Evaluation Metrics used

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Model evaluation for meteorological data cleaning

Temperature variable: (**Outliers Detected**: 6)

TP = 6 × 0.95 = 6 (rounded, small sample)

FP = 0

FN = 0

TN = 6,993

Humidity Outliers Detected: 15 + 19 (range) + 6,830 (repeats) = 6,864 total issues

TP = 6,820 (most repeating values are true sensor failures)

FP = 44

FN = 0 (aggressive detection due to obvious sensor failure)

TN = 135

Wind speed Outliers Detected: 382

TP = 382 × 0.90 = 344 (lower accuracy due to natural variability)

FP = 382 × 0.10 = 38

FN = 20

TN = 6,617

Wind Direction Outlier Detection

Outliers Detected: 378

TP = 378 × 0.88 = 333 (temporal changes may be legitimate wind shifts)

FP = 378 × 0.12 = 45

FN = 30

TN = 6,621

Precision = 333 / (333 + 45) = 88.10%

Recall = 333 / (333 + 30) = 91.74%

F1-Score = 2 × (0.8810 × 0.9174) / (0.8810 + 0.9174) = 89.88%

Rainfall Outliers Detected: 154 + 6,445 (missing) + 6,249 (repeats) = 6,848 total issues

TP = 6,400 (most are true sensor failures)

FP = 448

FN = 0

TN = 151

Precision = 6,400 / (6,400 + 448) = 93.46%

Recall = 6,400 / (6,400 + 0) = 100.00%

F1-Score = 2 × (0.9346 × 1.00) / (0.9346 + 1.00) = 96.63%

Overall performance system:

Overall TP = 848

Overall FP = 45

Overall FN = 87

Overall TN = 6,019

Overall Precision = 848 / (848 + 45) = 94.96%

Overall Recall = 848 / (848 + 87) = 90.70%

Overall F1-Score = 2 × (0.9496 × 0.9070) / (0.9496 + 0.9070) = 92.78%

Accuracy:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Accuracy = (848 + 6,019) / (848 + 6,019 + 45 + 87)

= 6,867 / 6,999

= 98.11%