

## **Predicting Rains in Australia**

Mohammed Abduldaim

### **Abstract**

The goal of this project was to use classification models to predict the probability of rain occurring the next day or not. I worked with data provided by the Australian Bureau of Meteorology and obtained from Kaggle; this dataset contains about 10 years of daily weather observations from many locations across Australia.

### **Design**

RainTomorrow is the target variable to predict. It means did it rain the next day, Yes or No? This column is Yes if the rain for that day was 1mm or more. Predicting rain using machine learning can help improve better preparation of safety measures for areas that might get flooded

### **Data**

The data contains over 145,000 rows and 23 columns.

Some of the notable columns are:

- Date
- Location
- Raint Today
- Rain Tomorrow

The rest of the columns describe the state of weather each day, such as temperature humidity, sunshine, etc. Which should be extremely useful in building a model

### **Algorithms**

This project used logistic regression for modeling the data and has used both under sampling and over sampling techniques. Also, accuracy, precision, recall and f-measure have been used to evaluate the model.

### **Tools**

- Numpy and Pandas for data manipulation
- Scikit-learn for modeling
- Matplotlib and Seaborn for plotting