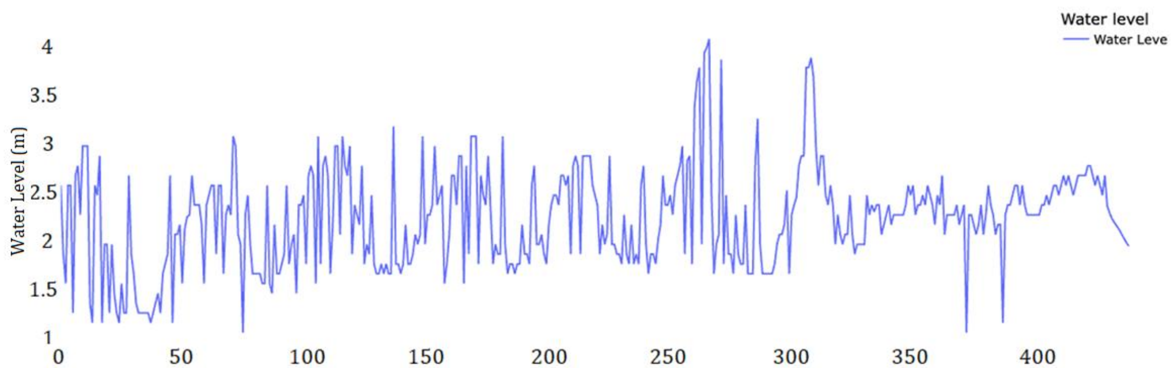


Figure 13: Plotting whole GWL with the next 10 days prediction, Sylhet zone by GRU

Observations: Water levels fluctuate (Figure 13) greatly over time, rising and falling frequently. Intermittent peaks occur, indicating periods of high-water levels. A slight decreasing trend is observed near the terminal index, indicating that the water level is gradually decreasing.

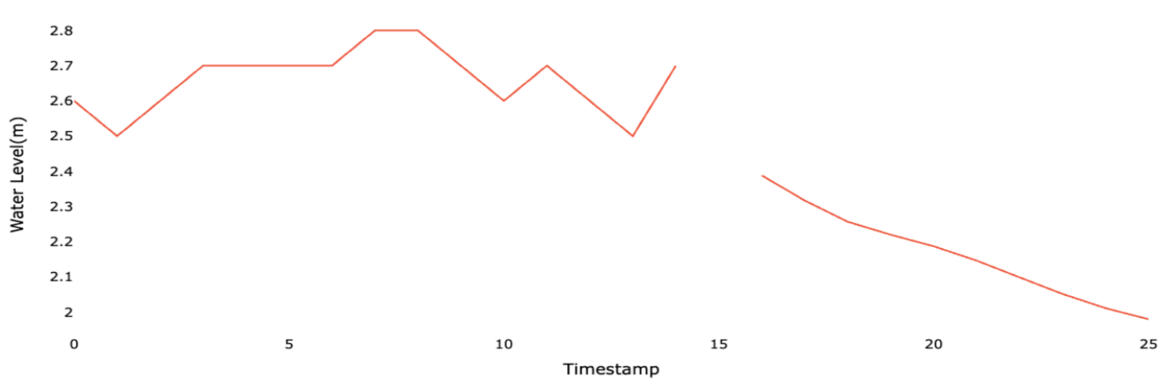


Figure 13a: Plotting last 15 days and next predicted 10 days of Sylhet by GRU

Observations: In the Figure 13a

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

LSTM + GRU Applications:

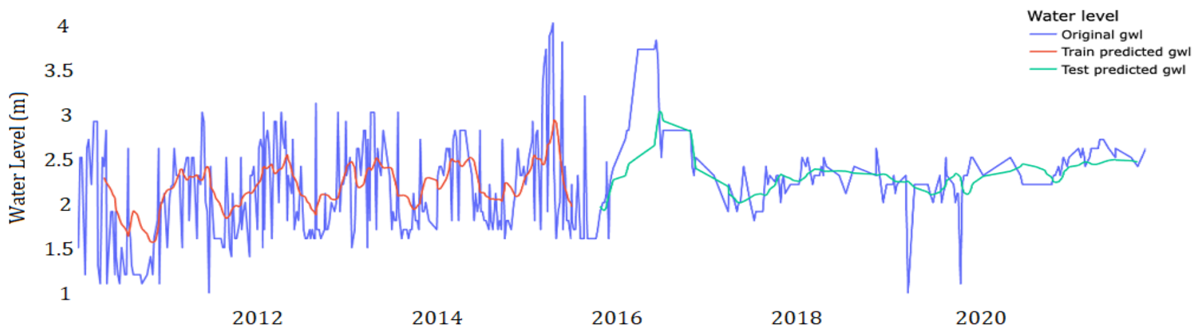


Figure 14: Plotting whole GWL with the next 10 days prediction, Sylhet zone by LSTM+GRU

Observations: The found spike (Figure 14) and dip occasions spotlight regions wherein the version may also gain from in additional refinement, along with incorporating outside elements like rainfall, recharge rates, or human activity.

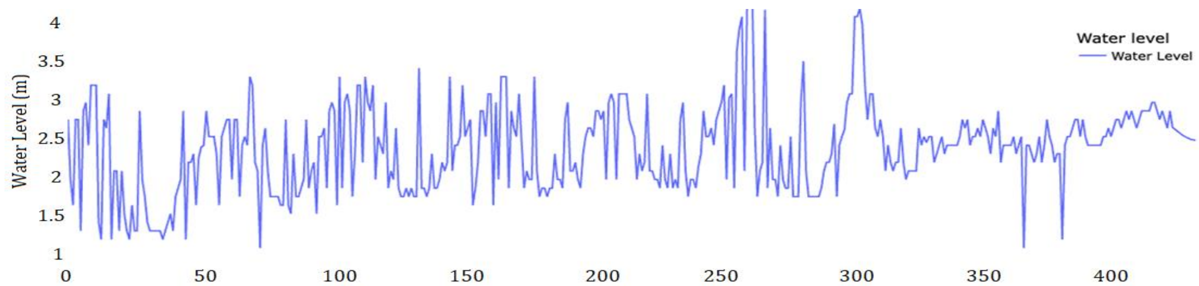


Figure 15: Plotting whole GWL with the next 10 days prediction, Sylhet zone by LSTM+GRU

Observations: The water stage demonstrates (Figure 15) full-size fluctuations in the course of the timeline. These versions may want to imply outside elements influencing the water stage, together with rainfall, inflow, or drainage events. Toward the cease of the timeline, the water stage seems to stabilize, with decreased fluctuation intensity.

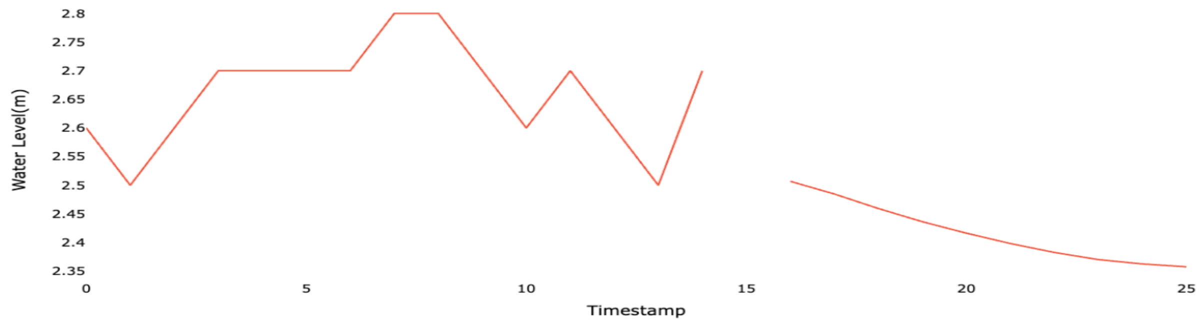


Figure 15a: Plotting last 15 days and next predicted 10 days of Sylhet by LSTM+GRU

Observations: The sharp rise (Figure 15a) and fall in the water levels during timestamps 0–10 could indicate variability in conditions affecting the water level. The steady decrease after timestamp 18 suggests a gradual drainage or lack of additional water input.

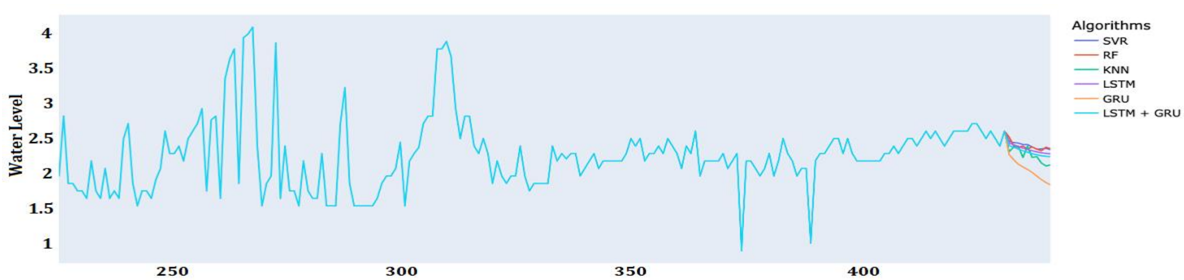


Figure 16: Plotting final chart with all algorithms and compare prediction to each other's

Observations: Predictive fashions (Figure 16) would possibly warfare on this vicinity because of the chaotic nature of the changes. The graph`s history has a mild blue shade, likely introduced for visible contrast, emphasizing the water degree range.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

Multivariate Time Series Forecasting for Groundwater Level, Rainfall, Temperature, Root and Surface Soil Witness, Depth to Groundwater level.

	waterLevel	temperature	humidity	rainfall	surface_soil_witness	root_soil_witness	profile_soil_moisture
0	1.4	16.40	8.79	0.0	0.73	0.68	0.71
1	1.5	16.38	8.18	0.0	0.70	0.66	0.69
2	2.6	18.68	9.03	0.0	0.69	0.65	0.67
3	2.6	18.87	8.12	0.0	0.66	0.62	0.66
4	2.6	16.80	7.69	0.0	0.65	0.61	0.63

Table 4: Multivariate dataset, Sylhet zone

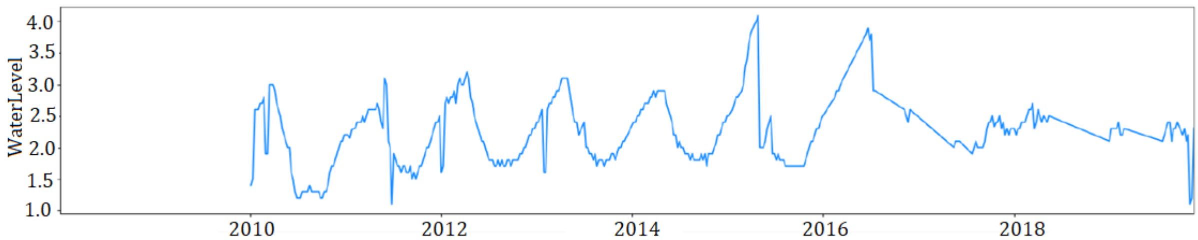


Figure 17: Multivariate analysis, Water Level, Sylhet zone

Observations: The maximum water degrees (Figure 17) are found round 2015–2016. This may want to suggest a duration of heavy rainfall, flooding, or modifications in water management. Water degrees display great fluctuations, with unexpected will increase and reduces in numerous years. After 2016, there's a slow decline withinside the common water degrees, indicating viable long-time period modifications withinside the system.

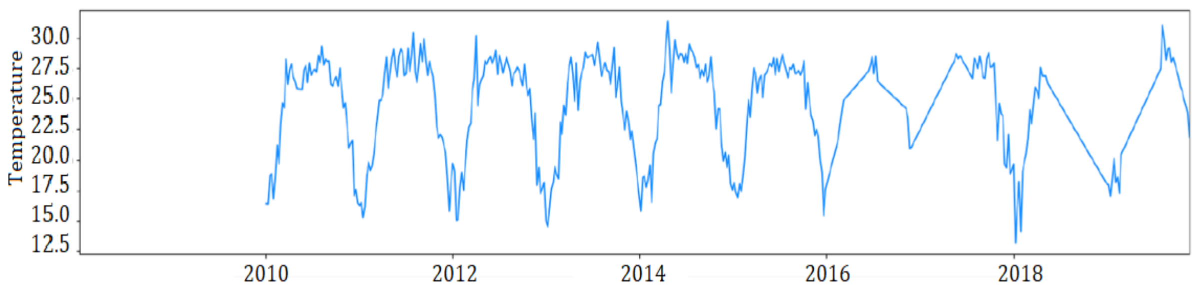


Figure 18: Multivariate analysis, Temperature, Sylhet zone

Observations**:** The sample of temperature modifications appears (Figure 18) regular over the years, suggesting strong seasonal dynamics. Towards 2018, there appears to be a moderate growth withinside the height temperature values as compared to in advance years, that could suggest warming traits or different environmental factors.

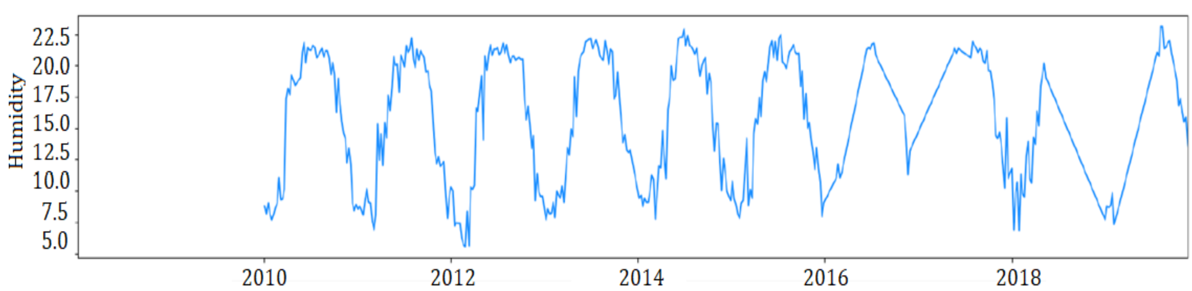


Figure 19: Multivariate analysis, Humidity, Sylhet zone

Observations: The seasonal peaks and troughs appear (Figure 19) regular throughout the years, indicating that the general climatic or environmental situations remained stable. Sharp dips and abnormal fluctuations are seen in a few years, in all likelihood because of severe climate events, modifications in information series methods, or neighborhood environmental factors.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

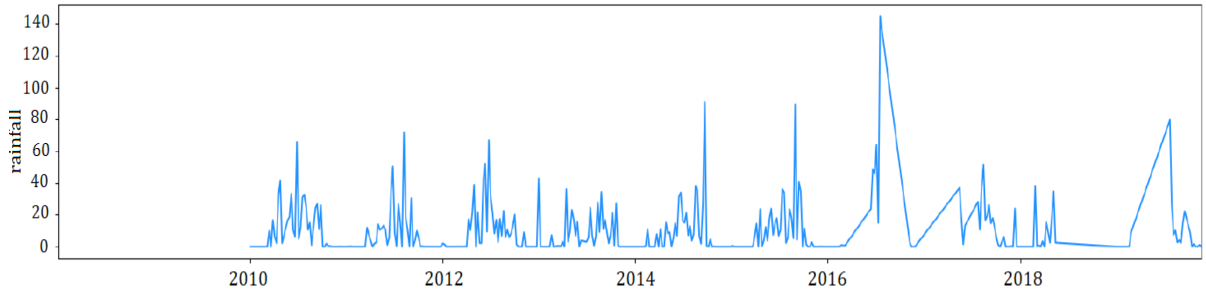


Figure 20: Multivariate analysis, rainfall, Sylhet zone

Observations: It appears to indicate (Figure 20) rainfall information over time, with a few sizeable peaks and valleys indicating intervals of better and decrease rainfall.

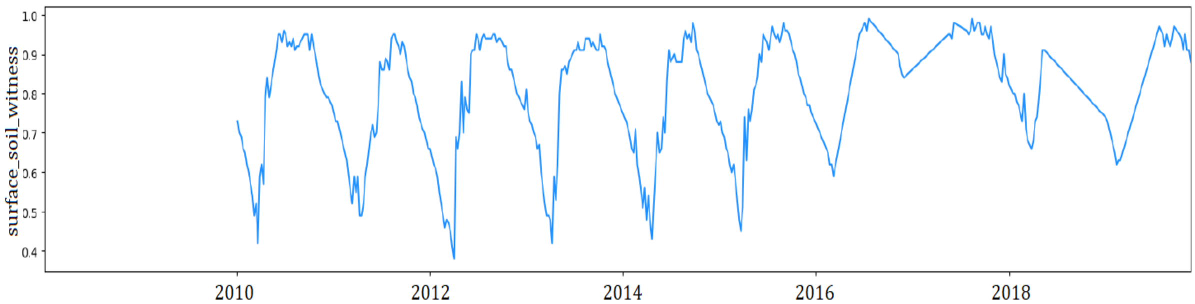


Figure 21: Multivariate Analysis, surface soil witness, Sylhet zone

Observations: The versions in wetness (Figure 21) appear constant over time, and not using an essential long-time period fashion as like as growing or reducing wetness. The peaks may correspond to wetter seasons refers, all through rainfall, whilst the troughs in all likelihood imply drier durations.

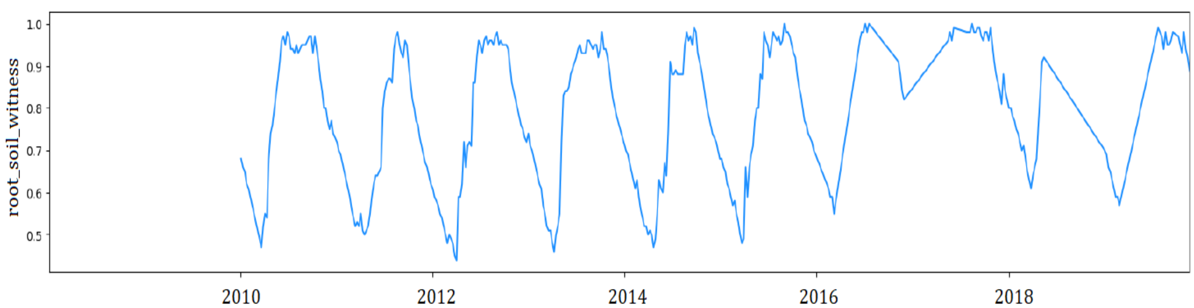


Figure 22: Multivariate Analysis, root soil witness, Sylhet zone

Observations: The photo appears (Figure 22) has a clean periodic pattern, with periodic rises and falls over time. While the general form of the cycles stays relatively dependable, there may be a few variant withinside the peaks and troughs. The plot covers almost a decade, signifying long-time period tracking of soil wetness.

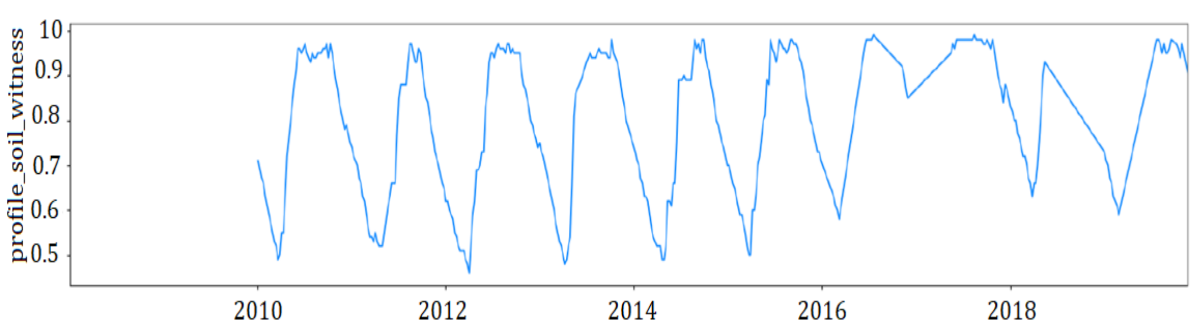


Figure 23: Multivariate analysis, profile soil witness, Sylhet zone

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

Observations: This plot displays (Figure 23) a clean habitual periodic cycle, with wetness peaking and declining in a predictable manner throughout years. The values continue to be dependably high; signifying is commonly greater stable.

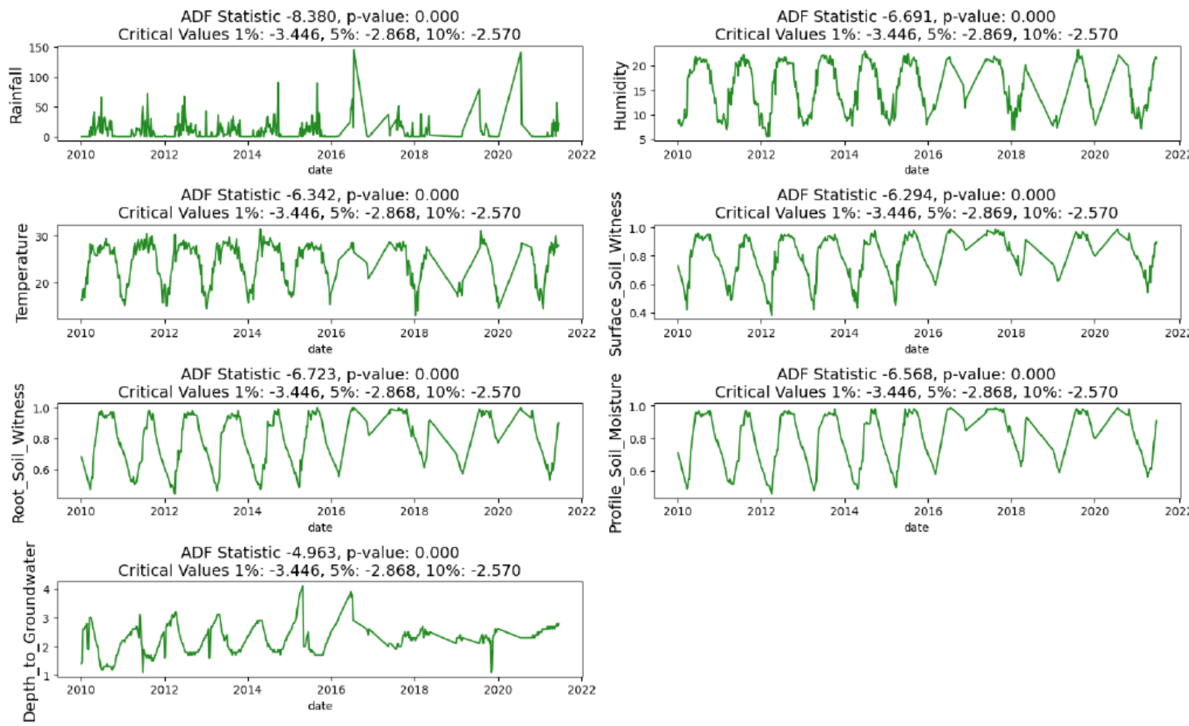


Figure 24: Multivariate analysis, Soil, Temperature, Humidity, Rainfall, Surface Soil witness, Sylhet zone

Observations: The graph displays (Figure24) time-collection information for numerous environmental variables (Rainfall, Humidity, Temperature, Surface Soil Wetness, Root Zone Soil Wetness, Profile Soil Moisture, and Depth to Groundwater) with related Augmented Dickey-Fuller (ADF) take a look at results, inclusive of statistics, p-values, and crucial values. Here are:

- **Trend and Seasonality:** Most variables showcase clean seasonal patterns (e.g., Humidity, Soil Wetness, and Profile Soil Moisture). Some variables, like Depth to Groundwater, show a long-time period growing trend, whilst others like Rainfall and Humidity showcase periodic spikes.
- **ADF Test Results:** The p-values for all variables are 0.000, indicating sturdy proof to reject the null speculation of non-stationarity. The ADF statistic is much less than the crucial values (1%, 5%, and 10% levels) for all collection, confirming that the information is desk bound or has been made desk bound.
- **Stationarity Confirmation:** Despite the presence of obvious developments and seasonality, the ADF take a look at shows the collection are desk bound, likely because of differencing or ameliorations implemented to the information.

Variable-Specific Observations:

- **Rainfall:** Characterized through excessive variability with irregular, sharp spikes. Most rainfall activities arise among 2010–2016, with fewer enormous spikes after 2016.
- **Humidity:** Displays a clean seasonal sample with periodic peaks and troughs, indicating cycles of moist and dry periods. The ADF statistic of -6.691 confirms sturdy stationarity.
- **Temperature:** Relatively stable, with minor seasonal fluctuations. ADF statistic of -6.342 shows the temperature collection is desk bound.
- **Surface Soil Wetness:** Shows wonderful seasonal conduct with everyday peaks and troughs. ADF statistic of -6.294 confirms stationarity.
- **Root Zone Soil Wetness:** Displays comparable seasonal cycles as floor soil wetness however with barely decrease magnitude. Stationarity is showed with an ADF statistic of -6.723.
- **Profile Soil Moisture:** Seasonal fluctuations are obvious and align carefully with floor and root region wetness developments. ADF statistic of -6.568 confirms stationarity.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

- **Depth to Groundwater:** Shows a long-time period upward trend, with periodic fluctuations superimposed. Despite the trend, the ADF statistic of -4.963 and p-price of 0.000 advocate the collection is desk bound after capacity ameliorations.

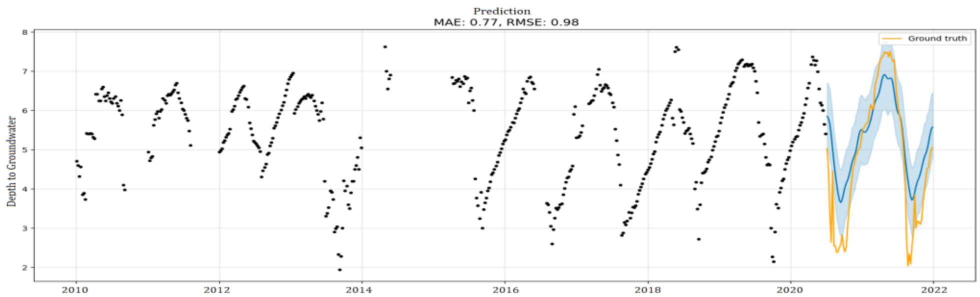


Figure 25: Prediction – MAE & RMSE, Depth to GWL analysis, Multivariate, Sylhet zone

Observations: In (Figure 25), MAE is 0.77 and RMSE is 0.98 values degree the accuracy of forecasts, with decrease values being better. Model seems to carry out nicely withinside the later part of the dataset 2020 onwards because the forecasts align strictly with the floor reality data.

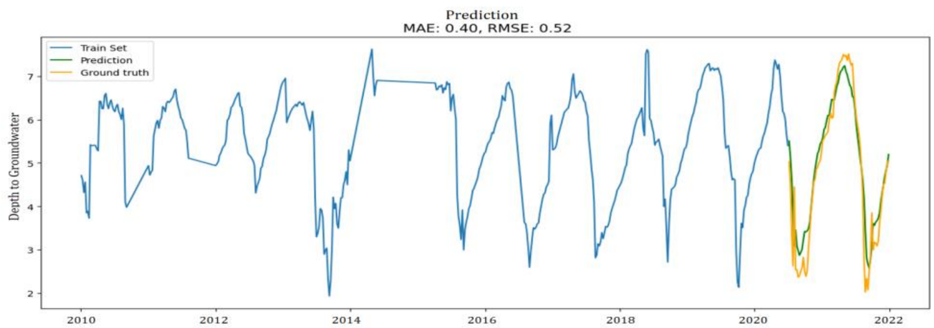


Figure 26: Time Series: FB Prophet Model Output, MAE, RMSE, Multivariate, Sylhet zone

Observations: Model Performance (Figure 26) MAE is 0.40 and RMSE is 0.52, metrics specify a perfection as compared to the primary image, with decrease mistakes values viewing that the version forecasts higher. The schooling records is explicitly pictured here, offering higher context for the version`s getting to know phase. The overall performance metrics, MAE and RMSE are substantially higher on this case, in all likelihood because of better records dealing with or version enhancements.

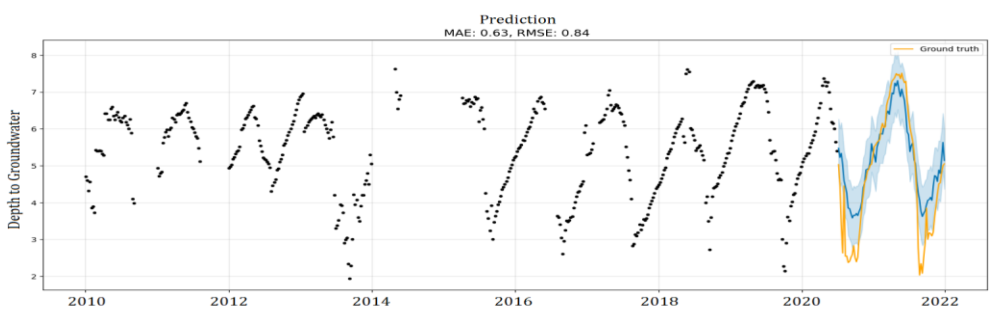


Figure 27 Multivariate Time Series Analysis: FB Prophet Model Output, Sylhet zone

***Observations*:** Here version overall performance (Figure 27) MAE is 0.63 and RMSE is 0.84, values are among the metrics visible inside the first plots, displaying higher overall performance related to the primary plot however barely worse than the second one. This plot makes a specialty of forecasts with experimental data, without showing the education set explicitly. The version's overall performance lies among the 2 earlier cases, signifying this could mirror a unique form of the version or adjustments in its parameters.

IV. Findings & Recommendations

A. Major Findings

- **Model Performance:** The DL fashions, in the main LSTM and GRU, outperformed conventional ML algorithms in taking pictures long-time period dependances and nonlinear shapes in each univariate and multivariate time collection datasets. The ML algorithms, including SVR, RF, and KNN, supplied strong overall performance on minor datasets and whilst foretelling less complicated patterns, however their correctness declined with growing records intricacy. Multivariate fashions leveraging DL

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

showed advanced analytical accuracy related to univariate fashions, as they correctly implemented the family members among a couple of variables.

- **Algorithm Suitability:** SVR imparts regular outcomes for linear and fairly nonlinear datasets however struggled with high-dimensional and multifaceted temporal family members. RF exhibited robustness in coping with noisy records and apprehending characteristic status, however, it lacked the capacity to version sequential sequences correctly. KNN even as easy and green for positive datasets, changed into penetrating to the top-quality of hyperparameters and lacked scalability for better datasets. LSTM and GRU were established to be the most real for time collection foretelling because of their potential to hold incidental proof and version temporal dependences. GRU, in precise, confirmed computational returns over LSTM for datasets with shorter arrangements.
- **Hybrid Approaches:** The combination of ML and DL fashions progressed the general accuracy and electricity of predictions. Hybrid strategies had been in the main real for multivariate prediction, wherein characteristic interactions required innovative examples and modeling performances.
- **Evaluation Metrics:** MAE, RMSE, and MAPE discovered that deep mastering fashions regularly condensed analytical mistakes related to standard ML approaches, in the main in complicated scenarios.

B. Major recommendations

- **Algorithm Selection:** For univariate time collection forecasting, LSTM and GRU are cautioned for apprehending long-time period dependencies, even as RF and SVR may be castoff for less complicated datasets or whilst computational sources are limited. For multivariate forecasting, DL fashions need to be organized because of their capacity to deal with complicated variable interactions.
- **Data Preprocessing:** Correct normalization, characteristic range, and elimination of outliers are vital to development the overall performance of each ML and DL fashions. For multivariate records, characteristic engineering strategies including correlation look at ought to be operating to perceive the maximum enormous variables.
- **Hybrid Modeling:** Leveraging hybrid fashions that integrate the strengths of ML and DL is usually recommended for top-quality overall performance, in the main in multifaceted or high-dimensional datasets.
- **Computational Efficiency:** For packages with restricted computational assets or real-time constraints, GRU ought to be preferred over LSTM because of its decrease computational necessities and equal overall performance. RF and KNN can function baseline fashions for speedy prototyping earlier than transitioning to extra computationally focused deep gaining knowledge of fashions.
- **Future Research:** Further exploration of ensemble techniques that integrate several ML and DL fashions is usually recommended to decorate robustness and simplification. The integration of interest mechanisms into DL fashions may want to similarly develop predicting correctness with the aid of using electively concentrating at the maximum suitable time steps or structures. The software of those performances to domain-precise datasets ought to be explored to validate their overall performance in real-global circumstances.

V. Results

The determination of show and execution valuation measurements ought to be unexpected on the highlights of the dataset and the exact necessities of the issue. For time arrangement evaluating, DL models like LSTM and GRU are regularly more agent for capturing chronological designs, whereas conventional ML models like SVR, RF, and KNN might work well when the relationship between inputs and yields isn't intensely time-dependent.

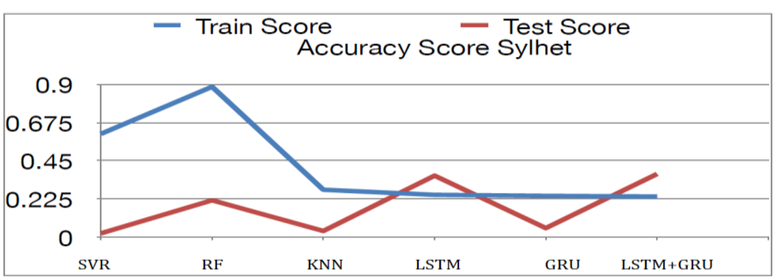


Figure 28: Accuracy Score Sylhet zone

A. Accuracy Score

Observations: The given chart compares (Figure 28) the preparation and test exactness scores for distinctive models connected to the Sylhet dataset.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

• **Overfitting in RF:** The Irregular Timberland show appears clear signs of overfitting, with a huge dissimilarity between preparation and test scores.

• **Superior Generalization in LSTM+GRU:** The combined LSTM+GRU performs superior on the test set, demonstrating that combining these two structures may adjust demonstrate complexity and generalization.

• **Underperformance in SVR and KNN:** These models underfit the information, as reflected by their reliable moo prepare and test scores.

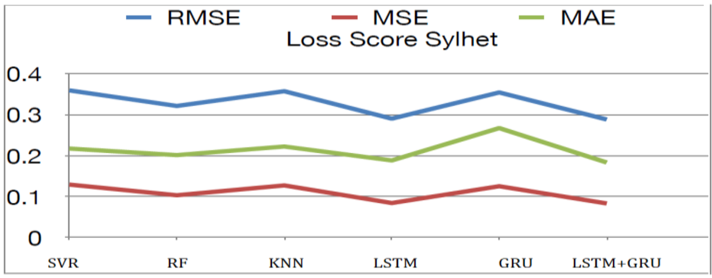


Figure 29: Loss Score Sylhet zone

B. **Loss Score**

Observations: The given chart shows (Figure 29) the misfortune measurements (RMSE, MSE, MAE) for different models on the Sylhet dataset.

Best Execution: The LSTM+GRU combined demonstrate beats others in terms of minimizing misfortune measurements (RMSE, MSE, MAE).

GRU Confinement: GRU alone battles, particularly with MAE, demonstrating that it might not be taking care of littler varieties within the information as well as LSTM or the combined approach.

SVR and KNN Underperformance: These fewer complex models result in reliably higher misfortune values, making them less appropriate for this dataset.

C. **Accuracy Score Heatmap**

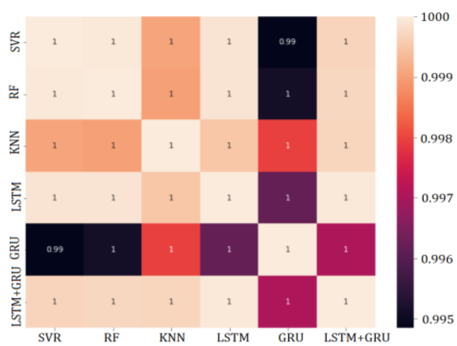


Figure 30: Accuracy Heatmap of Sylhet zone

Observations: The transferred picture (Figure 30) may be a heatmap speaking to the relationship or comparison between distinctive models. The cells contain values, likely speaking to relationship coefficients or execution measurements. Most values are near to 1, implying tall likeness or execution among the models. The as it were deviation shows up between SVR and GRU, where the esteem is 0.99, marginally lower than the others. The heatmap employments a slope color scale from dull (lower esteem) to shinning (higher esteem). The values are concentrated close the 1.0 check, reflecting a solid execution or comparison over all models. All models appear to perform exceptionally additionally, with near-perfect comes about. The pairing of SVR and GRU appears the foremost particular change but remains amazingly near to 1.

VI. Conclusion

The studies efficaciously demonstrated how an aggregate of Machine Learning (ML) and Deep Learning (DL) algorithms can cope with the encounters of time collection predicting. While conventional fashions like SVR and RF stay associated for less difficult tasks, superior methods like LSTM and GRU have set a new benchmark for comparing and forecasting complicated temporal relations, in particular in multivariate environments. The effects emphasize the significance of aligning the selection of logical techniques with the functions of the records and the predicting purposes. The incorporation of hybrid techniques that integrate the strengths of ML and DL

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIII, Issue XII, December 2024

has showed actual in improving version performance, specifically in occasions with excessive records complication. Assessment metrics along with MAE, RMSE, and MAPE make certain the better correctness of DL fashions in time collection predicting. This observe highlights the significance of indicating predicting fashions primarily based totally at the complexity of the records and the precise software requirements. While ML techniques stay valued for his or her simplicity and efficiency, DL fashions, specifically LSTM and GRU, are vital for tackling the demanding situations of present-day time collection predicting. Future observes have to emphasis on refining hybrid fashions, collaborating attention mechanisms, and smearing those performances to domain-precise troubles to in addition decorate their applicability and efficiency. By accepting the right methodologies, groups can resolve the entire capability of systematic analytics to make records-pushed picks with confidence.

References

1. Abdollahi, A., Bazrafshan, J., & Razmjooy, N. (2020). A deep learning method for temperature forecasting based on Bi-LSTM. *Applied Soft Computing*, 97, 106782.
2. Acharyya, A., (2017), Sustainable Development of Groundwater Resources in India: A Relook at Policy Initiatives, Vol-VIII.
3. Adhikari, U., & Ikeda, M. (2020). XGBoost and LSTM model for multivariate time series forecasting of groundwater level. Water, 12(11), 3082. DOI.
4. Aishwarya, R., & Vasudevan, V. (2023). A comprehensive review of multi-source data integration for groundwater modeling. Journal of Hydrology, 617, 128812. DOI.
5. Cheng, L., Li, G., & Castelletti, A. (2020). GWL forecasting with LSTM networks: A study in agricultural water management. Hydrological Sciences Journal, 65(9), 1505-1516.
6. Fawaz, H. I., Forestier, G., Weber, J., et al. (2019). Deep learning for time series classification: A review. Data Mining and Knowledge Discovery, 33(4), 917-963.
7. Gharbi, S., & Bouaziz, M. (2023). Real-time data assimilation for GWL prediction using machine learning. Water Resources Management, 37(5), 1461-1478. DOI.
8. Karthikeyan, L., Khosa, R., & Singh, V. P. (2020). Deep learning models for soil moisture retrieval from remote sensing data: A review. Environmental Earth Sciences, 79(7), 193.
9. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774).
10. Lim, B., & Zohren, S. (2021). Time series forecasting with deep learning: A survey. Philosophical Transactions Royal Society: Mathematical, Physical and Engineering Sciences, 379(2194), 20200209.
11. Mojid, M.A., Parvez, M.F., Mainuddin, M. and Hodgson, G., (2019), Water Table Trend- A Sustainability of GWL Development in North-West Bangladesh, Water, Vol 11, pp-1182.