

Project Proposal

Edison Murairi

October 16, 2023

1 Introduction

The goal of my project is to forecast weather. More precisely, I am interested in predicting temperature based on other weather indicators. To this end, I have selected the so-called ‘Jena Weather’ dataset. This dataset is a time series measurements of weather variations in Jena in Germany. The dataset consists of 420551 measurements, each separated by 10 minutes between January 1st, 2009 and January 1st 2016. There are columns in total, all of which are continuous variables. For my study, I will set the temperature as the predicted variables and all the other will be predictors. Below, I summarize the information about the dataset and provide the list of predictors and their physical units.

- Predicted variables: Temperature (Celsius)
- Number of measurements: 420551
- Number of continuous variables: 14
- Number of categorical variables: 0
- Time separation between each measurements: 10 minutes
- Start date: January 1st 2009
- End date: January 1st 2016

2 List of Numerical Features

1. Pressure (mbar): Atmospheric Pressure
2. Temperature (Kelvin) (colinear with the temperature in celsius),
3. Tdew (Celsius): Temperature in Celsius relative to humidity
4. rh (%): Relative Humidity
5. VPmax (mbar): Saturation vapor pressure
6. VPact (mbar): Vapor pressure
7. VPdef (mbar): Vapor pressure deficit
8. sh (g/kg): Specific humidity
9. H_2O C (mmol/mol): Water vapor concentration
10. ρ (g/m^3): Airtight
11. wv (m/s): Wind speed
12. max. wv (m/s): Maximum wind speed
13. wd (deg): Wind direction