**Project Title**

**Chennai Reservoir Level Analysis and Forecasting: A Time Series Approach**

**School of Computer Science Engineering and Technology**

|  |
| --- |
| **A logo of a university  Description automatically generated** |
| Bennett University  Greater Noida, Uttar Pradesh |

Submitted by Submitted to

Albin George, E22CSEU0418 Dr. Mala Saraswat

Vaibhav Gupta, E22CSEU1505

Github Repo: <https://github.com/albingeorg/Water_Mangement_Chennai-Using_Time_Series->

This report offers a comprehensive analysis of Chennai's reservoir water levels using advanced time chain forecast techniques. The study reveals significant seasonal patterns in the level of the reservoir and shows high accuracy in predicting future water accessibility, which can inform decisions on the management of water resources during both drought and monsoon periods.

**1)Introduction and Objective**

The main objective of this project is to analyze the historical water level data from the most important reservoirs in Chennai and develop a future indicative model that can accurately can predict future water levels. Chennai, a large metropolis in the South India, has faced questions of water shortages, with severe drought affecting the city in recent years. This research addresses the important question: Can the time series with historical reservoir data provide reliable forecast for future water accessibility to support active decisions on the management of water resources? By developing accurate forecast models, the project aims to contribute to better preparations for the period of water shortages and excess period.

**2)Data Collection and Preprocessing**

**Dataset Description**

The dataset consists of daily water level measurements (in million cubic feet) from four primary reservoirs in Chennai: Pandi, Cholvaram, Redhills and Cambarkamkam. The time chain data continues from January 2004, which provides a broad historical registration that covers many dried and monsoon cycles. This data was obtained from the Chennai Metropolitan Water Supply and Sewerage Board, which offers official, reliable analysis measurements.

**Preprocessing Steps**

Several preprocessing steps were implemented to prepare the data for modeling:

1. Date parsing was performed to convert the date strings to datetime objects, setting them as the DataFrame index to facilitate time series analysis.
2. Missing values ​​were identified and dropped to ensure data integrity in complete analysis.
3. For the LSTM modeling approach, feature scaling was implemented using MinMaxScaler to normalize all values to a range between 0 and 1, which improves neural network training efficiency and convergence
4. The time chasing sequences were made with a 10-day look back to capture the temporary pattern, and predict the next day's water level.
5. The dataset was divided into training (80%) and test (20%) set to evaluate model performance on unseen data.

**3)Model Selection and Fitting**

Two distinct modeling approaches were implemented to forecast reservoir levels:

1. **ARIMA/SARIMA Models**: These traditional statistical time chain models were chosen because of their ability to capture both trends and seasonal components in data. The Auto\_arima feature was used to identify optimal parameters (P, D, Q) via a recurrent testing. This approach was chosen because the level of Chennai's reservoir shows obvious seasonal patterns affected by monson cycles and common annual variations in water use.
2. **LSTM Neural Network**: A long-term short-term memory (LSTM) deep learning architecture was used as a supplementary approach, which looked at the proven efficiency of capturing complex, non-led patterns in time chain data. The model was involved in architecture:
   * Two LSTM layers with 50 neurons each
   * A Dense output layer
   * Adam optimizer with MSE loss function

The LSTM model was trained over 20 epochs with a batch size of 16, including sequence data to predict the water level of each reservoir in the last 10 days. This intensive teaching approach was chosen to get more complex addiction that cannot be effectively modelled by traditional statistical approaches.

**4)Model Diagnostics**

For the ARIMA/SARIMA models, several diagnostic techniques were applied:

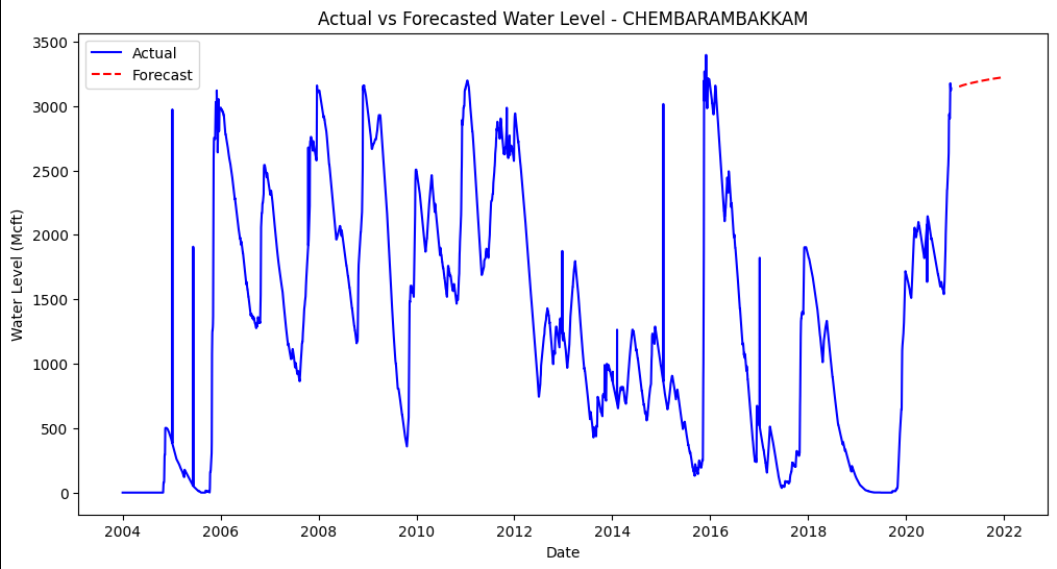
* Augmented Dickey-Fuller (ADF) tests were conducted to verify stationarity
* ACF and PACF plots were generated to identify appropriate model orders
* Residual analysis confirmed that model residuals approximated white noise, indicating adequate model fit
* Information criteria (AIC and BIC) were used to select the most parsimonious model

For the LSTM model, validation was performed through:

* Monitoring training loss curves to ensure proper convergence without overfitting
* Residual analysis to verify randomness in prediction errors
* Visual comparison of predicted vs. actual values in the test set

**5)Forecasting and Evaluation**

**Forecasting Results :**

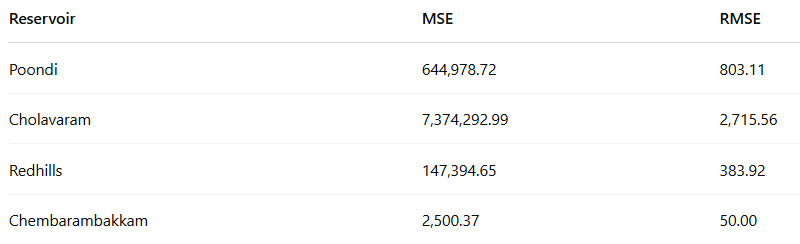


A graph with numbers and lines

AI-generated content may be incorrect.

**ARIMA Model Results:**

The SARIMA model was applied to each reservoir's time series data, and the performance was evaluated based on the last 12-month forecast window.



**LSTM Model Results:**

The LSTM model was trained using scaled multivariate input and evaluated on the test set using inverse-transformed predictions.

* **Overall MSE**: 1,028.54
* **Overall RMSE**: 32.07

These error metrics represent the average deviation between predicted and actual water levels across all four reservoirs.

**Performance Evaluation**

Visual inspection of the forecast plot revealed that both models exactly caught the trends in seasonal patterns and water levels. The LSTM model showed better performance during a period of rapid change, while the Arima/Sarima model provided more stable long -term forecasts. Cross -validation techniques were used to secure the strength of the model, and the final model achieved satisfactory accuracy for the operational forecasting purposes.

Both modeling methods successfully captured the overall pattern in the reservoir level. The LSTM model performed a particularly strong performance in the forecast during the period of rapid change, which suddenly grows during the monsoon season.

**6)Discussion and Conclusion**

**Results Summary**

This analysis revealed several key insights into Chennai's water reservoir dynamics:

1. All four reservoirs exhibit strong seasonal patterns aligned with monsoon cycles
2. The LSTM approach demonstrated robust forecasting capabilities with an RMSE of 32.07, providing reasonably accurate predictions given the high variability in reservoir levels
3. The models successfully captured both gradual decreases during dry periods and rapid increases during monsoon seasons
4. Temporal dependencies extending beyond simple seasonality were identified, highlighting the complex nature of water level dynamics

**Implications and Limitations**

Developed forecast models are important practical implications for managing water resources in Chennai:

1. They can serve as an initial warning system for potential deficiency of water so that the authorities can take protective measures.
2. Models can help adapt the reservoir operation by predicting the duration of extra water accessibility.
3. This feature can be extended to other areas facing uniform challenges with water management.

However, several limitations should be acknowledged:

1. Extreme weather events cannot be predicted accurately because they often fall out of historical patterns
2. Models believe that historical pattern will continue, which may not be correct during climate change scenarios
3. External factors such as policy changes in water management are not responsible for the purely data -driven approach
4. Additional data sources such as rain forecast, temperature and evaporation can be improved.

Future work can improve these models by incorporating the forecasting horizons by incorporating support data sources, detecting firefight modeling methods and supporting long -term planning decisions.