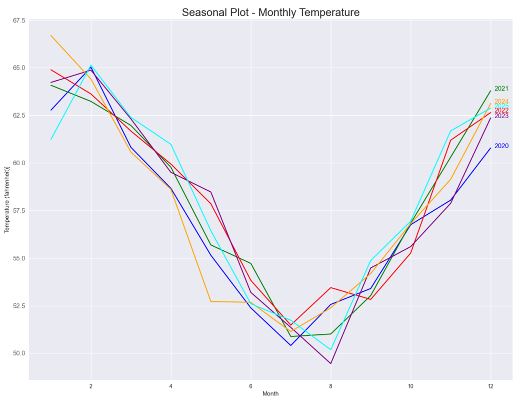
Findings: Deep Learning approach comparison between LSTM and Standard Neural Network

Hyacinthe Chemasle

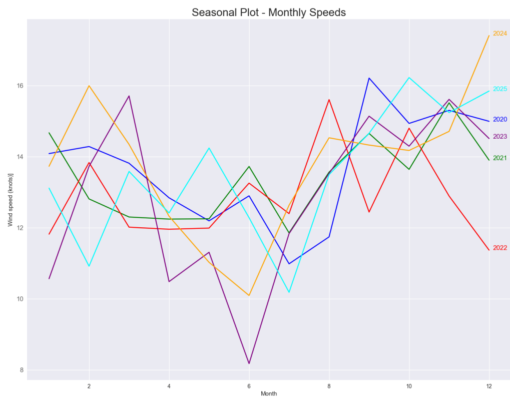
1 EDA (Exploratory Data Analysis)

For our research proposal we first started by carrying out some exploratory data analysis to find what the most insightful seasonal trends tend to be in the METAR reports.

We selected Visibility, Temperature and Wind speed. Wind speed had the most informative trends with seasonal spikes in the earlier and later parts of the month, whereas temperature showed a consistent parabolic relationship throughout the years. The earlier parts of the month saw consistent decrease to a stable trough in months 6-8 and then an increasing trend into the later parts of the year.



(a) Temperature



(b) Wind Speed

Figure 1: Seasonal plot for monthly temperature and wind speed statistics

2 Data Preprocessing

Before feeding our data into our LSTM and MLP model we need to make sure that the data is clean and equipped for the model. These preprocessing techniques were used:

1. Imputation Approaches
2. Deletion Approaches
3. Standardisation
4. Ordinal Encoding
5. Feature Selection

The weather reports conducted at Wellington International Airport contained a lot of columns with missing data. We had to drop a total of 18 columns.

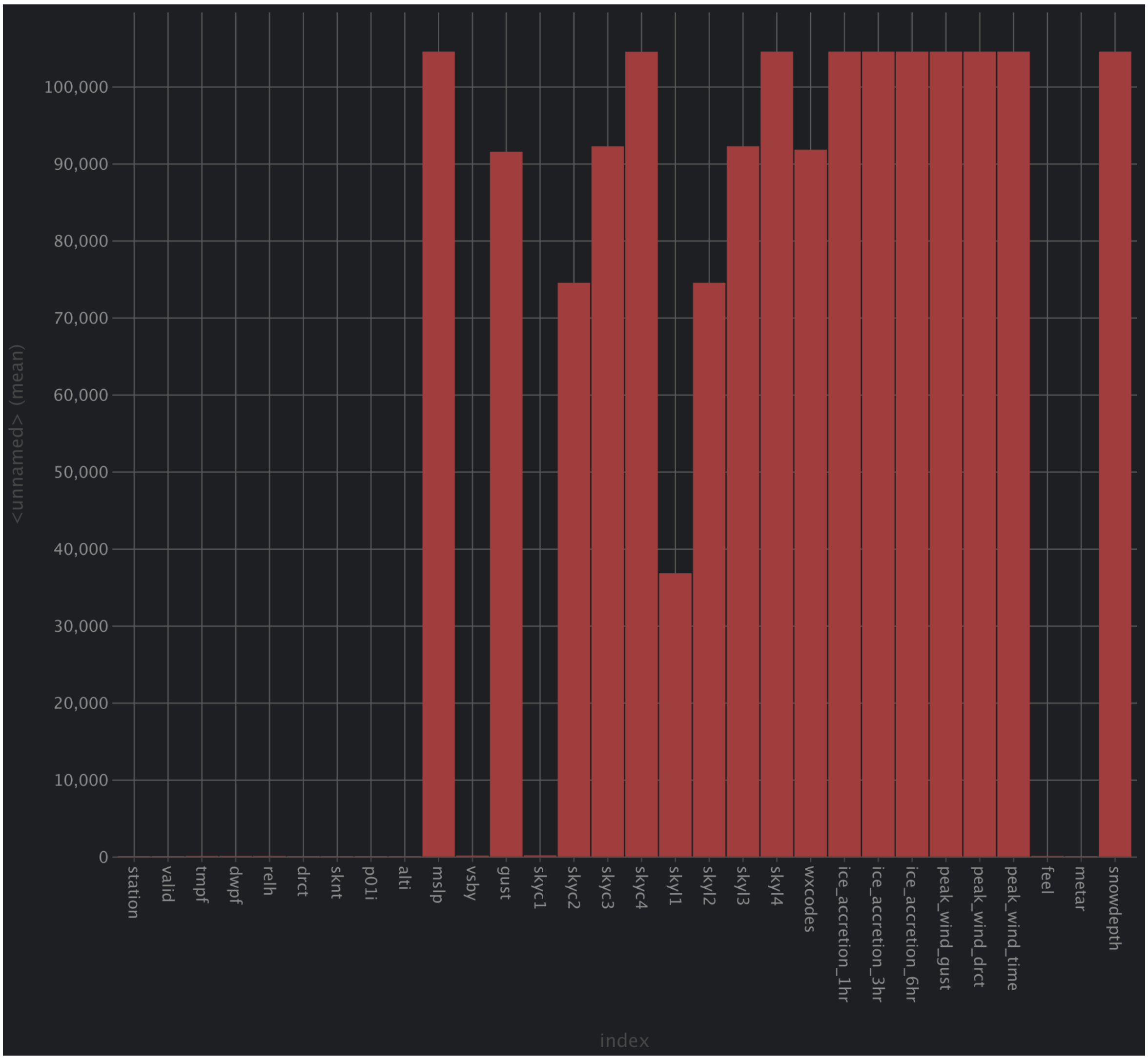


Figure 2: Summary Statistics for columns in weather reports

Imputation approaches were used for sky level coverage as around 30% of the data in this column was missing. Although not visually representative in the above figure, columns like temperature, visibility and feel had on average less than 32 missing instances which allowed us to use deletion approaches since it made up $< 1\%$ of data.

All Numerical features were standardised using the standard scaler. No categorical variables were utilised after feature ranking apart from sky cover which had a natural rank order to classify cloud clover: “Clear”, “Few” “Scattered”, “Broken”, “Overcast” “Vertical Visibility”. Ordinal encoding was utilised during the preprocessing stage to extract this order.

Pearson correlation was utilised to find relationships between features that contributed greatly to the final prediction of a target. This information was leveraged to give us a subset of five features that contributed the most meaningfully to the targets of: “Temperature”, “Visibility” and “Wind speed”.

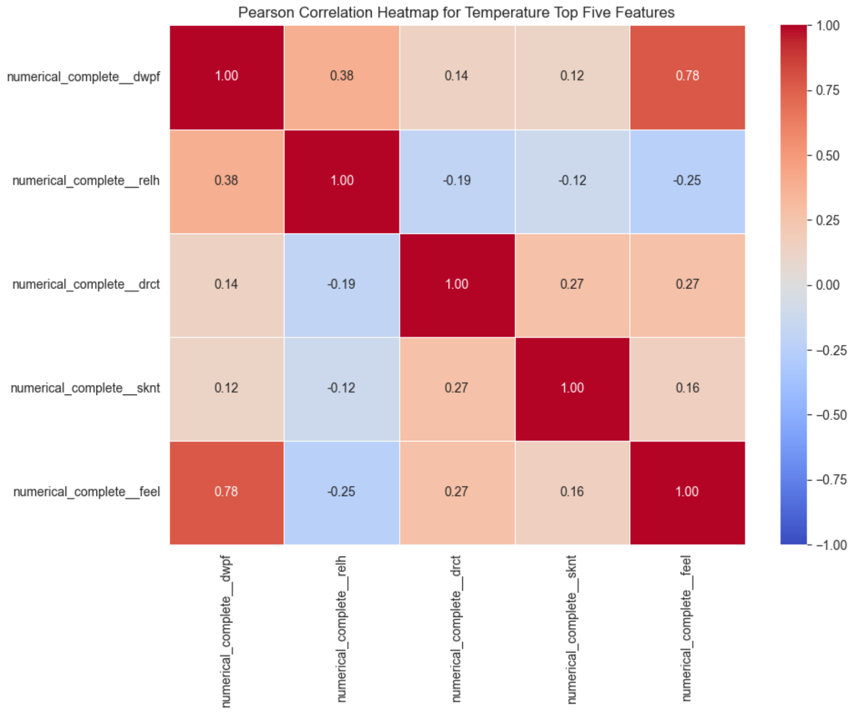
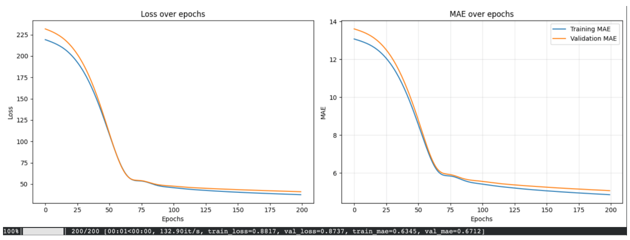


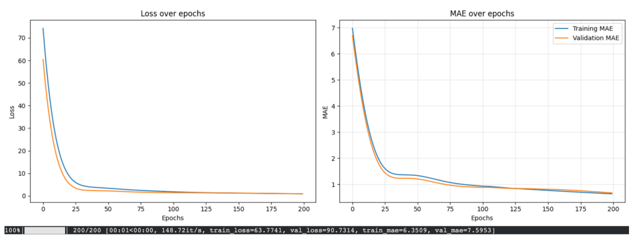
Figure 3: Heatmap visualisation of correlation between features for temperature

Top Selected Features for temperature: Dewpoint, Relative Humidity, Wind Direction, Wind Speed, Apparent Temperature

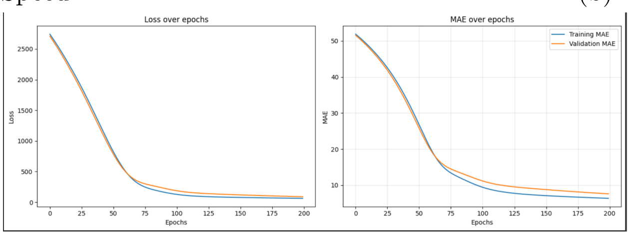
3 Training and Validation Curves for MLP Regressor



(a) Wind Speed



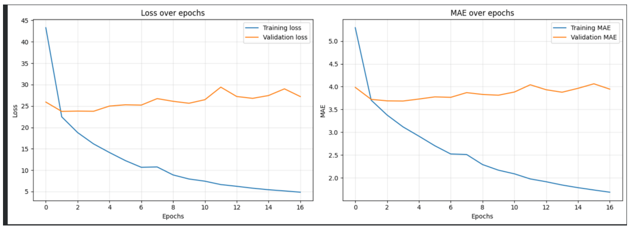
(b) Visibility



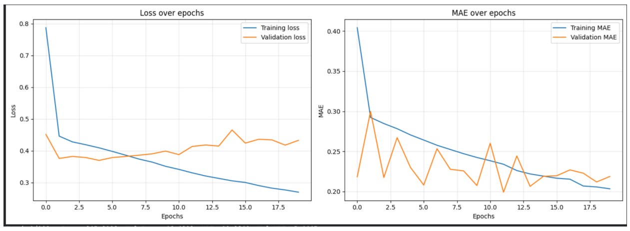
(c) Temperature

Figure 4: Loss and MAE for MLP Targets

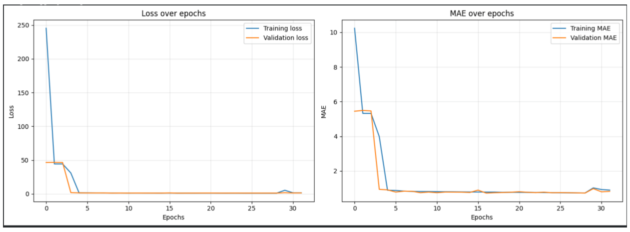
4 Training and Validation Curves for LSTM



(a) Wind Speed



(b) Visibility



(c) Temperature

Figure 5: Loss and MAE for LSTM Targets

5 Findings of MLP Regressor

For the MLP we tested two models labelled "Base" and "Extended" where the base model does not concern itself with extra hidden layers whereas the Extended model leverages this by creating a funnel shape. Both models utilise the MSE Loss function and Adam optimiser with a learning rate of 0.001.

5.1 Base Model Architecture

The base Neural network had an input size of 10 with one hidden layer of 64 neurons. Then comes the final layer being the one output where the model predicts a continuous value for either Wind speed, Visibility or Temperature.

Target	Wind Speed	Visibility	Temperature
MAE	6.659	1.400	45.600
MSE	67.100	2.956	2137.970
RMSE	5.048	1.710	46.230

Table 1: Base Model Metrics

5.2 Extended Model Architecture

The extended architecture contained the same input size with two extra hidden layers leveraging the ReLU activation.

- Input Layer: 10 Neurons
- Hidden Layer One: 64 neurons
- Hidden Layer Two: 32 neurons
- Hidden Layer Three: 16 neurons
- Output Layer: One Neuron (Prediction)

With this funnel-like architecture we can leverage more of the nature that data like wind speed, visibility and temperature have. We can scale up to more features to provide a more complex space to learn relationships between variables and scaling down to retain the most important relationship to then give an accurate prediction.

Target	Wind Speed	Visibility	Temperature
MAE	43.910	49.900	6.460
MSE	2064.550	2586.720	64.980
RMSE	45.430	50.850	8.060

Table 2: Extended Model Metrics

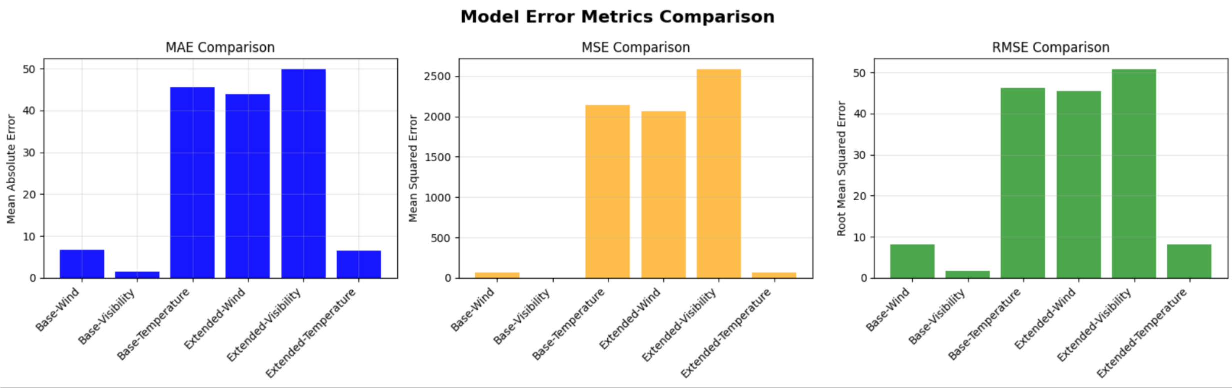


Figure 6: Model Error Metrics Comparison

Findings show that the Extended Model struggled to predict continuous values for wind speed and visibility. However, for the base model the discriminative ability was much higher for wind speed and visibility whereas temperature did tend to struggle.

When using the extended architecture temperature was the best performing improving its RMSE by 38.17 where the base model gave an RMSE of 46.23 whereas the extended architecture gave an RMSE of 8.06.

6 Findings of LSTM

For the second deep learning approach we utilised a Long-Short Term recurrent neural network approach that would be able to leverage the dynamics of time series. This was done by creating sequences in our data to be trained on in a 24-hour period. This would then help us to predict a continuous value for wind speed, visibility and temperature.

6.1 LSTM Architecture

Model consists of three hidden layers with 128 hidden neurons and an output size of one. We utilise dropout as a regularisation effect in case model overfits and split training data into batches with a batch size of 32. The Adam optimiser is chosen with a learning rate of 0.001 along with the MSE loss function.

Metric	Wind Speed	Visibility	Temperature
MAE	3.860	0.226	0.821
MSE	25.480	0.485	1.334
RMSE	5.048	0.696	1.155

Table 3: LSTM Performance Metrics

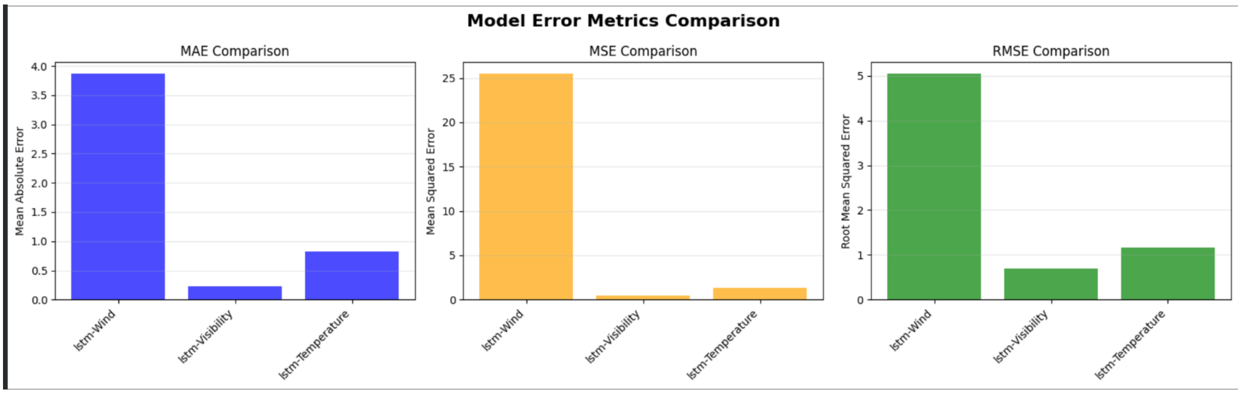


Figure 7: Model Error Metrics Comparison

The model had the greatest discriminative ability when it came to predicting visibility. This makes sense as a lot of the data has a value of 6.21 Statute miles indicating fair to good flying conditions which is the standard for most airport operations across the world.

Temperature was the second and most interesting performer with an RMSE of 1.155 compared to its extended MLP model of RMSE 8.06. As we discovered in our Initial EDA, temperature displayed a parabolic relationship with time across the year and seasons. These findings show that the LSTM has been able to retain an important relationship between time and temperature in our data at wellington international airport.

7 Comparison between MLP and LSTM

The LSTM outperforms the MLP regressor in all three targets. For example, the best RMSE for wind in the MLP section was 6.659 whereas the RMSE for the LSTM was 3.86.

8 Future Work

Based on these findings it is quite clear that the LSTM is the appropriate model for this forecasting task. The model performs the best on predicting temperature and less so on predicting wind speed. On further reflection, the visibility target doesn't seem to be informative in terms of finding a meaningful relationship that would be of importance to wellington international airport.

For future work it would be interesting to test targets such as Cloud Cover for categorical prediction. This would entail testing a Neural Network with SoftMax activation to predict across 6 classes of cloud cover categories and see how it compares to an LSTM for classification.

It is also worth mentioning that the early stopping that happened during training for our LSTM might suggest that it is learning too quickly or that it is overfitting on held out validation data. For the future it will be necessary to tune hyperparameters to see if we can yield a better model for the task of predicting Wind speed, Visibility or temperature.