Department of Artificial Intelligence and Data Sciences

Major Project
on
Mitigating Traffic Congestion through Time Series Analysis and Deep Learning

Submitted By:

Shweta Choudhary 05001172020

Ekta Gupta 07001172020

Bhumika Mahor 07101172020

B.Tech. CSE-AI

Supervisor :

Ms. Ritika Kumari

Assistant Professor

AI & DS Dept.

INDEX:-

INTRODUCTION :-

- Traffic congestion is a widespread issue in urban areas worldwide, leading to increased travel times, fuel consumption, and environmental pollution.

- In this project, we aim to explore the application of time series analysis and deep learning models, including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and other deep learning architectures, for traffic congestion control.

- By analyzing temporal traffic data and developing predictive models, we seek to identify congestion hotspots, forecast traffic flow patterns, and implement proactive measures to alleviate congestion and improve overall traffic efficiency.

LITERATURE REVIEW AND RESEARCH GAPS

S.NO.	AUTHOR	TITLE	YEAR	FINDINGS	RESEARCH GAPS
1.	Lokesh Chandra Das	Traffic Volume Prediction using Memory-Based Recurrent Neural Networks: A comparative analysis of LSTM and GRU	2023	The research paper compares LSTM and GRU models for traffic volume prediction using memory-based recurrent neural networks.	involves the implementation and evaluation of LSTM and GRU models on traffic volume prediction tasks using historical traffic data.
2.	Zheng, H., Lin, F., Feng, X., & Chen, Y.	Hybrid Deep Learning Model With Attention-Based Conv-LSTM Networks for Short-Term Traffic Flow Prediction	2020	it introduces an attention mechanism that helps in better modeling spatial dependencies in traffic data.	the need for more efficient hybrid deep learning models that effectively combine different architectures for short-term traffic flow prediction.
3.	AS Mihaita, H Li, and MA Rizoiu.	Traffic congestion anomaly detection and prediction using deep learning.	2020	the use of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) with respect to predicting the identification of abnormal traffic forecasting through their method.	regarding the anomaly detection and prediction of traffic congestion using deep learning methods, indicating a need for more advanced techniques in this area.

LITERATURE REVIEW AND RESEARCH GAPS

S.NO.	AUTHOR	TITLE	YEAR	FINDINGS	RESEARCH GAPS
4.	Zhao et al.	Traffic flow prediction using LSTM-based deep learning	2019	put forward the use of Long Short-Term Memory (LSTM) recurrent neural networks in capturing temporal dependencies present in traffic data and further proving the efficiency of the LSTM network for its suitable use in making accurate predictions about the dynamism of flow in traffic.	improvements in traffic flow prediction using LSTM-based deep learning methods, indicating potential gaps in the accuracy or efficiency of existing models.
5.	Jain, A. K., & Kumar, A. M.	Hybrid approach using ARIMA and artificial neural networks for predicting traffic situations	2018	introducing an optimization routine that enhances predictive performance and robustness.	utilizing ARIMA and artificial neural networks for predicting traffic situations, suggesting further exploration into optimizing the combination of these techniques.

RESEARCH OBJECTIVES :-

- Our objective is to utilize temporal traffic data to understand temporal patterns and trends in traffic congestion with the help of EDA.
- Then we proceed to Developing LSTM, GRU, SARIMA and other hybrid deep learning models to forecast traffic flow and predict congestion events. Identifying congestion hotspots and critical time periods using predictive analytics.
- Evaluating the performance of the developed models in terms of prediction accuracy, congestion detection, and effectiveness of congestion control strategies.
- This would help government authority for implementing proactive measures such as traffic signal optimization, route planning, and congestion pricing to alleviate traffic congestion.
- Providing actionable insights and recommendations for policymakers, transportation authorities, and urban planners to improve traffic management and enhance overall transportation efficiency.

Proposed Methodology :-

1. Data Collection
2. Exploratory Data Analysis
3. Feature Engineering
4. Deep Learning Model Building:

    4.1) LSTM
    4.2) GRU
    4.3) Stacked LSTM
    4.4) SARIMA
    4.5) Hybrid model (SARIMA + Stacked LSTM)
    4.6) Ensemble model (SARIMA + Stacked LSTM)

5. Results

    5.1) Individual Model
    5.2) Composite Model

6. Mitigation Techniques
7. Conclusion
8. Future Scope

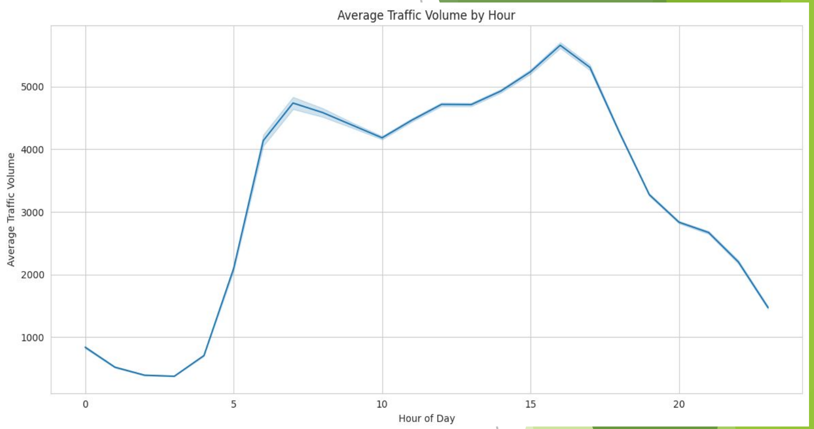
1. Data Collection :-

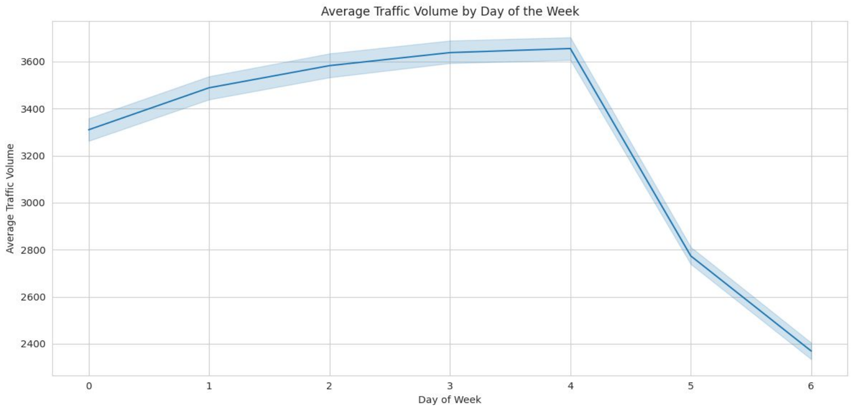
- The dataset contains 48,204 entries. It has 9 columns:
  'Holiday', 'temp', 'rain_1h', 'snow_1h', 'clouds_all', 'weather_main', 'weather_description', 'date_time', 'traffic_volume'.
- The dataset contains a mix of data types:
  - 4 object-type columns ('holiday', 'weather_main', 'weather_description', 'date_time').
  - 3 float64-type columns ('temp', 'rain_1h', 'snow_1h').
  - 2 int64-type columns ('clouds_all', 'traffic_volume').
- 'holiday' column: Missing values exist, with only 61 non-null entries.
- 'date_time' column: Provides temporal information for each observation.
- Dataset: Spans various dates and times, with each entry corresponding to a specific time interval.
- Summary statistics: Provided for numerical columns, including count, mean, standard deviation, minimum, 25th percentile, median, 75th percentile, and maximum values.

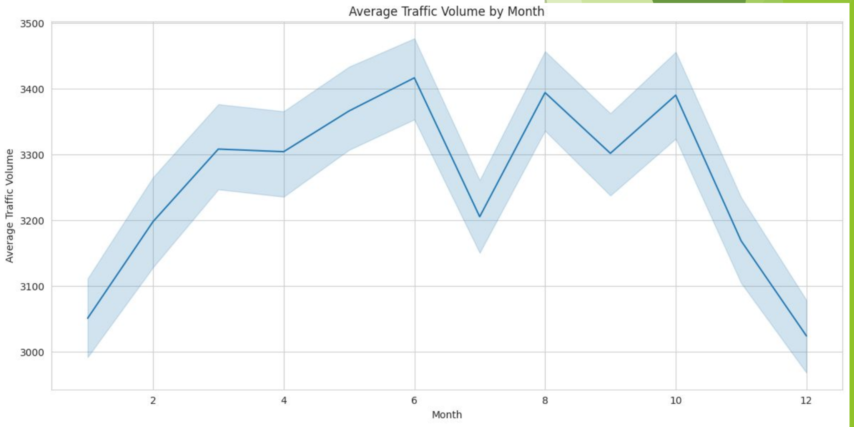
2. Exploratory Data Analysis :-

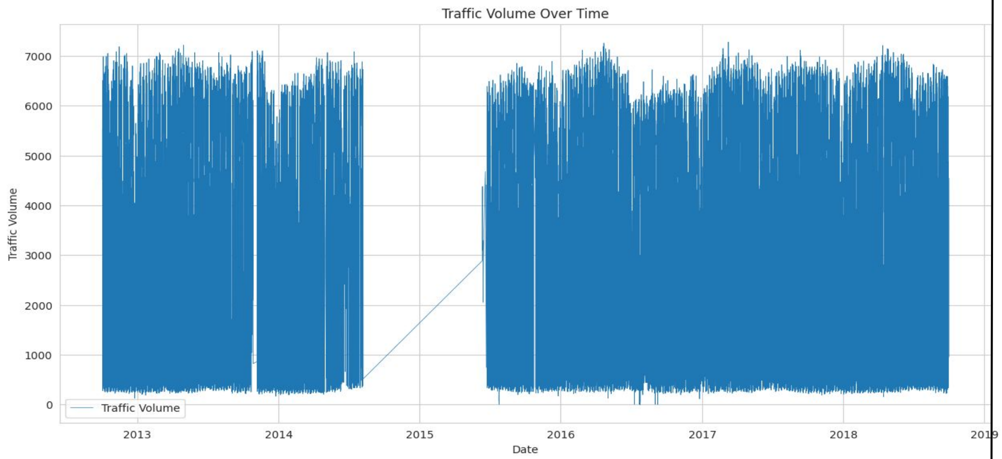
- **Objective of EDA:**
  - Identify patterns, anomalies, and relationships in the traffic data.
  - Establish foundational insights to guide further modeling efforts.
- **Temporal Dynamics Analysis:**
  - Hourly Trends: Identified peak traffic volumes during typical rush hours (7 AM to 9 AM and 4 PM to 6 PM).
  - Weekly Trends: Traffic progressively increases from Monday to Friday with a significant drop over the weekends.
  - Monthly Trends: Higher traffic volumes mid-year with drops in December, influenced by holidays and seasonal changes.
- **Impact of External Factors:**
  - Analyzed traffic under varying weather conditions and during holidays.
  - Major holidays like Christmas and New Year’s Day show considerable decreases in traffic volumes.

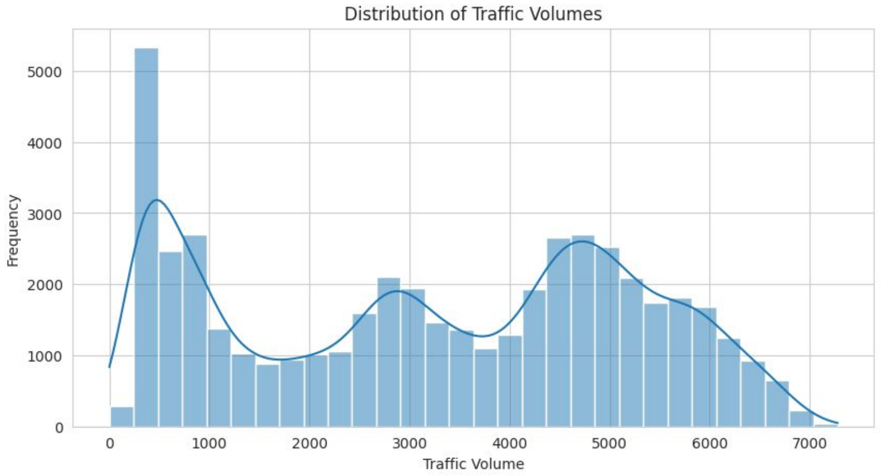
- **Key Findings:**
  - The need for dynamic traffic management strategies tailored to time-specific and seasonal variations in traffic volume.
  - Importance of incorporating weather and holiday impacts into traffic models for more accurate predictions.

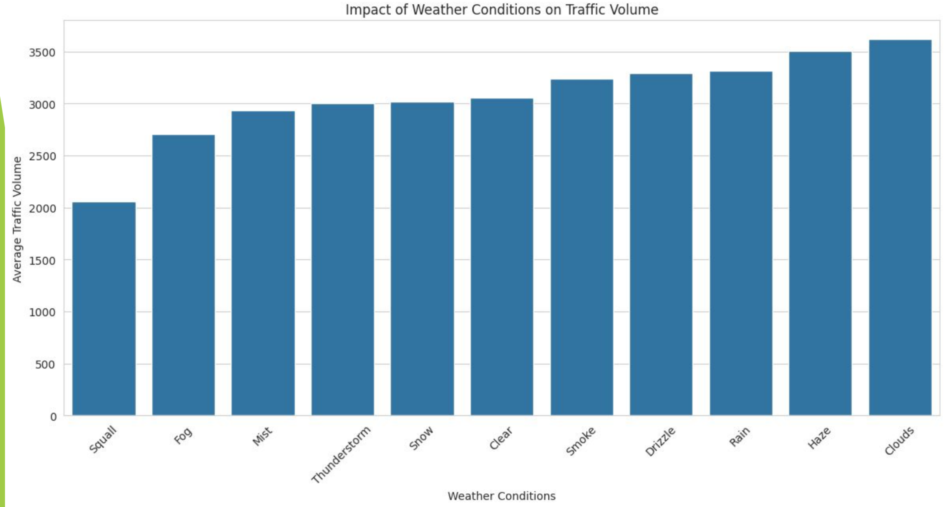


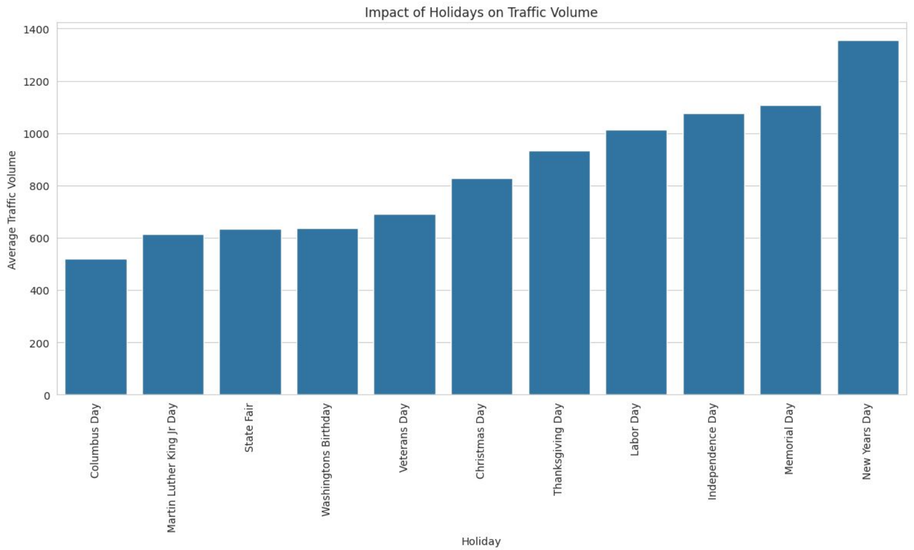





Traffic Volume Over Time
Traffic Volume
7000
6000
5000
4000
3000
2000
1000
0
2013
2014
2015
2016
2017
2018
2019
Date
Traffic Volume


Distribution of Traffic Volumes
Frequency
5000
4000
3000
2000
1000
0
0
1000
2000
3000
4000
5000
6000
7000
Traffic Volume


Impact of Weather Conditions on Traffic Volume
Average Traffic Volume
3500
3000
2500
2000
1500
1000
500
0
Squall
Fog
Mist
Thunderstorm
Snow
Clear
Smoke
Drizzle
Rain
Haze
Clouds
Weather Conditions


Impact of Holidays on Traffic Volume
Average Traffic Volume
1400
1200
1000
800
600
400
200
0
Columbus Day
Martin Luther King Jr Day
State Fair
Washingtons Birthday
Veterans Day
Christmas Day
Thanksgiving Day
Labor Day
Independence Day
Memorial Day
New Years Day
Holiday

3. Feature Engineering :-

- It includes extracting the temporal features such as hour, day of week, month.
- One-hot encoding for all values of weather conditions to convert string values to numeric. Then, joining the encoded weather data back to the main dataframe.
- Conversion of holiday into binary feature, where 1 represents holiday and 0 represents an non-holiday.
- Normalize continuous features of temperature (conversion of kelvin values to celsius)
- Scaling of features and target variable b/w 0 & 1 using MinMax scaler. Then, selection of relevant features for modelling and creation of suitable sequences for training of LSTM and GRU model.

4. Deep Learning Models :-

- In this section, we describe the techniques chosen to design forecasting models.

  Proposed Models-

  1. LSTM
  2. GRU
  3. Stacked LSTM
  4. SARIMA
  5. Ensemble Model = Sarima + Stacked LSTM
  6. Hybrid model= Sarima + Stacked LSTM

4.1) LSTM :-

- **Purpose:** Serves as a baseline for more complex modeling strategies.
- **Model Design:**
  - Configured with a layer of LSTM with 50 units, demonstrating the model's ability to process sequential data effectively.
  - Incorporates Dropout layer at a rate of 20% following LSTM layer to mitigate overfitting, ensuring the model generalizes well to new data.
- **Training Protocol:**
  - Utilized the Adam optimizer for efficient convergence during training, tailored to optimize the mean squared error (MSE) loss function.
  - The model was trained for 50 epochs with a batch size of 32, balancing computational efficiency with exposure to training data.
- **Training Outcomes:**
  - Validation strategies involved a 10% split of the training data to monitor and prevent overfitting, ensuring the model's robustness.
  - Systematic reduction in loss over training epochs signifies effective learning and adaptation to the complexity of traffic data.

4.2) GRU (Gated Recurrent Unit) :-

- Enhanced the GRU model by incorporating dropout regularization (dropout rate of 0.2) to mitigate overfitting and improve generalization.
- Model architecture consists of a single GRU layer with 50 units, followed by a dropout layer and a dense output layer.
- Compiled the model using the Adam optimizer and mean squared error loss function.
- Trained the model for 50 epochs with a batch size of 32, utilizing a validation split of 10%.
- Evaluated the model on the test set to obtain refined GRU test loss and performed statistical analysis to assess accuracy.
- Calculated mean squared error and root mean squared error metrics to quantify prediction error.
- Utilized histograms, scatter plots, and time series plots of residuals for visual analysis of model performance.
- Overall, the refined GRU model demonstrates improved predictive capabilities for traffic volume prediction based on historical data, aiding in traffic management and planning initiatives.

4.3) Stacked LSTM :-

- **Purpose:** Utilized for high accuracy in predicting traffic volume patterns by capturing complex temporal dependencies in the data.
- **Model Architecture:**
  - Two LSTM Layers: First layer with 50 units processes input sequences and returns sequences. Second layer with 50 units compacts the temporal information and does not return sequences.
  - Dense Layer: Single output unit for predicting traffic volume.
- **Training Protocol:**
  - Optimizer: Adam, chosen for efficient convergence.
  - Loss Function: Mean Squared Error, to minimize the prediction error in traffic volume.
  - Epochs and Batch Size: Trained for 50 epochs with a batch size of 32, incorporating a validation split of 10% to monitor model performance and avoid overfitting.
- **Training Outcomes:**
  - Emphasized the reduction in loss over epochs, indicating learning effectiveness and model stabilization.

4.4) SARIMA model :-

1. **Model Specification:** SARIMA extends ARIMA to handle seasonal components, with parameters (p, d, q) for non-seasonal and (P,D,Q)m for seasonal components.

2. **Model Fitting:** SARIMA is fitted to traffic volume data, accounting for daily patterns and seasonal effects, including holidays and events. Regressors like weather conditions are incorporated for improved predictability.

3. **Model Validation:** Rolling forecasting origin technique evaluates model effectiveness, with parameters re-estimated as new data arrives. Forecasting accuracy is assessed using RMSE, and MAE, comparing SARIMA performance to other machine learning models.

4. **Outcomes:**
   - SARIMA effectively forecasts traffic volume, capturing cyclicality and seasonal spikes.
   - SARIMA's strength lies in its easy interpretability and simplicity, making it suitable for preliminary analysis and forecasting in traffic management.
   - Combining traditional and advanced modeling techniques enhances traffic volume forecasting.

4.5) Ensemble Model of SARIMA and Stacked LSTM :-

- The ensemble model combines predictions from SARIMA and stacked LSTM models, averaging their forecasts based on their respective performance weights.

- SARIMA and stacked LSTM models are individually trained on the same dataset to capture distinct data patterns effectively.

- Regularization techniques are applied during training to prevent overfitting, particularly for the more complex stacked LSTM model.

- Performance metrics such as RMSE, MAE, and MAPE quantify the accuracy and reliability of ensemble predictions against actual traffic volumes.

- The combination of SARIMA and stacked LSTM models enhances prediction accuracy and stability in traffic forecasting.

4.6) Hybrid Model of SARIMA and Stacked LSTM :-

- The hybrid model combines forecasts from SARIMA and Stacked LSTM.
- SARIMA and Stacked LSTM are trained separately on the same dataset, with SARIMA parameters optimized for seasonality and Stacked LSTM fine-tuned to capture complex temporal dependencies.
- Utilizes rolling forecasting to assess real-world performance.
- Dropout and early stopping employed in Stacked LSTM to prevent overfitting.
- Outperforms individual models in forecasting accuracy across different traffic situations.
- The Hybrid SARIMA and Stacked LSTM model improves forecasting performance by leveraging unique qualities of both statistical and deep learning models. It enhances accuracy and produces solid predictions under varying traffic conditions.

5. Results :-

- the performance of the developed predictive models for traffic volume forecasting clearly varies across the different approaches used.
- The summary of the results is presented in Table with the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) for each model being reported.

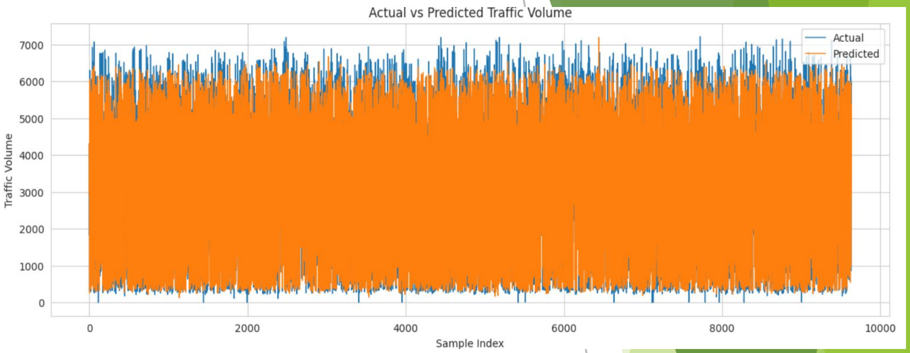
- Results:
  1. Individual model results
  2. Composite model results

5.1) Individual Model Result :-

5.1.1) LSTM

Mean Absolute Error (MAE): 341.516
Root Mean Squared Error (RMSE): 560.465

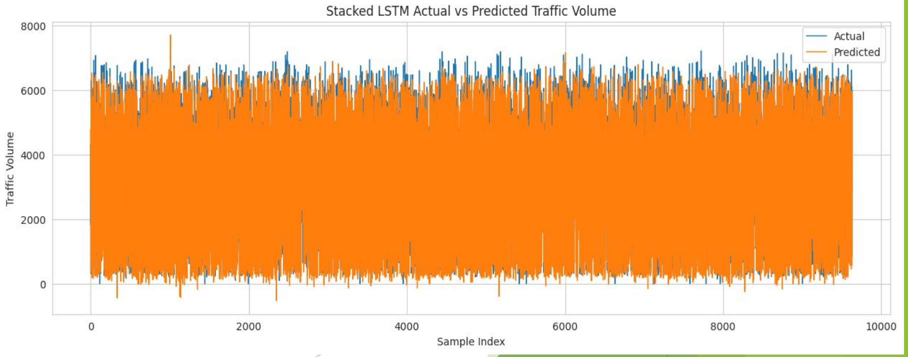
- Strong predictive capabilities, particularly in handling temporal dependencies inherent in traffic volume data.
- Potential for further improvement and optimization with different configurations of LSTM units and dropout rates.



5.1.2) Stacked LSTM

Mean Absolute Error (MAE): 338.6232
Root Mean Squared Error (RMSE): 535.6217

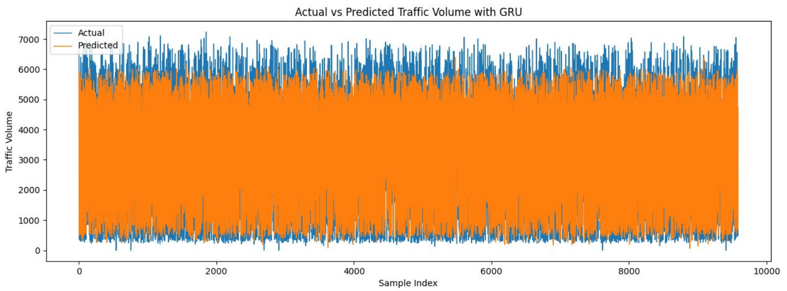
- Stacked LSTM outperformed LSTM and GRU in accuracy and reliability, showcasing its capability to capture complex temporal dependencies.
- Its robust performance established Stacked LSTM as the foundation for developing more sophisticated hybrid and ensemble models to leverage its strengths



5.1.3) GRU

Mean Absolute Error (MAE): 562.229
Root Mean Squared Error (RMSE): 779.127

- GRU underperformed compared to LSTM and Stacked LSTM model.
- Due to its inferior performance, GRU was not selected for further development into more complex models such as Stacked LSTM or hybrid configurations.
- This outcome steered the project towards leveraging LSTM architectures, which demonstrated better potential in handling the dataset's variability and complexity.

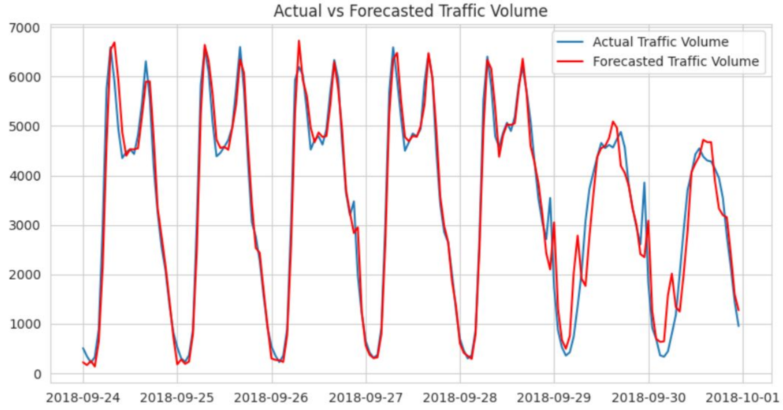


5.1.4) SARIMA

Mean Absolute Error (MAE): 308.807
Root Mean Squared Error (RMSE): 453.885

- SARIMA excels in capturing the cyclical nature of traffic volume, as evidenced by its lower RMSE (453.885) and MAE (308.807), closely mirroring the actual traffic patterns in the graph.

- Foundation for Hybrid Models: The better and comparable performance of SARIMA and Stacked LSTM led to the strategic decision to develop hybrid models, leveraging the strengths of both to enhance overall forecasting accuracy and robustness.



5.2) Composite Model Result :-

5.2.1) Hybrid (SARIMA + Stacked LSTM)

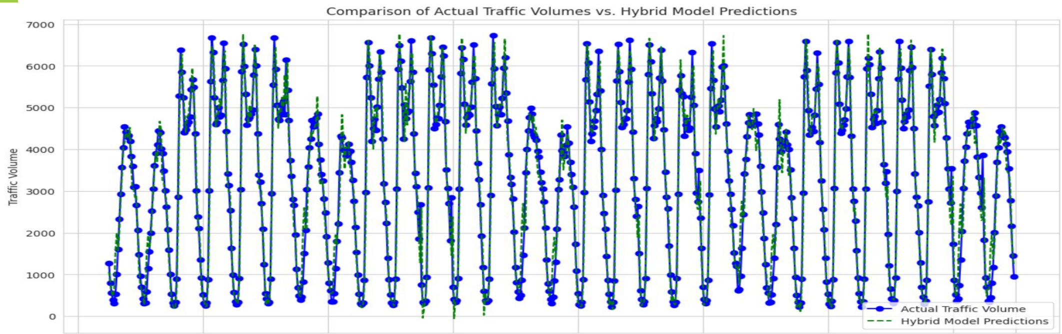
Mean Absolute Error (MAE): 273.266
Root Mean Squared Error (RMSE): 398.489

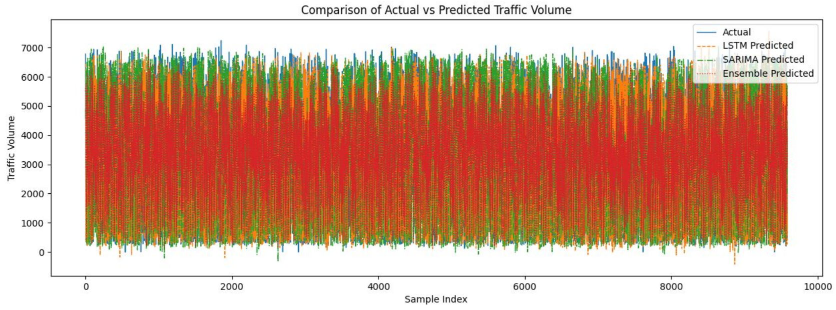
- Achieved the best performance among all models tested.
- Utilized SARIMA for initial accurate seasonal predictions, with Stacked LSTM refining these predictions by adjusting for residuals, enhancing overall forecast precision.
- This model emerged as the most effective, combining SARIMA's cyclical prediction strengths with Stacked LSTM's capability to handle complex temporal dependencies.

5.2.2) Ensemble

Mean Absolute Error (MAE): 1198.145
Root Mean Squared Error (RMSE): 1474.715

- Exhibited the highest RMSE and MAE, indicating a significant underperformance compared to other models.
- Due to its poor initial performance, further development on the ensemble model was halted in favor of focusing resources on the more successful hybrid model.





MODELS	Evaluation Metrics	
	RMSE	MAE
LSTM	566.40	363.53
GRU	779.12	562.22
Stacked LSTM	535.62	338.62
SARIMA	453.88	308.80
Ensemble (SARIMA + Stacked LSTM)	1474.71	1198.14
Hybrid (SARIMA + Stacked LSTM)	398.48	273.26

***TABLE 9.1:** Performance Metrics for Predictive Models*

6. Mitigation Techniques:-

Prediction and Early Intervention:
Traffic Flow Prediction: Your models, especially the hybrid SARIMA and LSTM model, have proven effective in predicting traffic flow and congestion events with high accuracy. This predictive capability is crucial for early intervention. By accurately forecasting traffic volume and identifying potential congestion before it happens, traffic management systems can implement measures to prevent congestion from occurring in the first place.

Dynamic Traffic Management:
Real-Time Adjustments: With real-time data and predictive insights from your models, traffic management systems can dynamically adjust signal timings and traffic flow directions to alleviate congestion points before they become problematic. This proactive approach shifts from reactive to preventive traffic management.

Optimization of Traffic Signals and Routes:
Signal Optimization: Using the predictive outputs, traffic signals can be optimized to ensure smoother flow of traffic. For example, extending green light durations at busy intersections predicted to experience high traffic volumes can reduce the likelihood of bottlenecks.

7. Conclusion :-

- Various predictive modeling techniques were evaluated for predicting traffic volumes in cities, including LSTM, GRU, Stacked LSTM, SARIMA, Ensemble, and Hybrid models.

- The SARIMA model performed well in capturing seasonal patterns, while the stacked LSTM model excelled in modeling complex, nonlinear relationships within the data.

- The Hybrid model, combining SARIMA with stacked LSTM, demonstrated the best performance, leveraging the strengths of both models.

- The ensemble model did not perform as expected, suggesting that integrating outputs from different models may not always lead to optimal results.

- Further investigation is needed to confirm anomalously reported low errors by the GRU model before considering it for practical applications.

8. Future Scope :-

- **Model Refinement:** Future research can refine the hybrid modeling approach by using alternative methods of combination, such as weighted hybrid models or meta learning from the advanced techniques of machine learning, which holds potential to increase the accuracy and robustness of the model.

- **Real-Time Application:** Developing real-time traffic forecasting systems using these models could provide immediate benefits to urban traffic management systems, helping to dynamically adjust traffic signals and manage congestion based on predicted traffic volumes.

- **Deployment in Smart City Solutions:** Models could be implemented within the larger smart city scenarios; predictive analytics could play a major role not only in traffic management but also urban infrastructure planning and public transport service planning.

REFERENCES :-

- Das, L. C. (2023). Traffic Volume Prediction using Memory-Based Recurrent Neural Networks: A comparative analysis of LSTM and GRU. arXiv preprint arXiv:2303.12643.4
- Wu, Y., Tan, H., Qin, L., Ran, B., & Jiang, Z. (2018). *A hybrid deep learning based traffic flow prediction method and its understanding. Transportation Research Part C: Emerging Technologies, 90, 166–180*. doi:10.1016/j.trc.2018.03.001
- Wang, J., Gu, Q., Wu, J., Liu, G., & Xiong, Z. (2016). Traffic Speed Prediction and Congestion Source Exploration: A Deep Learning Method. 2016 IEEE 16th International Conference on Data Mining (ICDM). doi:10.1109/icdm.2016.0061
- Abdullah, S. M., Periyasamy, M., Kamaludeen, N. A., Towfeek, S. K., Marappan, R., Kidambi Raju, S., Alharbi, A. H., & Khafaga, D. S. (2023). Optimizing traffic flow in smart cities: Soft GRU-based recurrent neural networks for enhanced congestion prediction using deep learning. *Sustainability*, 15(7), 5949. doi:10.3390/su15075949
- Zheng, H., Lin, F., Feng, X., & Chen, Y. (2020). *A Hybrid Deep Learning Model With Attention-Based Conv-LSTM Networks for Short-Term Traffic Flow Prediction. IEEE Transactions on Intelligent Transportation Systems, 1–11*. doi:10.1109/tits.2020.2997352
- AS Mihaita, H Li, and MA Rizoiu (2020). "Traffic congestion anomaly detection and prediction using deep learning." arXiv preprint arXiv:2006.13215.

THANKYOU