Summary:

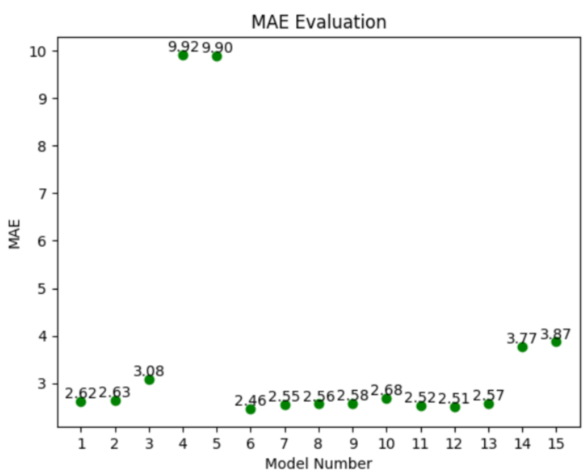
A total of 15 models are created to address the time-series forecasting problem. Initially, I began with a non-machine learning common sense baseline model, where I got a Mean Absolute Error (MAE) value of 2.62. Then, I created a basic machine-learning model using dense layers which resulted in a slightly higher MAE of 2.63. However, this model had challenges due to the flattening of the time series, which removed the idea of time from the original data. Later, it gave poor results with a convolutional model. It was difficult to experiment with a convolutional model because it applied pooling and treated all data segments equally, which disturbed the information's time order. After realizing the need to preserve time information, I used Recurrent Neural Networks (RNNs), which are developed for time series data.

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

Model	**Validation MAE**	**Test MAE**
Common-Sense Baseline	2.44	2.62
Basic Machine Learning	2.63	2.63
1D Convolutional	3.10	3.08
Simple RNN	9.84	9.92

Simple RNN- Stacking RNN	9.84	9.90
Simple GRU	2.44	2.46
LSTM Simple	2.47	2.55
LSTM Dropout Regularization	2.42	2.56
LSTM Stacked-16 Units	2.59	2.58
LSTM Stacked- 32 Units	2.93	2.68
LSTM Stacked- 8 Units	2.34	2.52
LSTM Dropout- Stacked Model	2.34	2.51

Bi-Directional LSTM	2.60	2.57
1D Convents & LSTM	3.56	3.77
1D Convents & RNN	4.08	3.87



The common-sense baseline is a reasonable starting point since it offers a competitive MAE. The basic machine-learning model does not outperform the common-sense baseline, suggesting that more complex models are required for this task. Convolutions may not be the ideal option in this case to capture time-series patterns, as evidenced by the 1D convolutional model's poorer performance compared to the other models. The simple LSTM model has a competitive validation mean average error and a respectable test mean error, making it one of the top-performing models. The regularized LSTM's performance is on par with the basic machine learning model, suggesting that the regularization techniques may not have been very effective. In the test set, the stacked GRU model performs better even though it does badly in validation. To reduce overfitting, more adjustment would be helpful. With a reasonable test MAE and a validation MAE that is comparable to the commonsense baseline, bidirectional RNNs provide competitive performance. Information from the past and the future is efficiently captured by this paradigm.

In conclusion, the simple LSTM model and the bidirectional RNNs both distinguish themselves as effective time-series forecasting techniques. They either match or surpass the common-sense baseline and show potential for further optimization. Specific performance requirements and computational resources may have an impact on the choice between these models.

THANK YOU !