Concept for Data Mining: Classification Accuracy

- Typical Error Matrix:

TRUE POSITIVE	FALSE POSITIVE
FALSE NEGATIVE	TRUE NEGATIVE

Reference Data

Classified Data

TRAINING DATA (actual classes)		
Class–A	Class–B	Totals
2834 (TP)	173 (FP)	3007
318 (FN)	3103 (TN)	3421
3152	3276	6428

(rows labeled: Class–A, Class–B, Totals)

- Diagonals represent sites classified correctly according to reference data.

- Off-diagonals were misclassified.

Classification Accuracy...

❖ **Overall Accuracy** is essentially tells us out of all of the reference sites what proportion were mapped correctly.

$$\text{Overall Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

❖ **Individual Class Accuracy** Calculated by dividing the number of correctly classified pixels in each category by either the total number of pixels in the corresponding *column; Producer's accuracy,* or *row; User's accuracy.*

➢ **Producer's Accuracy** is the map accuracy from the point of view of the map maker (the producer).

$$\text{Producer's Accuracy (Class A)} = \frac{TP}{(TP+FN)}$$
$$\text{Producer's Accuracy (Class B)} = \frac{TN}{(FP+TN)}$$

➢ **User's Accuracy** is the accuracy from the point of view of a map user, not the map maker. The User's accuracy essentially tells use how often the class on the map will actually be present on the ground. This is referred to as reliability.

$$\text{User's Accuracy (Class A)} = \frac{TP}{(TP+FP)}$$
$$\text{User's Accuracy (Class B)} = \frac{TN}{(TN+FN)}$$

Classification Accuracy...

	Class-A	Class-B	Totals
Class-A	2834 (TP)	173 (FP)	3007
Class-B	318 (FN)	3103 (TN)	3421
Totals	3152	3276	6428

Overall Accuracy = (TP+TN)/(TP+TN+FP+FN)
Producer's Accuracy (Class A) = TP/(TP+FN)
Producer's Accuracy (Class B) = TN/(FP+TN)
User's Accuracy (Class A) = TP/(TP+FP)
User's Accuracy (Class B) = TN/(TN+FN)

Accuracy on preceding slide:
• Overall Accuracy = 92.4%
• Producer's Accuracy (Class A) = 89.9%
• Producer's Accuracy (Class B) = 94.7%
• User's Accuracy (Class A) = 94.2%
• User's Accuracy (Class B) = 90.7%