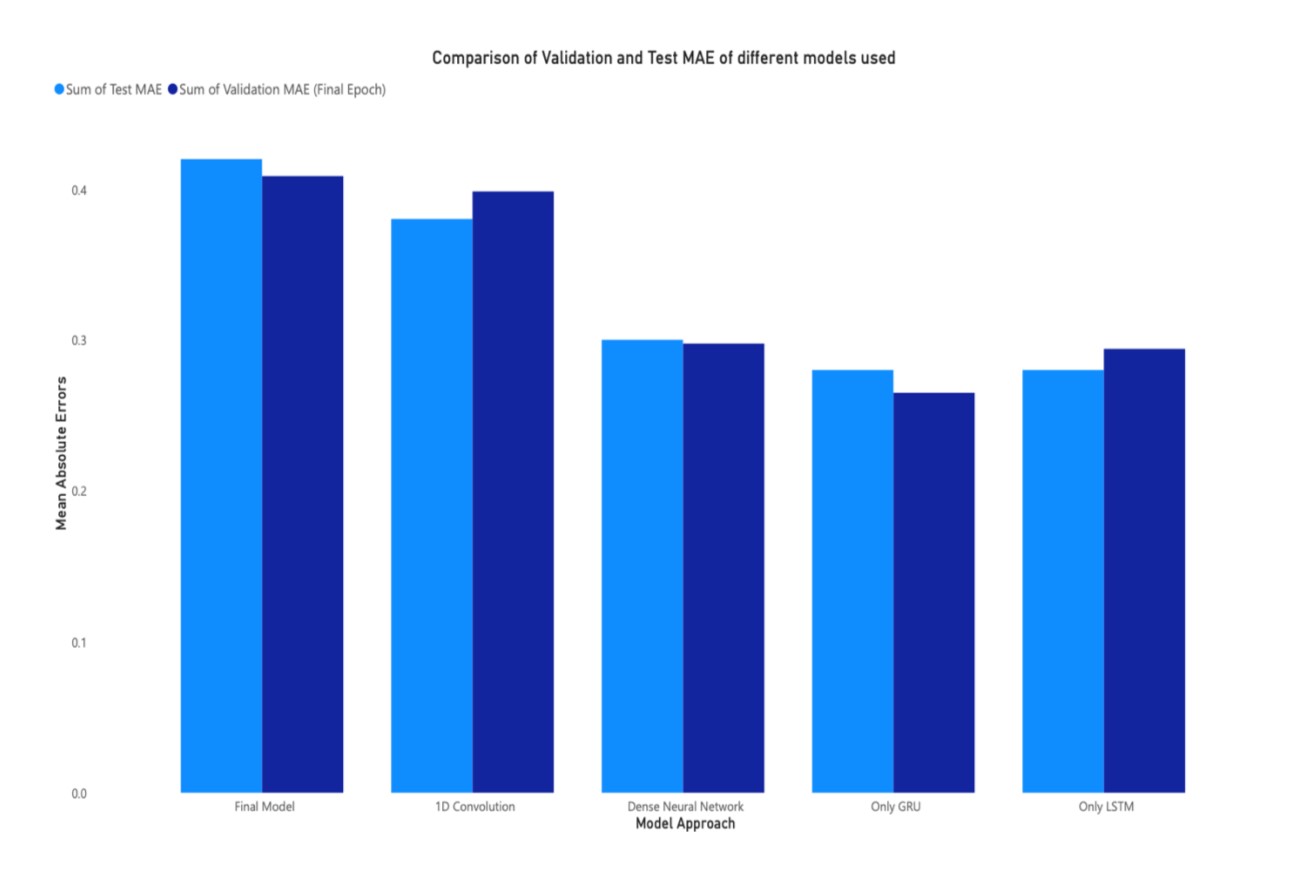
**REPORT**

**Group – 13** (Phani Varshitha, Durga Chowdary)

The project uses machine learning models to predict temperature based on historical data from the Jena Climate dataset, which spans 2009 to 2016. The project investigated a variety of modeling techniques, including dense neural networks, 1D convolutions, LSTM (Long Short-Term Memory), and GRU (Gated Recurrent Unit) layers, with the goal of minimizing mean absolute error (MAE) in temperature predictions.

Here's a table summarizing the key results from the different modeling approaches used in the weather forecasting study:

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Approach** | **Training MAE (Final Epoch)** | **Validation MAE (Final Epoch)** | **Test MAE** |
| Dense Neural Network | 0.2314 | 0.2975 | 0.30 |
| 1D Convolution | 0.2769 | 0.3986 | 0.38 |
| Only LSTM | 0.2262 | 0.2940 | 0.28 |
| Only GRU | 0.2883 | 0.2649 | 0.28 |
| Final Model | 0.3807 | 0.4088 | 0.42 |



**Key Findings:**

Dense Neural Network: Initial attempts with a simple dense network yielded a test MAE of 0.30.

1D Convolution: Employing a 1D convolution method resulted in a slightly higher test MAE of 0.38, suggesting that this approach may not be as effective for the given time-series data.

Only LSTM: The application of LSTM models showed promising results, reducing the test MAE to 0.28, indicating a better capability to capture temporal dependencies.

Only GRU: Introduction of GRU layers and dropout to combat overfitting achieved a test MAE of 0.28, consistent with LSTM performance but with a simpler architecture.

**Final Model Construction:**

In a stacked model setup, the ultimate approach combined LSTM layers with additional units, recurrent dropout, 1D convolutions, and dropout layers. Despite the complexity, this model achieved a slightly higher test MAE of 0.42, which could be attributed to overfitting or the increased model complexity not translating to better generalization.

**Conclusion:**

The exploration of different neural network architectures to forecast weather demonstrated that LSTM and GRU models, with their ability to capture sequential data dependencies, were more effective than traditional dense networks and 1D convolutions. However, the final complex model combining LSTM and 1D convolutions did not significantly outperform simpler LSTM or GRU models, highlighting the importance of model simplicity and the diminishing returns of increased complexity.

**Visualization**

To better illustrate the learning process and model performance over time, we plotted the mean absolute error (MAE) of each model approach across epochs for both the training and validation datasets.

A graph of training and validation

Description automatically generated

The graph above depicts the mean absolute error (MAE) for both training and validation sets using the various modeling approaches used in the study: Dense Neural Network, 1D Convolution, LSTM, and the final complex model.   
  
Dense Neural Network and 1D Convolution approaches have a noticeable gap between training and validation MAE, indicating overfitting, particularly in the dense network model.   
  
The 1D Convolution validation MAE is higher and less stable, indicating that it may not be the best approach for this time-series data.   
  
LSTM models show a closer alignment between training and validation MAE, indicating improved generalization capabilities. The LSTM approach also has lower MAE values than the dense and 1D convolution models, demonstrating its suitability for time series forecasting.   
  
The Final Model, despite being more complex, does not outperform the simpler LSTM model in terms of MAE. The training and validation MAE are higher, which could be attributed to the model's complexity, potentially leading to overfitting or failing to capture temporal dependencies as effectively as expected.   
  
Finally, the LSTM model emerges as the most effective approach for this particular weather forecasting problem, striking the best balance between accuracy and generalization. The exploration of model complexity with the final model demonstrates that more layers and parameters do not always imply better performance on unseen data, emphasizing the importance of model selection and tuning in machine learning projects.