

\*The gray boxes are what is different from the basic RNN model

\*\*The bold entry in the Test MAE column is the lowest Test MAE

One interesting conclusion that I found while working with the daily temperature data set is that GRU models outperform their more complex LSTM counterparts in Test MAE. And this appears to apply even when adding Stacked Layers. My original assumption was that LSTM would have been the better of the two models, but perhaps the complexity increase caused LSTM to be more overfit then the GRU models.

It also appears that the second Stacked layer has a neutral or in some cases negative impact on the Test MAE scores of models. And for the LSTM one stacked layer significantly increased the Test MAE. The stacked layer likely increased overfitting by adding onto the already overly complex model. On the other hand one stacked layer appears to be the ideal amount for GRU models allowing them to train on the data better without overfitting the model.

The RNN and 1D Convnets model was by far the worst preforming model likely do to the two approaches having conflicting affects instead of working together.

It appears that for this data set 16 hidden units is the most successful number. Going below that to 8 and above that all the way to 128 only decreased the success of the most successful model I created.

Another surprising finding from the report was the impact of dropout on the Test MAE. It appears to have been relatively helpful in mitigating the over fitting of the LSTM model. But the GRU model only had a slight improvement in Test MAE. It should also be noted that the Dropout models had an increase in epoch time of over 10 times. That indicates an increase in computational need for the Dropout models that is disproportionate to its improvement in Test MAE.

In the end it seems that a GRU model with 4 hidden layers and one stacked layer with 16 hidden units is the most affective model for this data set.