**Report:**

* **Write in a suitable for non-technical audience**
* **Report findings (strengths/weaknesses of potential models for use with new instances)**
* **EXPLAIN DECISIONS**

**What we still need to do:**

* **Decide on what we will say we’ve done (can just say in accordance w/ contact)**
* **Select a model**
* **Make predictions (in valid format)**
* **Save predictions**
* **Submit:**
  + **.ipynb**
  + **PDF version of notebook (w/ output and markdown)**
  + **A clear concise PDF summary report (recommended 6 pages)**
  + **predictions.csv text file**
  + **Individual statements on individual contribution (maximum 1 page)**

**19-20: An exceptionally clearly presented notebook and summary report, demonstrating extensive successful implementation of all the basic ML process elements, very well explained decisions and analysis of results. Assignment conveys significant insight into data preprocessing, and important features in the success of the model.**

Introduction:

* The Big Picture: predicting temperature, why it’s useful
* The dataset: ERA5, size, common use

The ERA5 is a publicly available dataset from the Centre for Medium-Range Weather Forecasts (ECMWF). This dataset provides hourly estimates for a large range of atmospheric, ocean-wave and land-surface quantities recorded over many years. Data pertains to a specific geolocation which is identifiable by a given longitude and latitude.

Predicting the temperature based off of the many features included within this dataset is a problem which proves especially relevant to weather forecasting. Machine learning methods can be used to obtain a prediction of temperatures for data fed into the model

Goal: Create models for predicting temperature given observations about weather

Methods:

* Exploration of data structure
* Detail preprocessing/ data cleaning
* Considerations of making use of available data
* Feature engineering (methods of nearby & others) – why are nearby features useful
* Self-made validation set
* Cross validation for hyperparameter tuning
* Construction of several machine learning regression models
* Selection of one model to make predictions on the test data
* Evaluation methods

Results:

* Training/validation scores of all models
* Improvements from nearby features
* Thorough evaluation on whether models are over- or under-fitting training data
* Thorough justification of hyperparameters (i.e. why grid/random search parameters are applicable)
* Why we’re selecting the main one (need to justify)
* Results from final model on test set
* What these mean in business terms

Discussion:

* Extra things we explored
* Other methods of nearby features and why we didn’t use them
* Issues we had with the models
* What could be added to make the project better
* More powerful machine for faster fine tuning
* More complex neural network architectures (regularisation methods such as dropout/early stopping with callbacks)

Conclusions

References