Data Pipeline Process

*To feed R Monitoring App Module*

# Download

Download .csv file from Australian BOM

# Prepare

Move all files from raw\_archive to incoming (including the downloaded one)

# Data processing

D:\Work\Projects\Australia Rain Tomorrow\Recover\Version.**x** - Complete Dataset\recurrent\_check

Run this notebook:

1 - CSV Data Preprocessing.ipynb

- Perform preprocessing of raw .csv data

- Skip unwanted lines

- Rename columns to fit with original raw data

- Add missing columns to fit with original raw data

Output: preprocessed/formated\_data.csv

*>> copy to D:\Work\Projects\sydney-rain-forecast\dataset\formated*

# Feature Engineering

Run this notebook:

2 - Feature engineering - V1.ipynb

- Perform feature engineering (add, merge columns...)

- Fill NaN values based on strategies

Output: processed/ processed\_dataset\_v1.csv

*>> copy to D:\Work\Projects\sydney-rain-forecast\dataset\processed*

*>> rename to M****x****\_processed\_dataset\_v1*

Prediction & Check Pipeline

*To feed R Script*

Manual BS

*In reccurrentCheck.R*

# -- Observe reactiveVal: predictions

observeEvent(predictions(), {

….

Add this

debug\_\_predictions <<- predictions

Launch app and go to monitoring

>> select model and run predictions

Exit app

R Console:

formated\_data <- read.csv("dataset/formated/formated\_data.csv")

formated\_data$Prediction <- debug\_\_predictions$Raw.Prediction

write.csv(formated\_data, "output/df\_with\_predictions", quote = F, row.names = F)