

Architectural Decision Document for Dunedin Weather Predictor.

- Data quality & feature engineering

I assessed the original weather data by visualising it with pandas and pygal, which are fast and efficient. i.e. I immediately found that the data contained two types of station, missing data points, and the 4 types of weather data provided. Furthermore, visualising these features with pygal showed that there is a positive linear relationship between air temperatures each year, where humidity had an inverse relationship with temperature. Both relationships were moderate, which is why I chose last year temperature and humidity as the features used in the model.

For the half year model, I used only temperature data, but from the past 3 years in order to take into account the rising temperature that affected the accuracy of the full year model.

- Why have I chosen a specific algorithm?

I used the random forest regressor as the prediction algorithm as it proved to be the most accurate in capturing temperature trends.

- Why have I chosen a specific framework?

I chose python 3.8 as the framework as it provides data processing efficiency and is easy to use. It also have many useful open source packages that helps with visualising data and making models.

- Why have I chosen a specific model performance indicator?

I chose to use R2 and Mean squared error from the sklearn package, they are accurate and representative indicators of the accuracy of continuous data.