

1) Table 1

| Day | Ground Truth | Predictions |
|-----|--------------|-------------|
| 1 | Shower | Shower |
| 2 | Clear | Shower |
| 3 | Shower | Clear |
| 4 | S | S |
| 5 | C | S |
| 6 | S | S |
| 7 | C | S |
| 8 | C | C |
| 9 | C | C |
| 10 | S | S |

Confusion Matrix

| | | |
|----------------|----------------|------------|
| | C ₁ | $\neg C_1$ |
| C ₁ | 4 (TP) | 1 (FN) |
| $\neg C_1$ | 3 (FP) | 2 (TN) |

$$\text{Accuracy} = \frac{6}{10} = 0.60$$

$$\text{Precision} = \frac{4}{7} = 0.571$$

$$\text{Recall} = \frac{4}{5} = 0.80$$

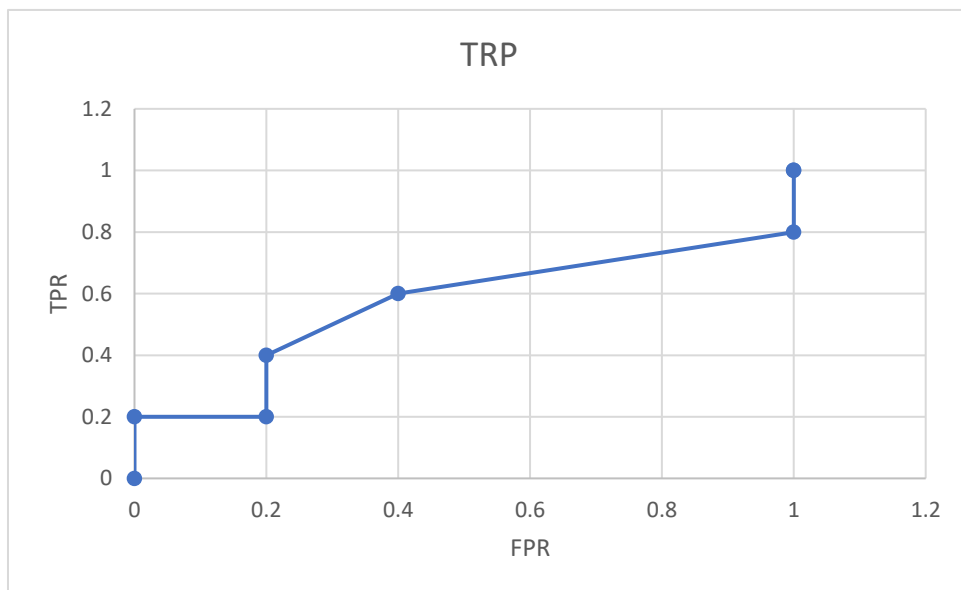
2) Table 2

| Day | Ground Truth | Predictions |
|-----|--------------|-------------|
| 1 | S | 0.95 |
| 2 | C | 0.85 |
| 3 | S | 0.78 |
| 4 | S | 0.66 |
| 5 | C | 0.6 |
| 6 | S | 0.55 |
| 7 | C | 0.53 |
| 8 | C | 0.52 |
| 9 | C | 0.51 |
| 10 | S | 0.4 |

$$F_1 = \frac{2 \times \frac{4}{7} \times \frac{4}{5}}{\frac{4}{7} + \frac{4}{5}} = \frac{2}{3} \approx 0.6667$$

| TP | FP | TN | FN | TJR | FPR |
|-----|-----|-----|----|-----|-----|
| ## | ## | | | 1 | 1 |
| ### | ### | | | 1 | 1 |
| | ### | | 1 | 0.8 | 1 |
| | | | | 0.6 | 0.4 |
| | 1 | | | 0.4 | 0.2 |
| 1 | 1 | | | 0.2 | 0.2 |
| 1 | | ### | | 0.2 | 0 |
| | | ## | ## | 0 | 0 |

| FPR | TRP |
|-----|-----|
| 1 | 1 |
| 1 | 1 |
| 1 | 0.8 |
| 0.4 | 0.6 |
| 0.2 | 0.4 |
| 0.2 | 0.2 |
| 0 | 0.2 |
| 0 | 0 |



$$\text{Difference of RMSE} = \sqrt{0.00890} - \sqrt{0.01326}$$

$$\text{pooled_sd} = \sqrt{\frac{(10-1) \times (0.055) + (10-1) \times (0.1287)}{10+10-2}} = 0.0208$$

③

| | Residuals for R1 | Residuals for R2 | d |
|----|--------------------|------------------|-------|
| 1 | 0.09 | -0.11 | 0.2 |
| 2 | 0.12 | 0.41 | 0.02 |
| 3 | -0.07 | -0.12 | 0.05 |
| 4 | -0.12 | 0.09 | -0.21 |
| 5 | -0.08 | 0.14 | -0.22 |
| 6 | -0.1 | 0.13 | -0.23 |
| 7 | -0.08 | -0.12 | 0.04 |
| 8 | -0.08 | 0.11 | -0.19 |
| 9 | -0.1 | 0.13 | -0.23 |
| 10 | -0.09 | -0.09 | 0 |
| | $\bar{d} = -0.077$ | | |

$$S_d^2 = \frac{1}{K-1} \sum (d_i - \bar{d})^2 \quad K=10$$

$$\begin{aligned} \sum = & (0.2 - (-0.077))^2 + (0.02 - (-0.077))^2 + (0.05 - (-0.077))^2 \\ & + (-0.21 - (-0.077))^2 + (-0.22 - (-0.077))^2 + (-0.23 - (-0.077))^2 \\ & + (0.04 - (-0.077))^2 + (-0.19 - (-0.077))^2 + (-0.23 - (-0.077))^2 \\ & + (0 - (-0.077))^2 = 0.21961 \end{aligned}$$

$$S_d^2 = \frac{0.21961}{9} = 0.02440$$

$$t = \frac{\bar{d}}{\frac{\sqrt{S_d^2}}{K}}$$

$$t = \frac{-0.077}{\sqrt{\frac{0.02440}{910}}} = -1.5588$$

~~$$P\text{-val} = 0.1534$$~~

$$df = 9$$

$$\text{Confidence} = 95\%$$

$$\alpha = \frac{1 - 0.95}{2} = 0.025 \quad \text{critical } t\text{-val} = 2.26$$

Since our t value is not greater than our critical t value, we can't reject the null hypothesis. So regressor 1 is not significantly better or worse than regressor 2.

Tutorial 2

Worksheet 5 Question 2:

Loading and splitting the dataset:

```
from sklearn.model_selection import train_test_split

data = load_breast_cancer()
df = pd.DataFrame(data.data, columns=data.feature_names)
df['target'] = data.target
X = df.drop('target', axis=1)
y = df['target']

train_X, test_X, train_y, test_y = train_test_split(X, y, test_size=0.2, random_state=42)
```

Benchmarking:

Decision Tree:

```
Decision Tree Benchmarking:
Accuracy: 0.9473684210526315
Precision: 0.9488079172289698
Recall: 0.9473684210526315
F1: 0.9468057045386604
AUC-ROC score: 0.9348182115951523
```

Random Forrest:

```
Random Forest Benchmarking:
Accuracy: 0.9649122807017544
Precision: 0.9652053622194477
Recall: 0.9649122807017544
F1: 0.9647382344750767
AUC-ROC score: 0.9580740255486406
```

Bagging Classifier:

```
Bagging Benchmarking:
Accuracy: 0.956140350877193
Precision: 0.9560881370091896
Recall: 0.956140350877193
F1: 0.9560357083576897
AUC-ROC score: 0.9510317720275139
```

AdaBoost Classifier:

```
AdaBoost Benchmarking:  
Accuracy: 0.9736842105263158  
Precision: 0.9737190197716513  
Recall: 0.9736842105263158  
F1: 0.9736214250146138  
AUC-ROC score: 0.969701932525385
```

XG Boost Classifier:

```
XGBoost Benchmarking:  
Accuracy: 0.956140350877193  
Precision: 0.9560881370091896  
Recall: 0.956140350877193  
F1: 0.9560357083576897  
AUC-ROC score: 0.9510317720275139
```

On this given dataset and the training and test split, the best performing model in every category was the AdaBoost Classifier.

The code for this bench marking is hosted [here](#)