DSE 2242– Fundamentals of Machine Learning Lab
Week 5 – Date: 4th January 2025

EXER 1

Consider the dataset 'weatherAUS' called Rain in Australia. It describes the weather characteristics on different dates and locations. This dataset contains about 10 years of daily weather observations from many locations across Australia.

1. Preprocess the data, remove the attributes which were are not useful to predict rain. Also, remove rows with at least one missing value for each of them.
2. Calculate the Variance Inflation Factor (VIF) value. VIF is a number that determines whether a variable has multicollinearity or not (starts from 1, and it has no upper limit. If the number gets larger, it means the variable has huge multicollinearity on it.).
3. Remove multicollinearities by creating new features. Find the features that have paired values and create the new feature which is the difference value between those pairs.
4. Remove features that have a VIF value above 5.
5. Build a regression model to perform the Rain prediction. Also, tabulate accuracy of the prediction models, before and VIF computation.