- Use Case

  This data can be used to predict either a full or half year of air temperature in Dunedin, New Zealand, the data set can be changed to other regions to predict the air temperature of other regions.

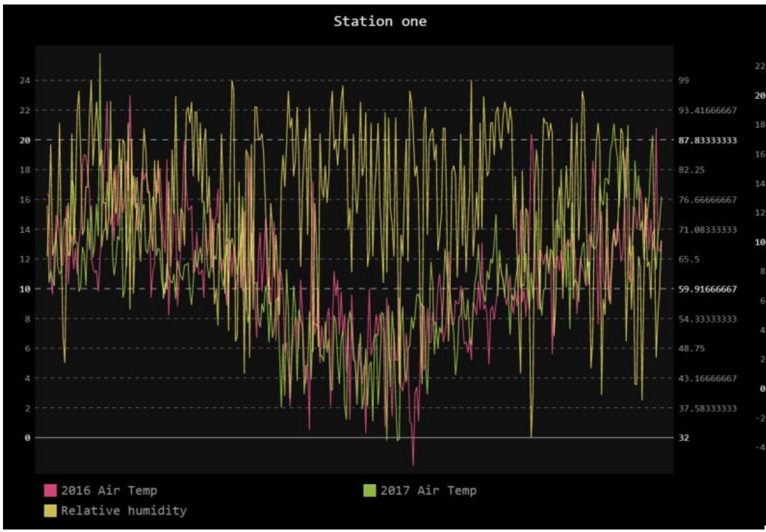
- Data Set

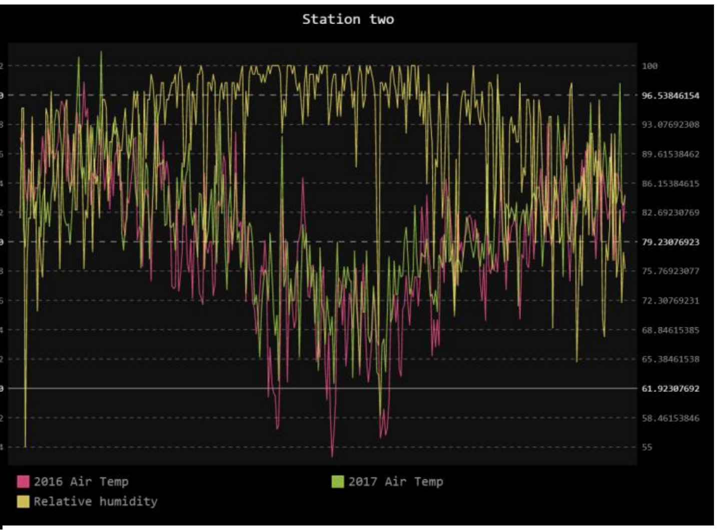
  The project used daily weather data sets collected from cliflo.niwa.co.nz to train, test and evaluate the model.

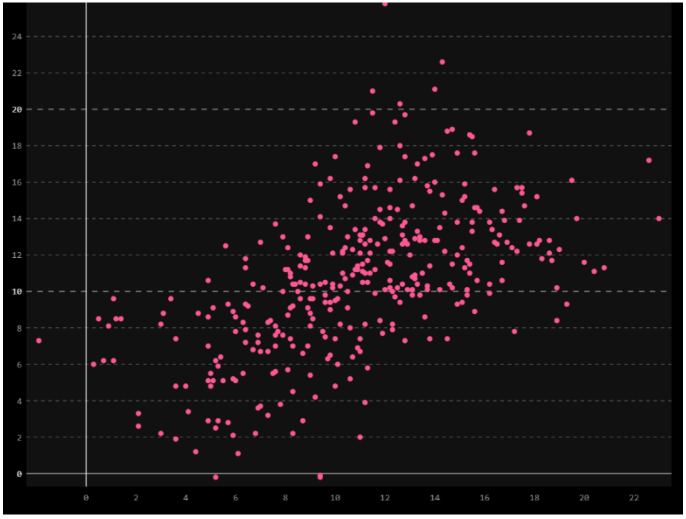
- Data Quality Assessment

  We used pandas to process and clean the data. Overall, the data is high quality as it comes from a credible organisation, but there were a few missing data which had to be dropped.

- Data Exploration & Data Visualisation





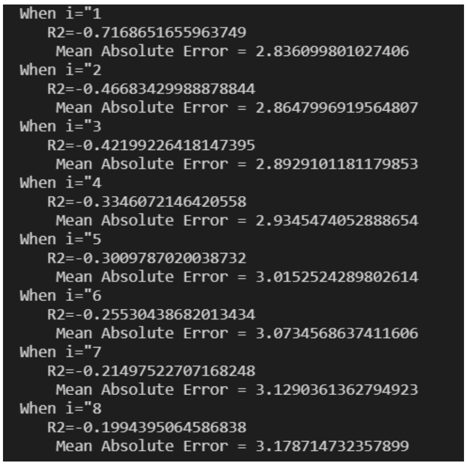


- At least one Feature Engineering (e.g. imputing missing values) applied:

  Pandas was used to extract the data of two different stations in Dunedin. Furthermore, missing values were dropped, and we aligned the dates of the independent and dependent variable’s data.

- Selection and justification of Model Performance Indicator (e.g. F1 score)

  We used R2 and mean squared error as they accurately measure the accuracy of predicted continuous variable.



```
When i="1
   R2=-0.7168651655963749
    Mean Absolute Error = 2.836099801027406
When i="2
   R2=-0.46683429988878844
    Mean Absolute Error = 2.8647996919564807
When i="3
   R2=-0.42199226418147395
    Mean Absolute Error = 2.8929101181179853
When i="4
   R2=-0.3346072146420558
    Mean Absolute Error = 2.9345474052888654
When i="5
   R2=-0.3009787020038732
    Mean Absolute Error = 3.0152524289802614
When i="6
   R2=-0.25530438682013434
    Mean Absolute Error = 3.0734568637411606
When i="7
   R2=-0.21497522707168248
    Mean Absolute Error = 3.1290361362794923
When i="8
   R2=-0.1994395064586838
    Mean Absolute Error = 3.178714732357899
```

- At least one traditional Machine Learning Algorithm and one DeepLearning Algorithm applied and demonstrated

  We used the classic Linear Regression Model and a machine learning random forest model.

- Model performance between different feature engineering and models compared and documented

  The performance of Linear regression was worse than that of random forest, it had a lower R2 and mean absolute error than random forest.