Table of Contents

Go to:

- Introduction
- Automation
- Preprocessing 🗓️
- Modelling 📉
- Prediction 📈
- Conclusion ❕

Rain in Australia

Preprocessing Steps

- delete features with over 10% of missing values
- replace Nans for categorical variables with mode
- replace Nans for numerical variables with median
- delete Date column since it is not used for modelling (note: The date is deleted for making the model easier. One should keep in mind that the seasons in fact have an influence on the weather. Therefore for advanced modelling the date/month should be considered)
- encode RainToday and RainTomorrow in binary variable
- encode location and variables for wind direction with get_dummies (note: Since there are a lot of Locations in the dataset, this step leads to an enormous increase of the number of features)
- Scaling of numerical features by vector normalization

Pressing the button "Make dataset" randomly chooses 20% of the original data to create a new dataset

Make dataset

Pressing the button "Preprocess" preprocesses the newest dataset

Preprocess

Preprocessing done!