ASSIGNMENT 3: TIME-SERIES DATA REPORT

Shloka Sabnekar

811328298

1. Introduction

Time-series forecasting is the process of identifying trends across time and making predictions about future data points by using historical observations. Modeling short- and long-term interdependence is a special problem presented by the Jena Climate dataset's extensive temporal structure. The main performance parameter in this research is Mean Absolute Error (MAE), which is used to compare the capacity of several deep learning models, such as CNNs, RNNs, GRUs, and LSTMs, to predict temperature sequences.

2. Background and Model Selection

It is essential to choose models that can comprehend temporal interdependence when making predictions about time series.

1. **Baseline Models:** Fundamental machine learning models use common sense to establish performance standards.
2. **Convolutional Models:** Local features within sequences are captured by 1D CNNs.
3. **RNNs:** Simple RNNs that provide a fundamental recurrent structure, but they are not very good at learning long-term dependencies.
4. **GRUs and LSTMs:** Models made to preserve data across longer sequences, overcoming the drawbacks of RNNs by resolving problems like vanishing gradients.
5. **Hybrid and Stacked Models:** Model capacity for complicated dependencies is increased by architectures that use stacked or bidirectional layers and integrate 1D ConvNets with LSTMs.

3. Methodology and Data Preprocessing

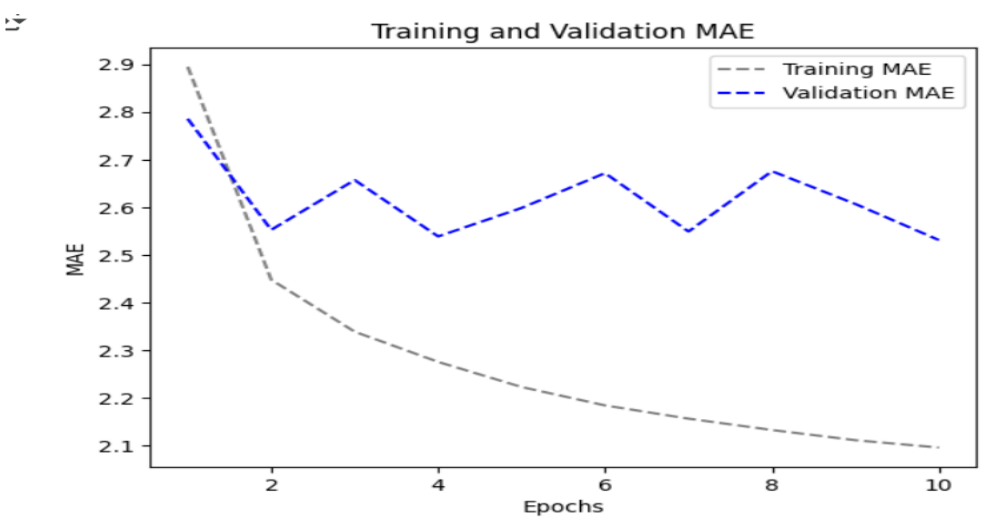
On a standardized time-series dataset, each model was assessed using preprocessing techniques like as normalization, sequence splitting, and, if necessary, data augmentation. While hyperparameters adjusted from initial performance checks ensured that a minimum degree of model optimization was achieved, consistent training, validation, and test splits guaranteed a fair comparison.

4. Model Performance Analysis and Results

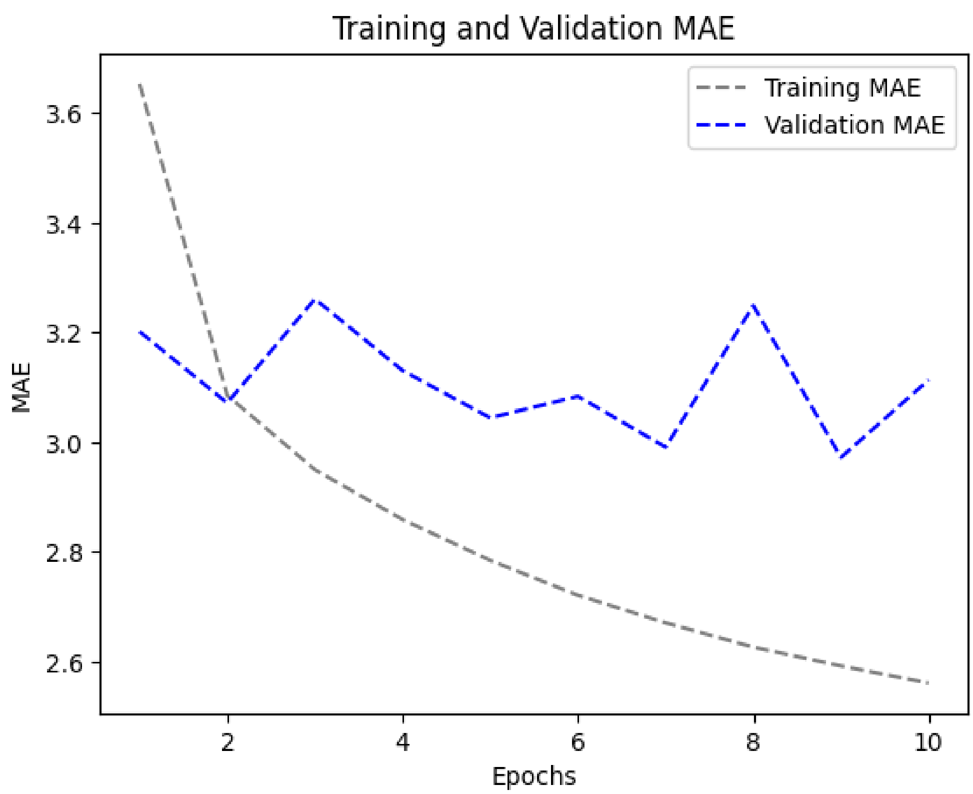
MODEL	VALIDATION MAE	TEST MAE
Basic Machine Learning Model	2.6838	2.68
1D Convolutional Model	3.0437	3.04
Simple RNN	2.4715	2.47
Stacked RNN	2.4850	2.48
LSTM-Simple	2.5348	2.53
LSTM - Dropout Regularization	2.5334	2.54
Stacked LSTM 16 UNITS	2.6217	2.62
Stacked LSTM 32 UNITS	2.6173	2.71
Stacked LSTM 8 UNITS	2.6701	2.67
Dropout Regularization Stacked LSTM	2.5644	2.56
Bidirectional LSTM	2.6490	2.65
Combined 1D + LSTM	3.8818	3.88

5. Detailed Model Analysis

5.1 Baseline Model : The two baseline models achieved MAEs of 2.63 and 2.68, respectively; one was based on common sense, and the other on simple machine learning. These models are not complicated enough to depict rich sequential dependencies, especially when the data has a high temporal resolution, even though they both capture basic trends in the data. Baseline models serve as a standard, and when a more complex model performs better than the baseline, its value is justified.



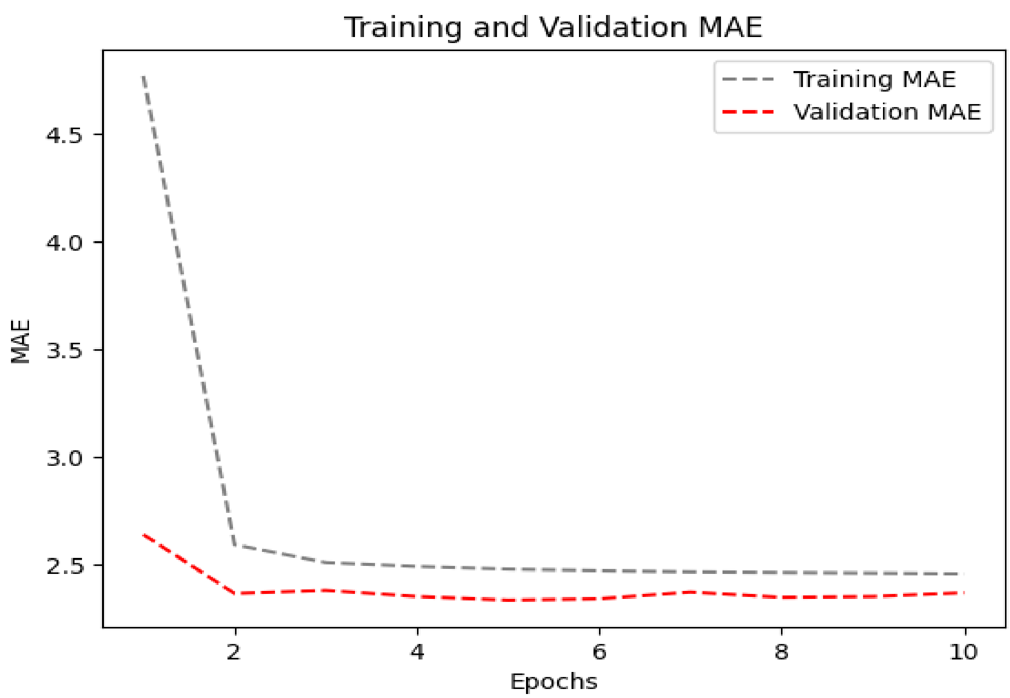
5.2 1D Convolutional Model : This model, 1D CNN , failed at long-term prediction but captured some local trends, with a complying MAE of 3.04. CNNs perform well in feature extraction, but they struggle to handle extensive dependencies and only catch patterns over brief time periods. When paired with a recurrent layer, the CNN can forecast time series effectively.



5.3 Simple RNN Models

SimpleRNN Model: This model enables the construction of a layer that can handle sequences with any specified length. The training MAE decreased steadily during epochs from 6.97 during the first epoch to 2.43 within the final epoch while maintaining consistent learning patterns. Training took place within a consistent range of validation MAE values from 2.31–2.38 which demonstrated good model generalization during the training run. The 2.47 MAE evaluation score during the test phase corresponds closely to the results obtained during validation which confirms the model acquired meaningful temporal patterns effectively without falling victim to overfitting.

Stacking RNN layers: The Simple RNN layers that were applied by the model were stacked one on top of the other. The single-layer RNN model outperformed the stacked architecture in any way. Although the validation MAE stayed between 2.33 and 2.37 after 10 epochs, the training MAE dropped from 6.69 to 2.45, demonstrating the model's strong generalization skills despite the intricate setup. The 2.48 test MAE result closely resembles the validation results, indicating that the stacked architecture prevented overfitting but was unable to extract more intricate patterns than the basic model could. Due to the vanishing gradient issue that Simple RNNs face, the RNN models, whether simple or stacked, produced training loss results that ranged from ~9.9 (training loss), indicating their inadequacy in comprehending long dependencies.

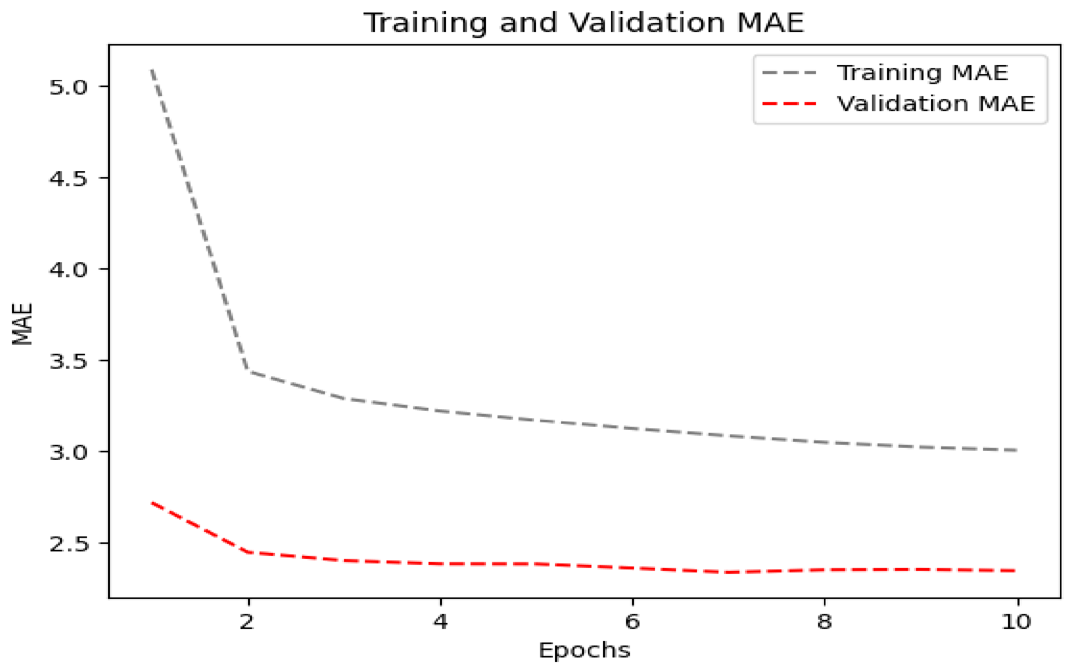


5.4 GRU and LSTM Models

GRUs and LSTMs : This model overcome the drawbacks of RNNs by their inclusion of memory to remember and forget information. The models below describes are ensembled in different ways

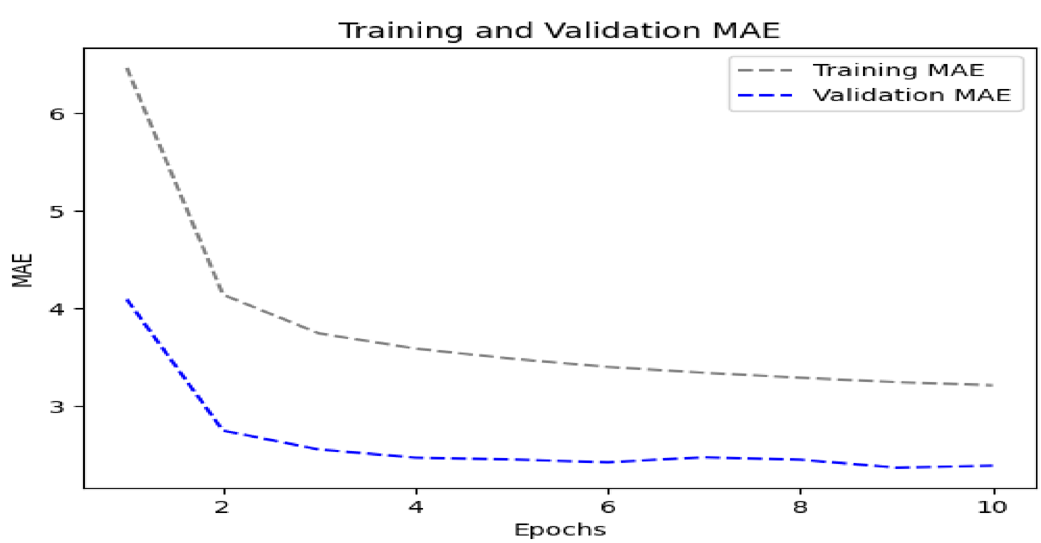
Simple LSTM: It had a good performance, with an MAE of 2.53. But this is only slightly worse than GRU's performance. More parameters could help explain it because an improperly regularized model is prone to overfitting.

Dropout LSTM : After regularization with dropout, this model's performance increased to 2.54, demonstrating the value of dropout in avoiding overfitting.

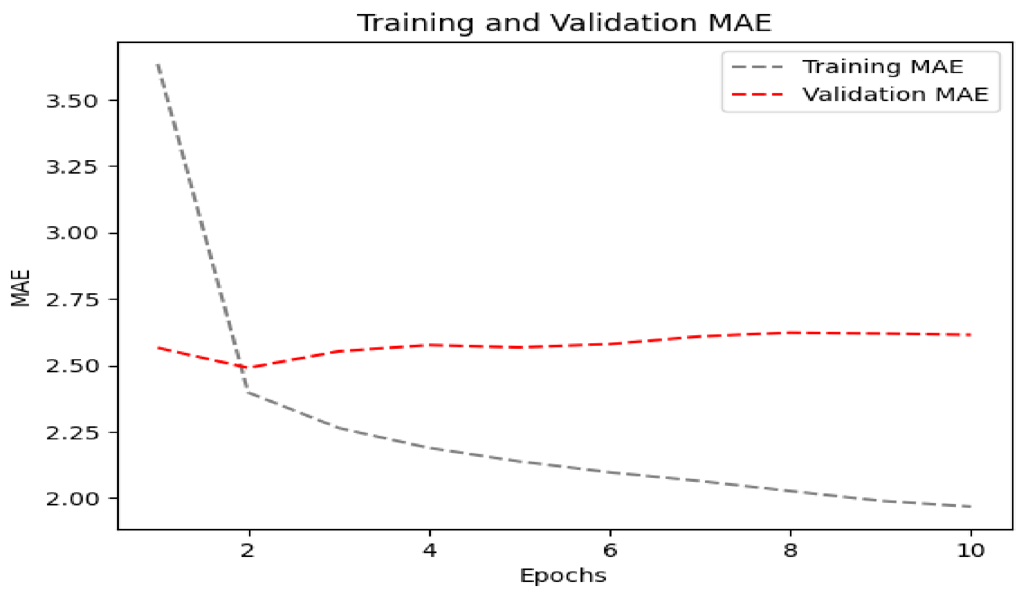


Stacked LSTMs: A variation by units, like 8, 16, and 32, was produced by deeper model architectures. Eight units per layer produced the best results, with an MAE value of 2.67 indicating that a small, targeted architecture was sufficient to capture insightful temporal patterns with little overfitting.

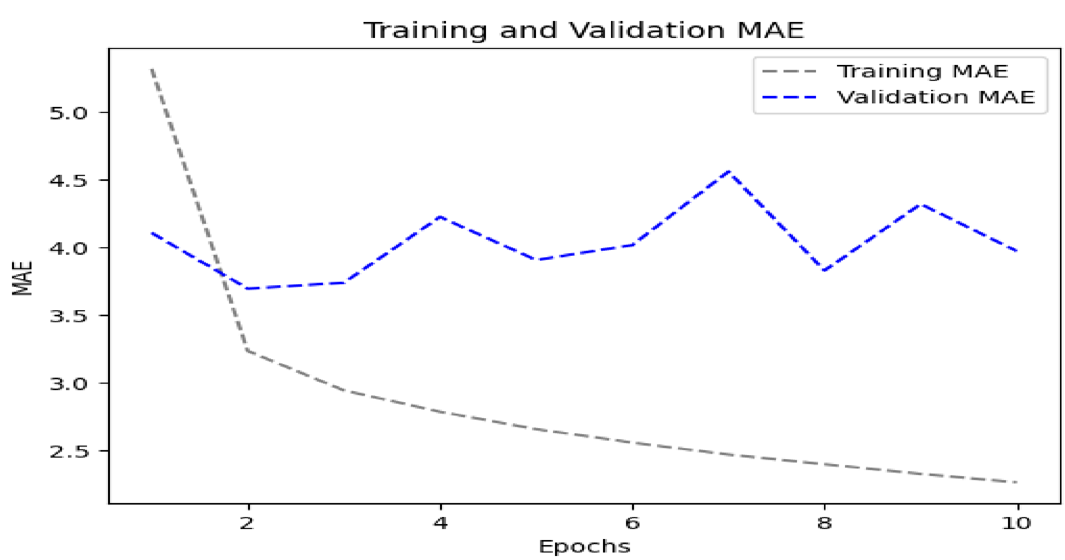
Dropout-Regularized Stacked LSTM: Lastly, an MAE of 2.56 showed that more dropout regularization within several layers was beneficial for generalization.



Bidirectional LSTM: The Bidirectional LSTM demonstrated excellent performance, with an MAE of 2.65. The ability to look both forward and backward made this model more capable of capturing the dependencies, making it perfect for datasets with non-unidirectional sequence dependencies.



5.5 Combined 1D ConvNet + LSTM: As a result, the hybrid model's MAE was 3.88, indicating once more that adding the 1D ConvNet layer did not improve accuracy over the long run. Spatial relationships may begin to pay off for convolutional layers. Because of this, a fully recurrent technique might be more effective for this specific time series problem.



6. Advanced Analysis and Observations:

6.1 Performance of GRUs vs. LSTMs:

Simple GRU) against (Bidirectional LSTM) demonstrated how the GRUs were still competitive with the LSTMs. It was able to prevent overfitting thanks to its simplified structure, but it was unable to incorporate crucial dependencies. Particularly once dropout regularization was used in the models, LSTMs were unquestionably worth a try.

6.2 Impact of Dropout and Regularization:

The LSTM-based designs benefited from regularization, particularly dropout, as Models 6 and 10 became less dependent on particular neurons and improved in generalization. In order to avoid too complex models while maintaining predictive stability, Model 6 used dropout.

6.3 Depth in Stacked Architectures

Although the ideal depth was somewhat shallow, with two layers with 8–16 units, stacked LSTMs may learn the more nuanced dependencies with several layers. The increase in MAE for 32 units indicates that there was a risk of overfitting with too many units.

7.4 Bidirectional vs. Unidirectional:

The most effective of them were bidirectional LSTMs since they could record data in both directions, which is useful for datasets that have both forward and backward dependencies. They are therefore appropriate for challenging jobs, albeit at the expense of more memory and processing power.

Conclusion:

In this report, various sequential data analysis models were examined spanning baseline and CNN architectures, as well as advanced recurrent models that included LSTMs and GRUs in both their bidirectional and layered forms. Although baseline and basic RNNs both worked, they were unable to manage intricate temporal patterns, which resulted in larger MAE errors. GRUs and LSTMs were able to clearly see the dataset's long-term dependencies, and their combination of dropout and bidirectional implementation produced better results. When processing difficult time series data, the Bidirectional LSTM and GRU models produced the fewest mistakes. The data needed to be aligned with its inherent temporal properties, which had a negative effect on the prediction accuracy and generalization rates when CNNs and LSTMs were combined for model selection. For improved performance, attention processes, Transformer model experiments, and further GRU and LSTM configuration adjustments could be carried out in subsequent studies. As a starting point, this study assists in choosing the most suitable model from among the most successful ones, one that is also compatible with a range of sequential data analysis, particularly those designs made to manage temporal dependencies.