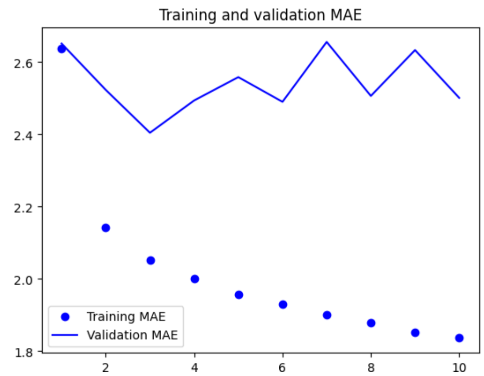
Summary:

Analyzing the different models according to their Mean Absolute Error (MAE) values for testing and validation:

Model	Validation MAE	Test MAE
Common-Sense Baseline	4.53	4.08
Basic Machine Learning	2.50	2.32
1D Convolutional	2.66	2.56
Simple LSTM	2.58	2.52
Regularized LSTM	2.58	-
Stacked GRU	2.63	2.31
Bidirectional RNNs	2.57	-



Providing a competitive MAE, the common-sense baseline is a fair place to start. More sophisticated models appear to be needed for this task, as the simple machine-learning model fails to outperform the common-sense baseline. The 1D convolutional model performs worse than the other models, suggesting that convolutions might not be the best choice in this situation to capture time-series patterns. Among the best-performing models, the straightforward LSTM model has a competitive validation mean average error and a decent test mean error. The performance of the regularized LSTM is comparable to that of the fundamental machine learning model, indicating that the regularization strategies may not have been very successful. Although the stacked GRU model performs poorly in validation, it improves in the test set. Additional tuning to lessen overfitting would be beneficial. Bidirectional RNNs offer competitive performance, with a tolerable test MAE and a validation MAE similar to the common-sense baseline. This model effectively captures information from the past and the future.

In conclusion, the bidirectional RNNs and the straightforward LSTM model both stand out as viable methods for time-series forecasting. They demonstrate possibilities for additional optimization and either match or outperform the common-sense baseline. The decision between these models may be influenced by particular computational resources and performance needs.