



Text Processing Model Evaluation

How do we evaluate a text classification model?

| | correct | not correct |
|--------------|---------|-------------|
| selected | tp | fp |
| not selected | fn | tn |

Is Accuracy a Good Measure?

$$(tp + tn) / (tp + tn + fp + fn)$$

| | correct | not correct |
|--------------|---------|-------------|
| selected | tp | fp |
| not selected | fn | tn |

Is Accuracy a Good Measure?

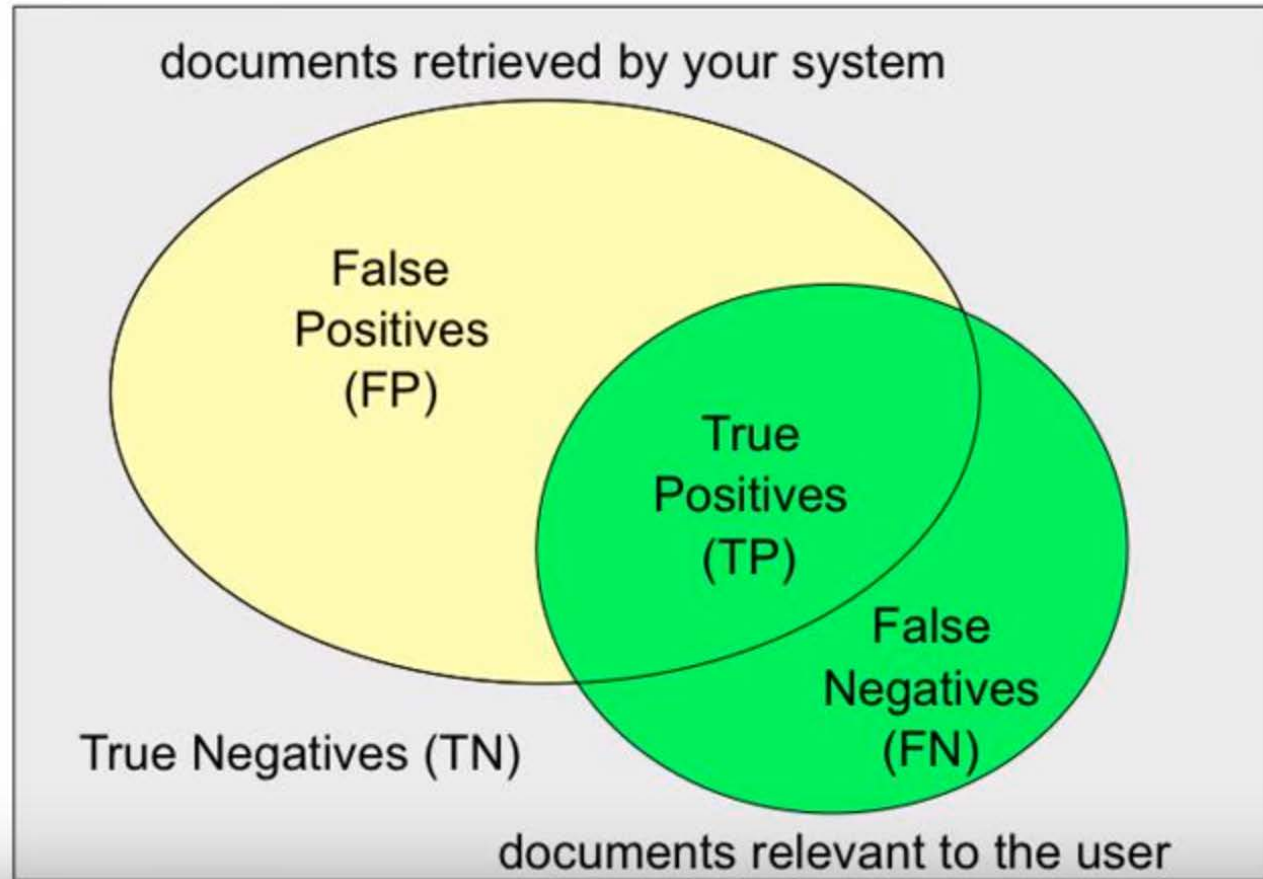
$$(tp + tn) / (tp + tn + fp + fn)$$

| | correct | not correct |
|--------------|---------|-------------|
| selected | 0 | 0 |
| not selected | 10 | 99,990 |

all documents in the test corpus

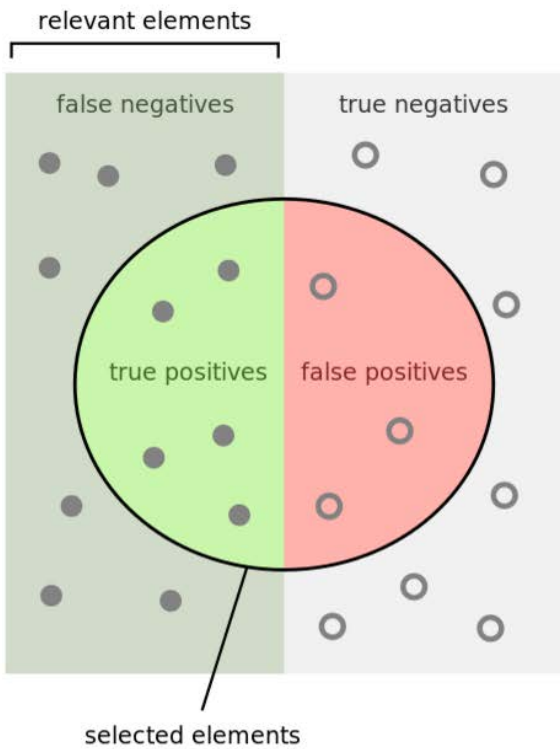
| Relevant? | Retrieved? | |
|-----------|------------|----|
| | Yes | No |
| Yes | TP | FN |
| No | FP | TN |

Confusion matrix /
contingency table



Source: "Evaluation 6: precision and recall," Victor Lavrenko, <https://www.youtube.com/watch?v=mctizdBujk4>

Sources: https://en.wikipedia.org/wiki/Precision_and_recall; "6 - 7 - Precision, Recall, and the F measure - Stanford NLP," Professor Dan Jurafsky & Chris Manning, <https://www.youtube.com/watch?v=2akd6uwtowc>



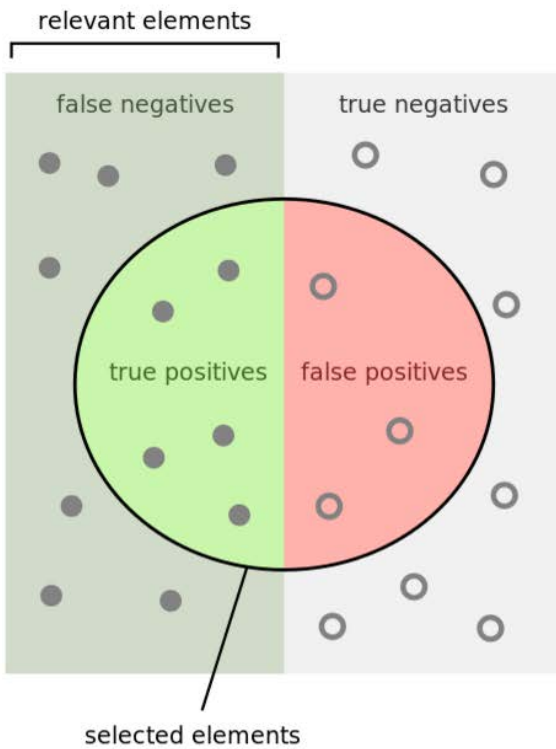
How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$



To Do: Calculate Precision, Recall, and F measure

selected
not selected

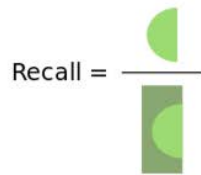
| | correct | not correct |
|--------------|---------|-------------|
| selected | 0 | 0 |
| not selected | 10 | 99,990 |

How many selected
items are relevant?



Precision =

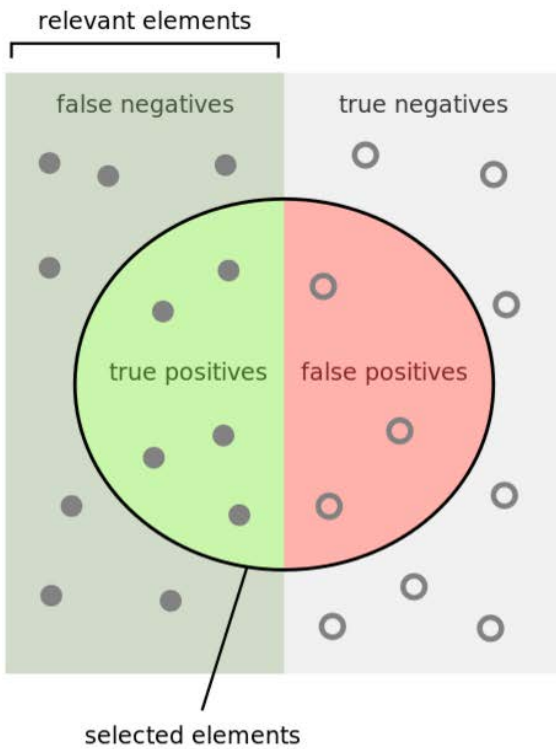
How many relevant
items are selected?



Recall =

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Sources: https://en.wikipedia.org/wiki/Precision_and_recall; "6 - 7 - Precision, Recall, and the F measure - Stanford NLP," Professor Dan Jurafsky & Chris Manning, <https://www.youtube.com/watch?v=2akd6uwtowc>

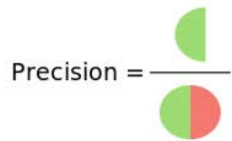


To Do: Calculate Precision, Recall, and F measure

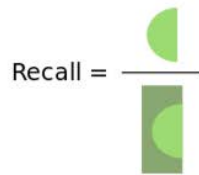
selected
not selected

| | correct | not correct |
|--------------|---------|-------------|
| selected | 8 | 32 |
| not selected | 2 | 99,960 |

How many selected
items are relevant?



How many relevant
items are selected?



$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Sources: https://en.wikipedia.org/wiki/Precision_and_recall; "6 - 7 - Precision, Recall, and the F measure - Stanford NLP," Professor Dan Jurafsky & Chris Manning, <https://www.youtube.com/watch?v=2akd6uwtowc>

Demonstration: RTextTools