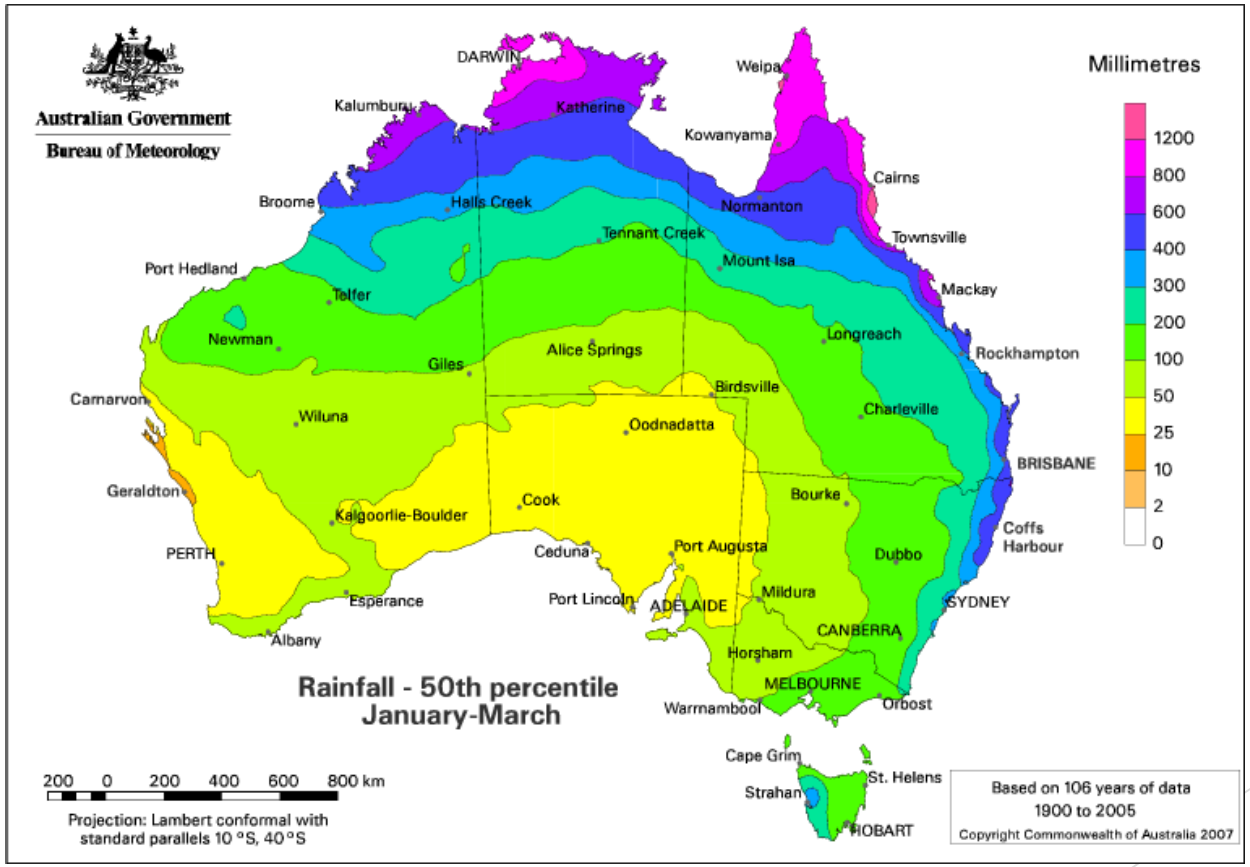
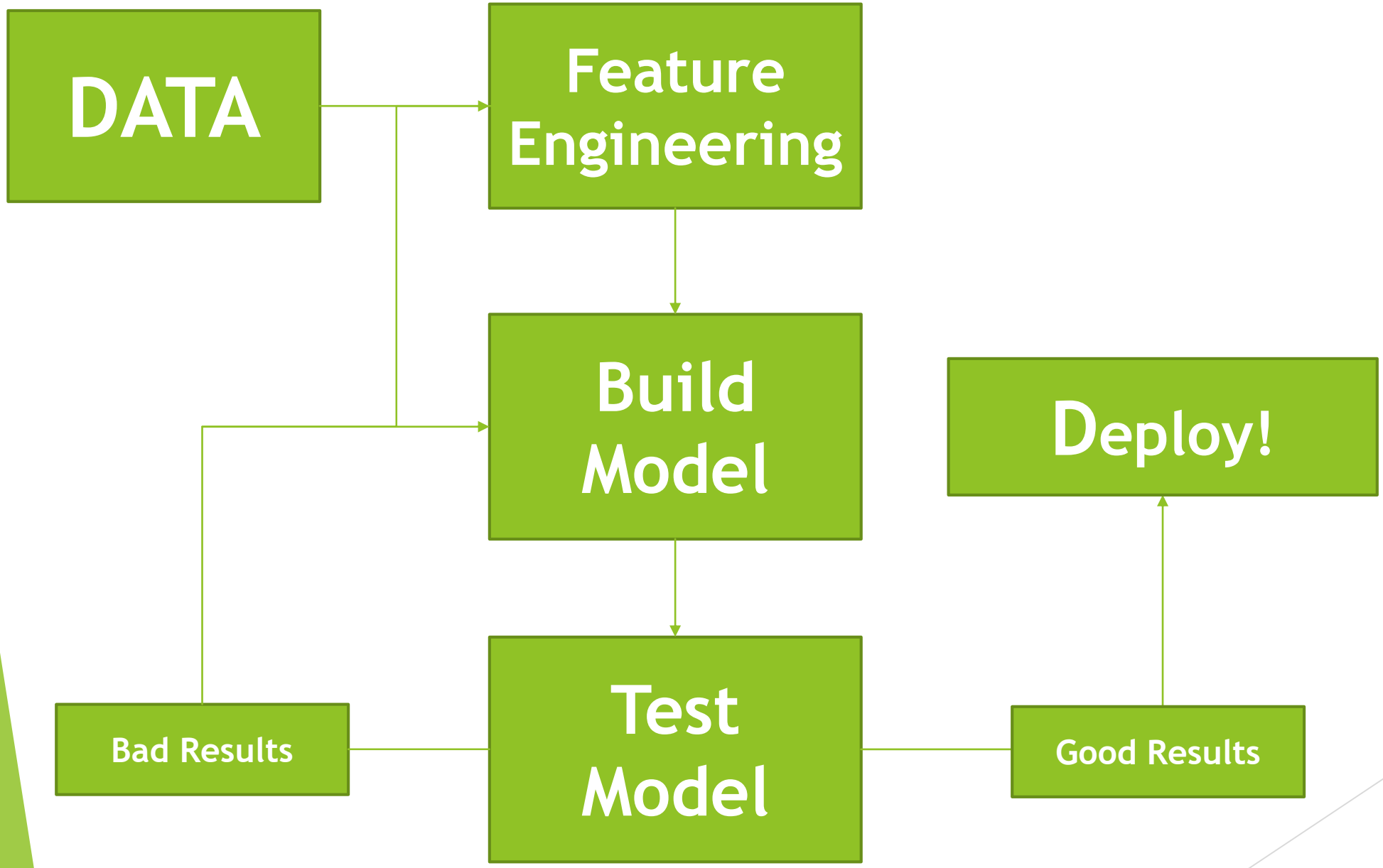
Predicting if it Rains

Australia



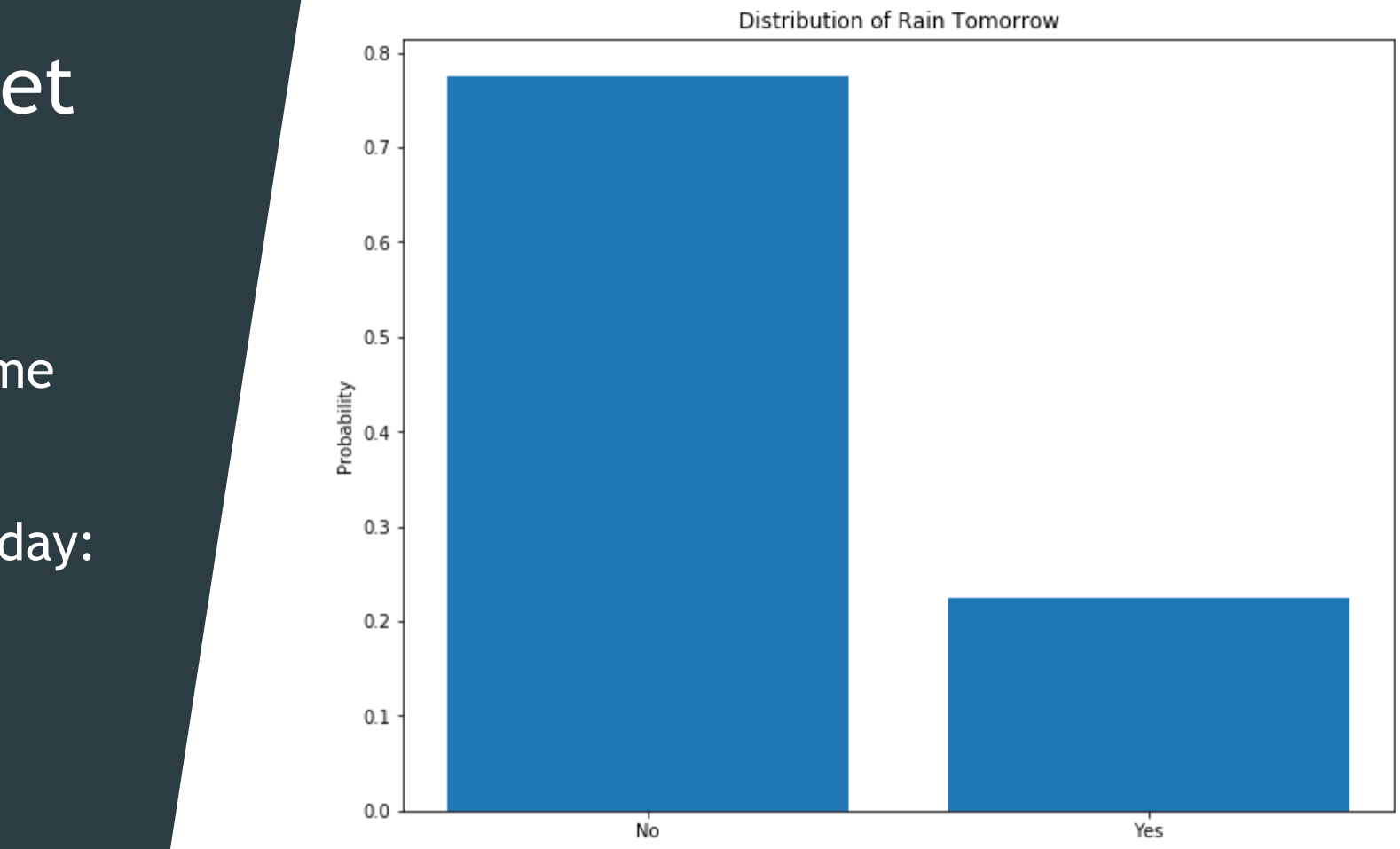
The Data

- Target:
  - Rain Tomorrow (Yes vs. No)
- Input Data
  - Temperature
  - Wind (speed, direction)
  - Clouds, Sun
  - Location
  - Rain Today (Yes vs. No)


DATA
Feature Engineering
Build Model
Test Model
Bad Results
Good Results
Deploy!

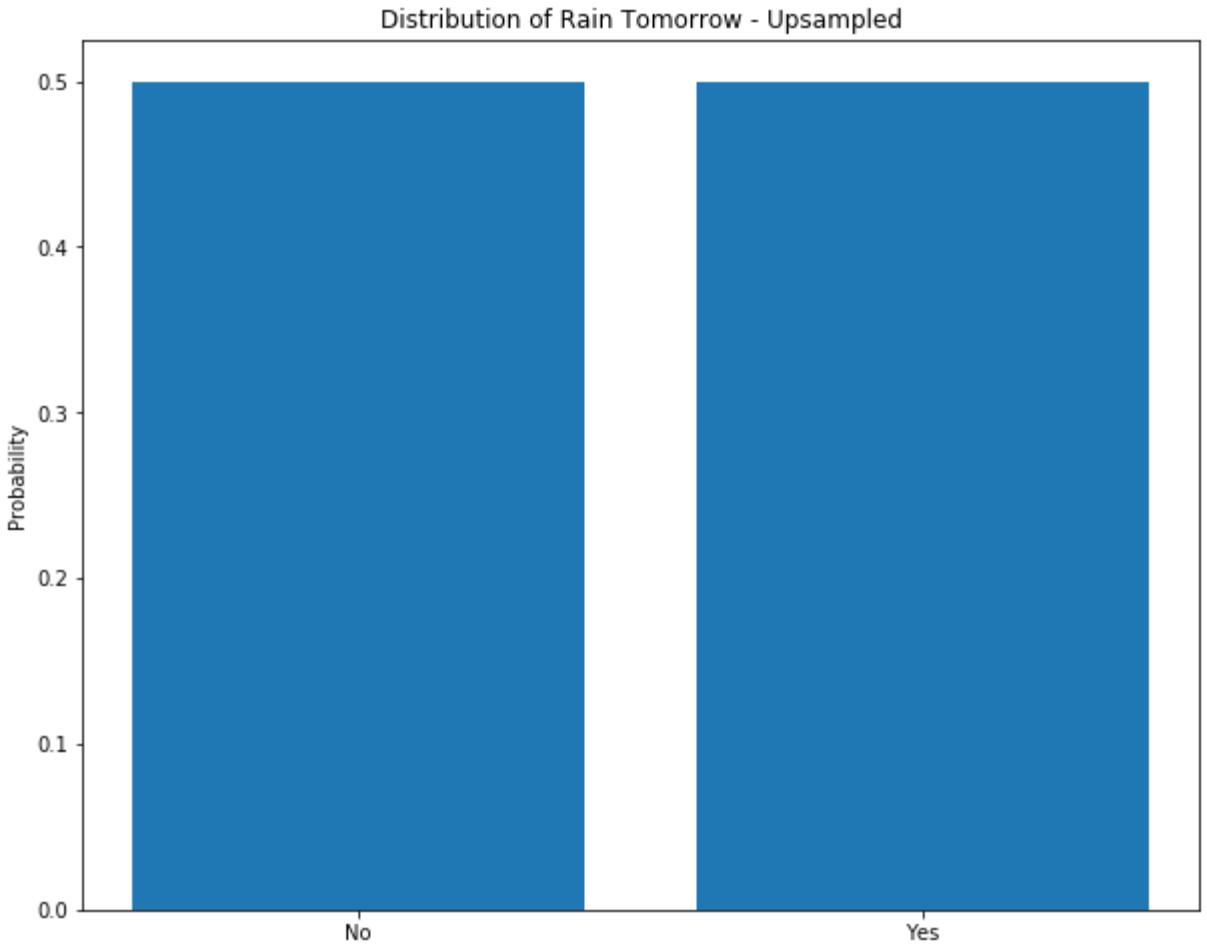
Problem:
Imbalanced Target

- Rains only 22% of the time
- Predicting no rain everyday:
  - 78% accurate!



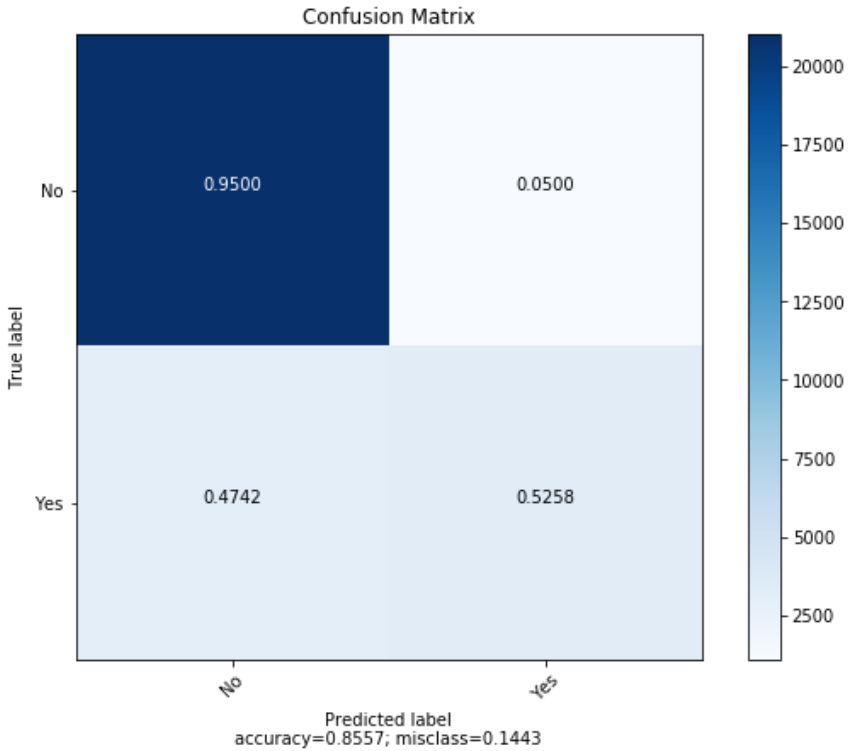
Solution: Upsampling

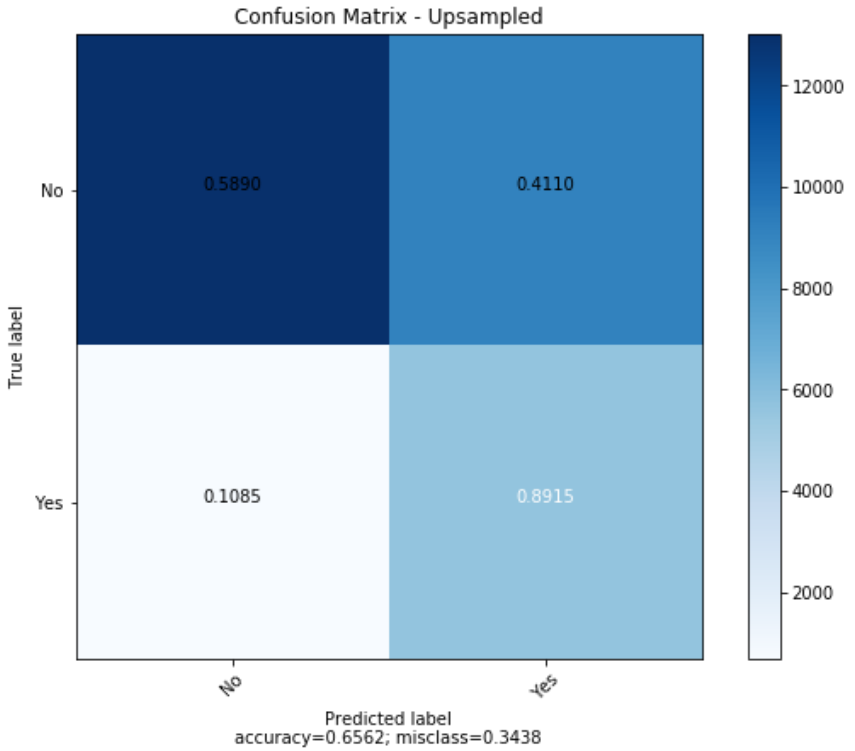
- Create realistic synthetic data
- Balance Target



Upsampling – Pros vs. Cons

- Pros:
  - When it does rain, correct more often
  - Predict rain, actually Sun – Happy Surprise
- Cons:
  - Model predicts rain too often
  - Unrealistic





Solution: Build Two Models - Compare

Recommendations

- Model Options:
  - **Model 1 - Non-Upsampled:**
    - Very Accurate
    - But, when it rains, only right half the time
    - Overall: **Realistic**

Recommendations

- Model Options:
  - **Model 2 – Upsampled:**
    - Not very accurate
    - But, when it rains, right 90% of the time
    - Overall: **Rain Focused**

Recommendations

- Model Choice:
  - Depends on Goals!
- For Meteorology:
  - Choose **Realistic** Model
  - Correct more on average
- If we care more about Rain:
  - Choose **Rain Focused** Model
  - Correct more when it rains