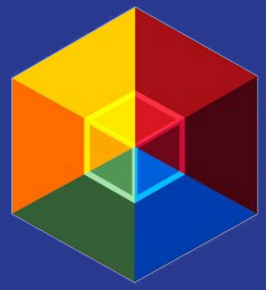
Evaluation Measurement in Machine Learning

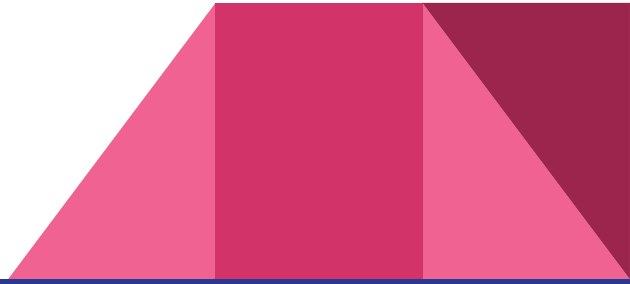
Tim Analisis
Oct 13 2016

KMKLabs - Senayan City

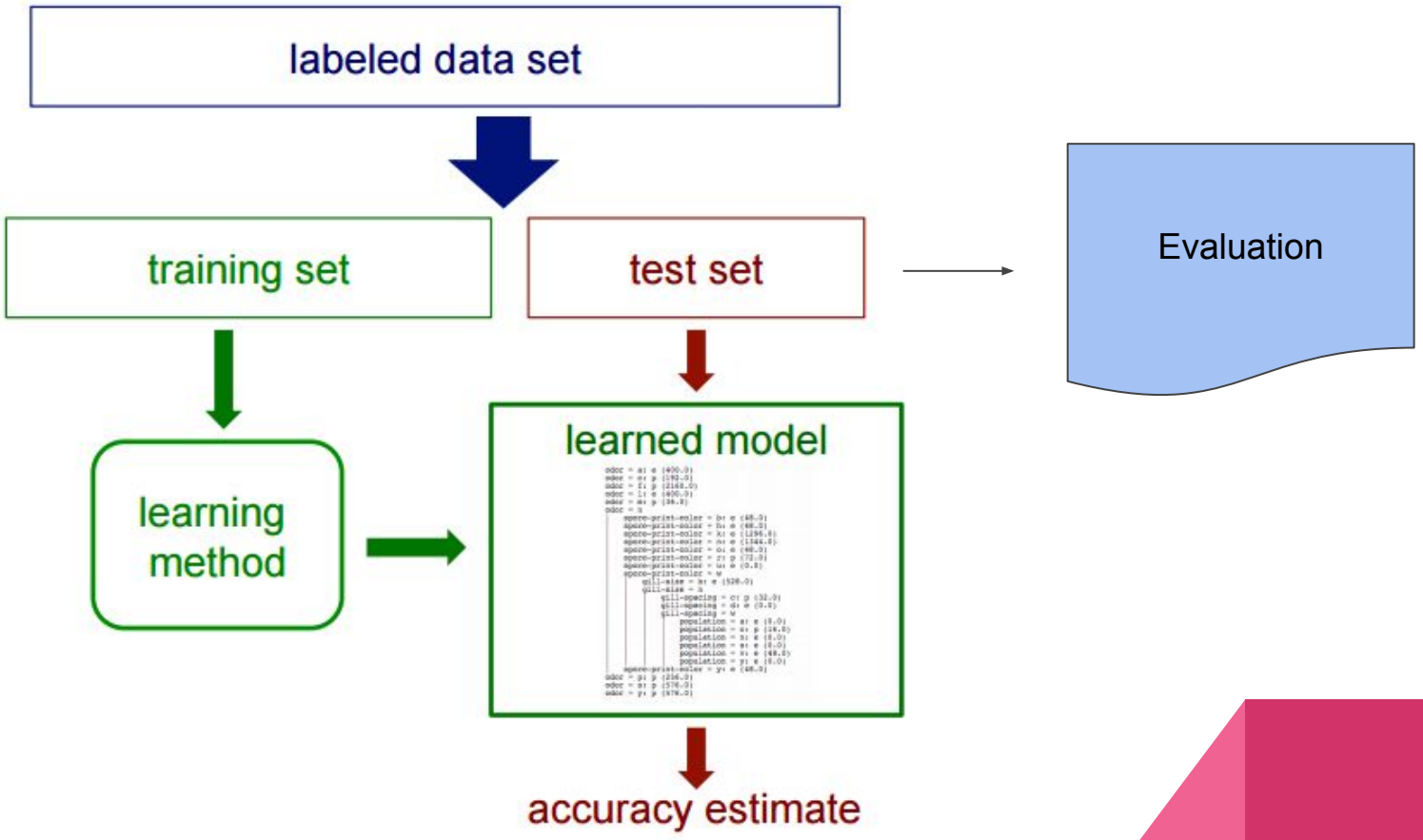


Outline

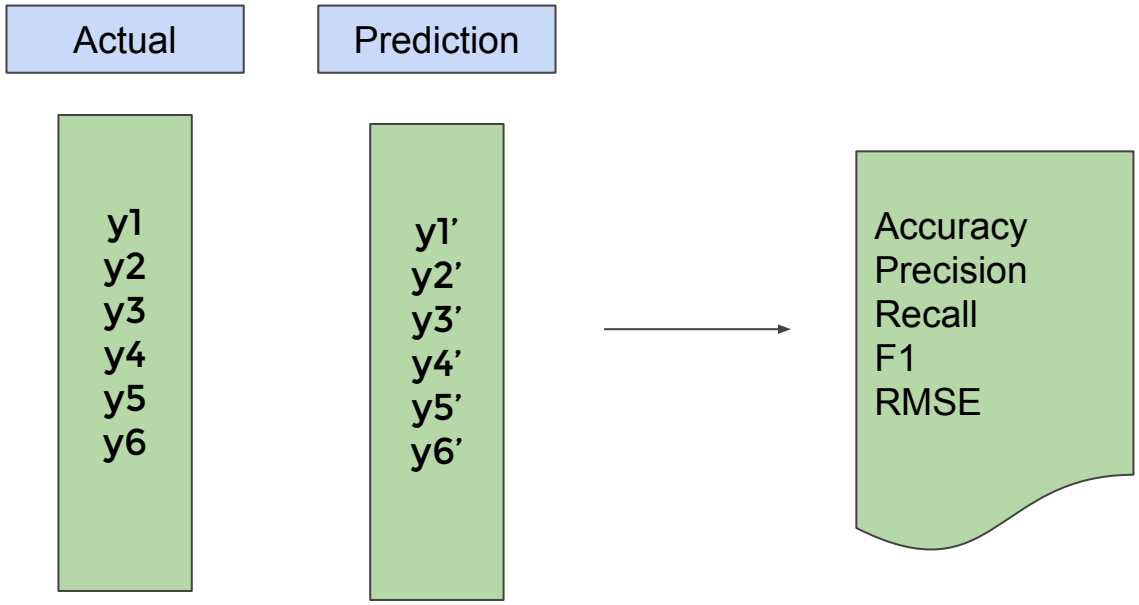
1. Overview of Machine Learning Procedure
2. Cross-validation
3. Confusion Metrics
4. Accuracy, Precision, Recall
5. Imbalance dataset measurement
6. Measurement in Data Construction
7. HomeWork



1. Overview of Machine Learning Procedure

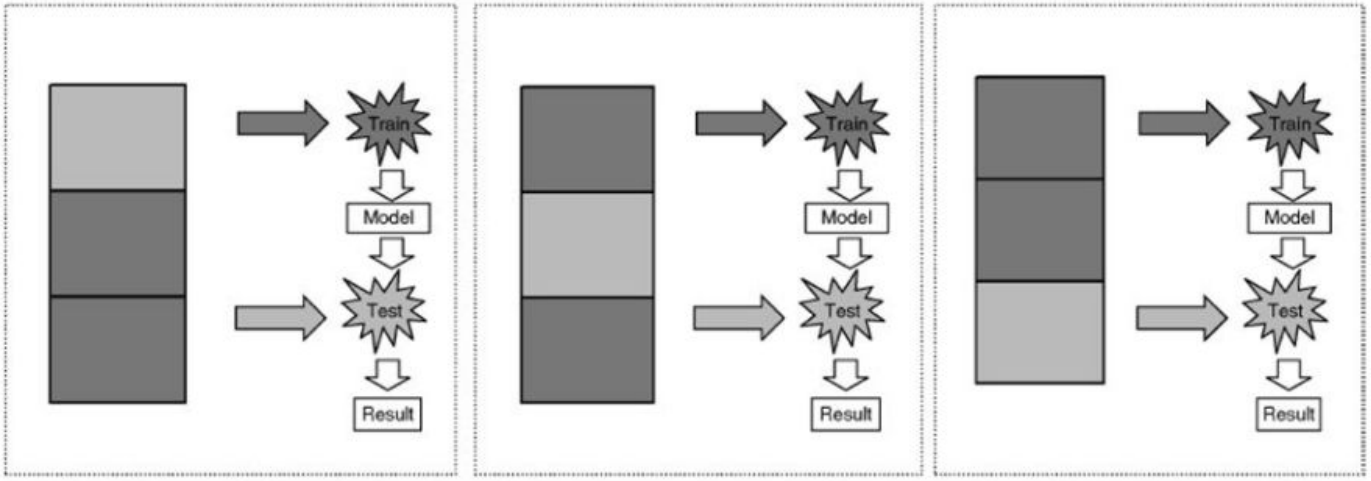


1. Overview of Machine Learning Procedure

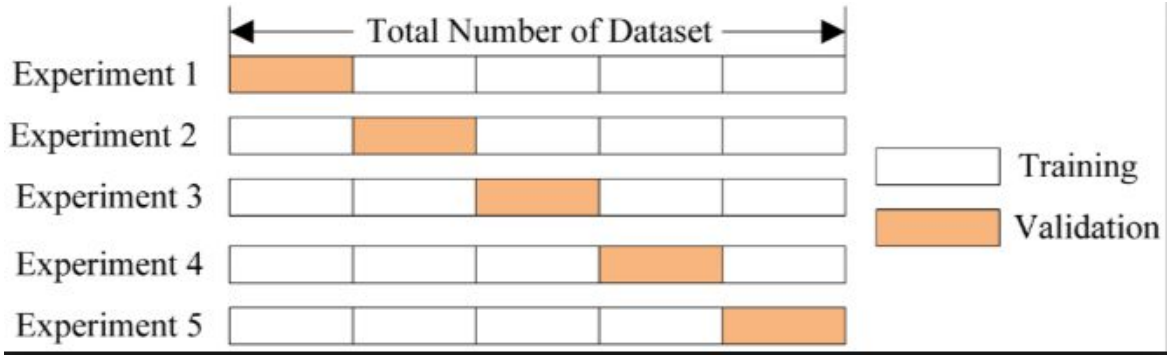


- *Comparing between actual label with prediction label*
- *Evaluation bisa dilakukan pada supervised ataupun unsupervised classification, selama dataset yang ada memiliki label / anotasi*

2. K-fold Cross validation



Cross-Validation. Figure 1. Procedure of three-fold cross-validation.



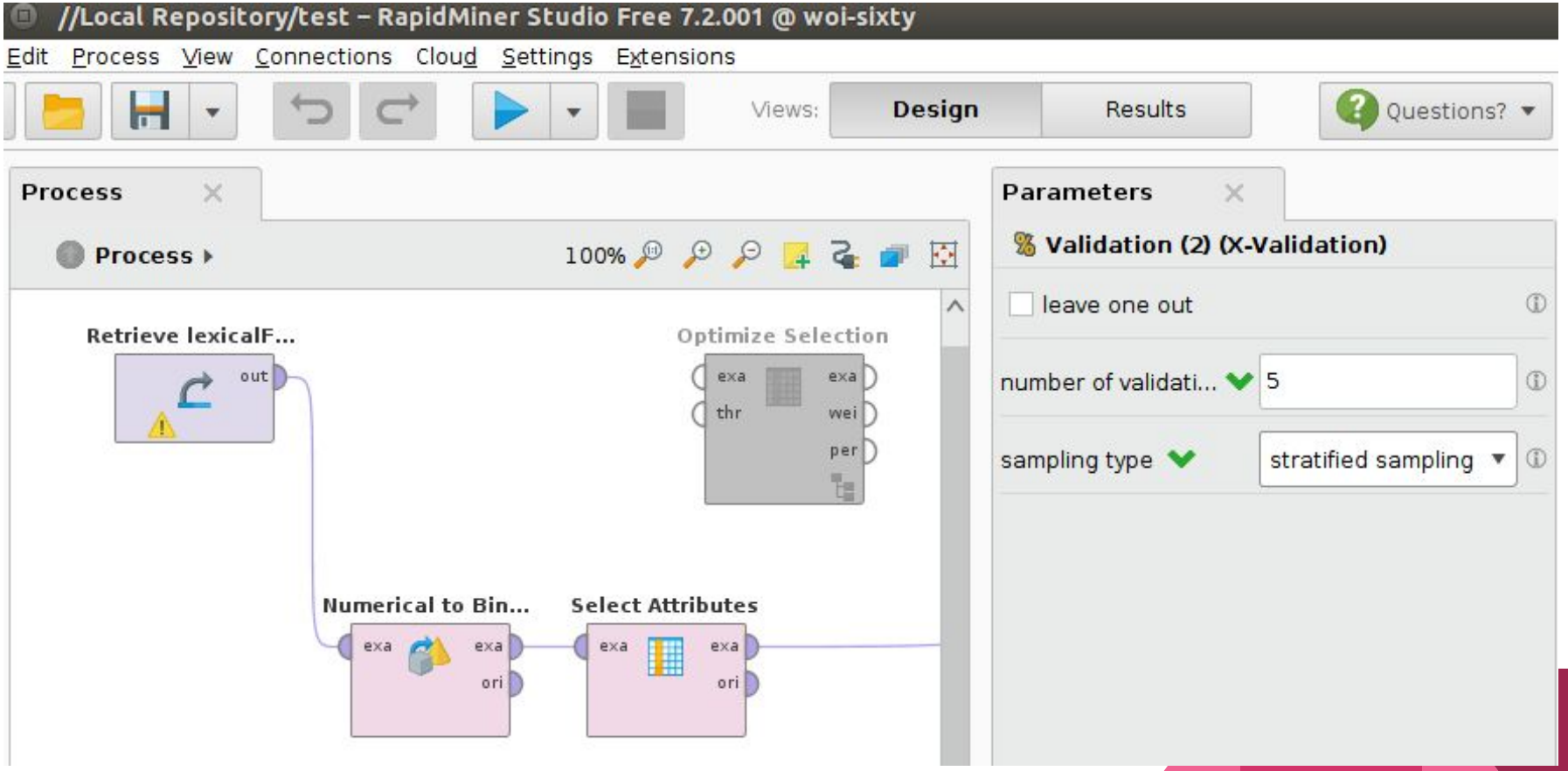
2. K-fold Cross validation

Suppose we have 100 instances, and we want to estimate accuracy with cross validation

iteration	train on	test on	correct
1	s_2 s_3 s_4 s_5	s_1	11 / 20
2	s_1 s_3 s_4 s_5	s_2	17 / 20
3	s_1 s_2 s_4 s_5	s_3	16 / 20
4	s_1 s_2 s_3 s_5	s_4	13 / 20
5	s_1 s_2 s_3 s_4	s_5	16 / 20

accuracy = 73/100 = 73%

2. K-fold Cross validation in Rapid Miner



2. K-fold Cross validation - Sampling

1. Linear Sampling

The Linear sampling simply divides the ExampleSet into partitions without changing the order of the examples i.e. subsets with consecutive examples are created.

2. Shuffle Sampling

The Shuffled sampling builds random subsets of the ExampleSet. Examples are chosen randomly for making subsets.

3. **Stratified Sampling** ✓

The Stratified sampling builds random subsets and ensures that the class distribution in the subsets is the same as in the whole ExampleSet.

2. K-fold Cross validation - Sampling

3.1.1. Computing cross-validated metrics

The simplest way to use cross-validation is to call the `cross_val_score` helper function on the estimator and the dataset.

The following example demonstrates how to estimate the accuracy of a linear kernel support vector machine on the iris dataset by splitting the data, fitting a model and computing the score 5 consecutive times (with different splits each time):

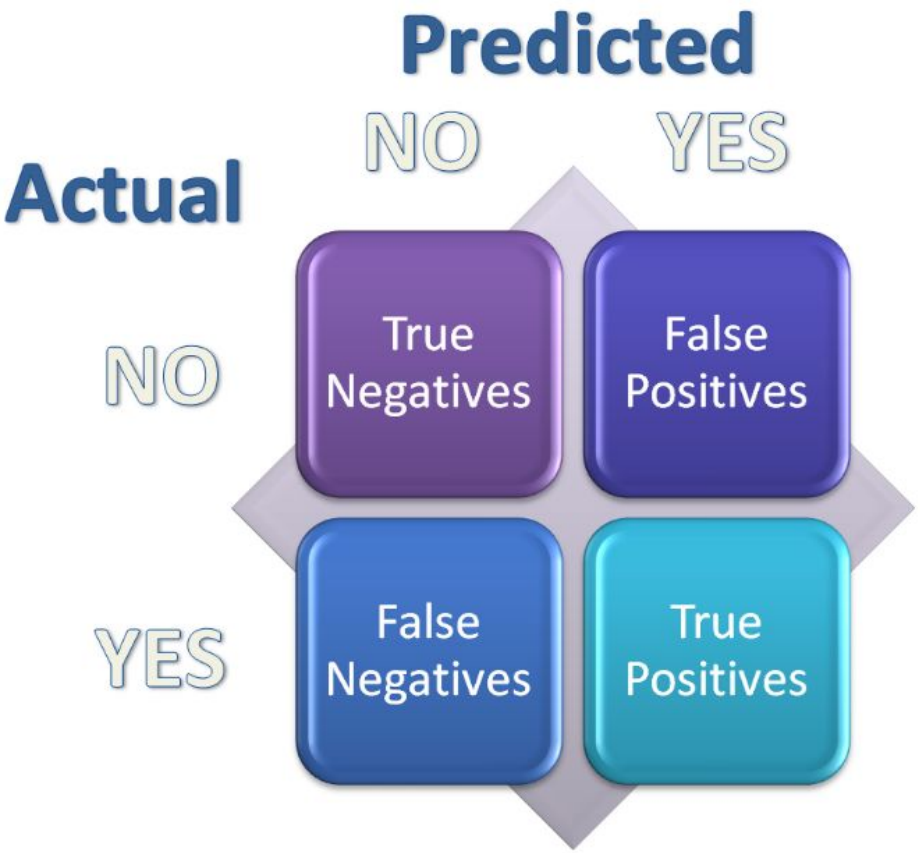
```
>>> from sklearn.model_selection import cross_val_score
>>> clf = svm.SVC(kernel='linear', C=1)
>>> scores = cross_val_score(clf, iris.data, iris.target, cv=5)
>>> scores
array([ 0.96...,  1.  ...,  0.96...,  0.96...,  1.        ])
```

The mean score and the 95% confidence interval of the score estimate are hence given by:

```
>>> print("Accuracy: %0.2f (+/- %0.2f)" % (scores.mean(), scores.std() * 2))
Accuracy: 0.98 (+/- 0.03)
```

10-fold cross validation is common, but smaller values of *n* are often used when learning takes a lot of time

3. Confusion Metrics

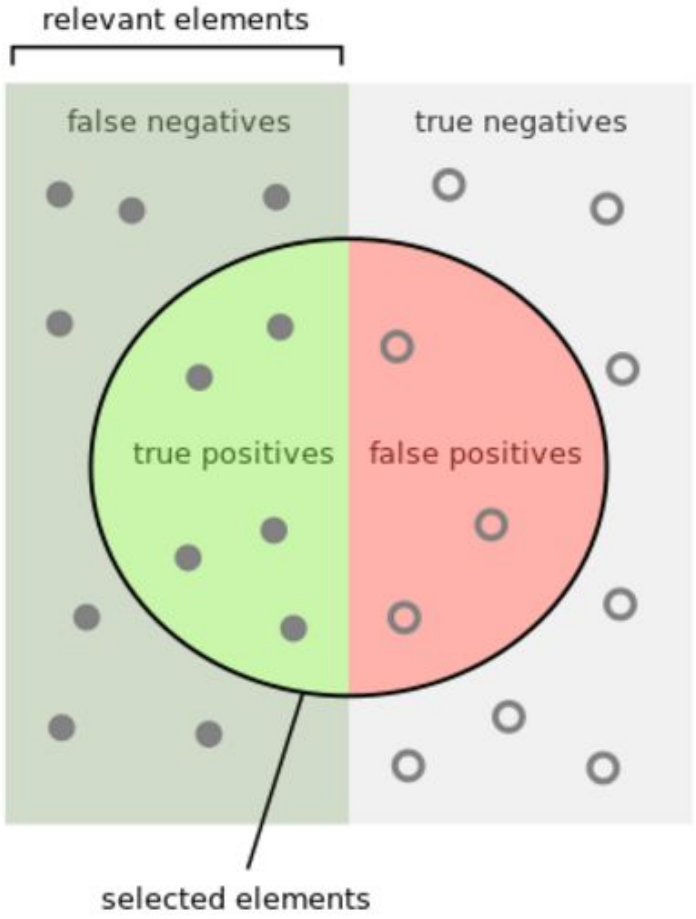


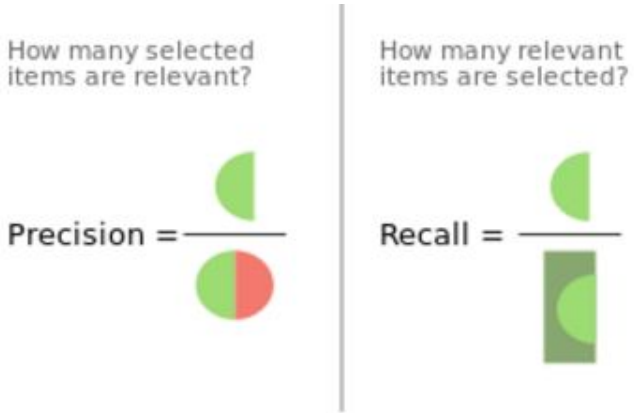
Pred: 1 0 1 0 1 1 1 1

Act : 1 1 1 0 0 1 0 1

		Predicted	
		0	1
Actual	0	TN: 1	FP: 2
	1	FN: 1	TP: 4

4. Precision, Recall, Accuracy



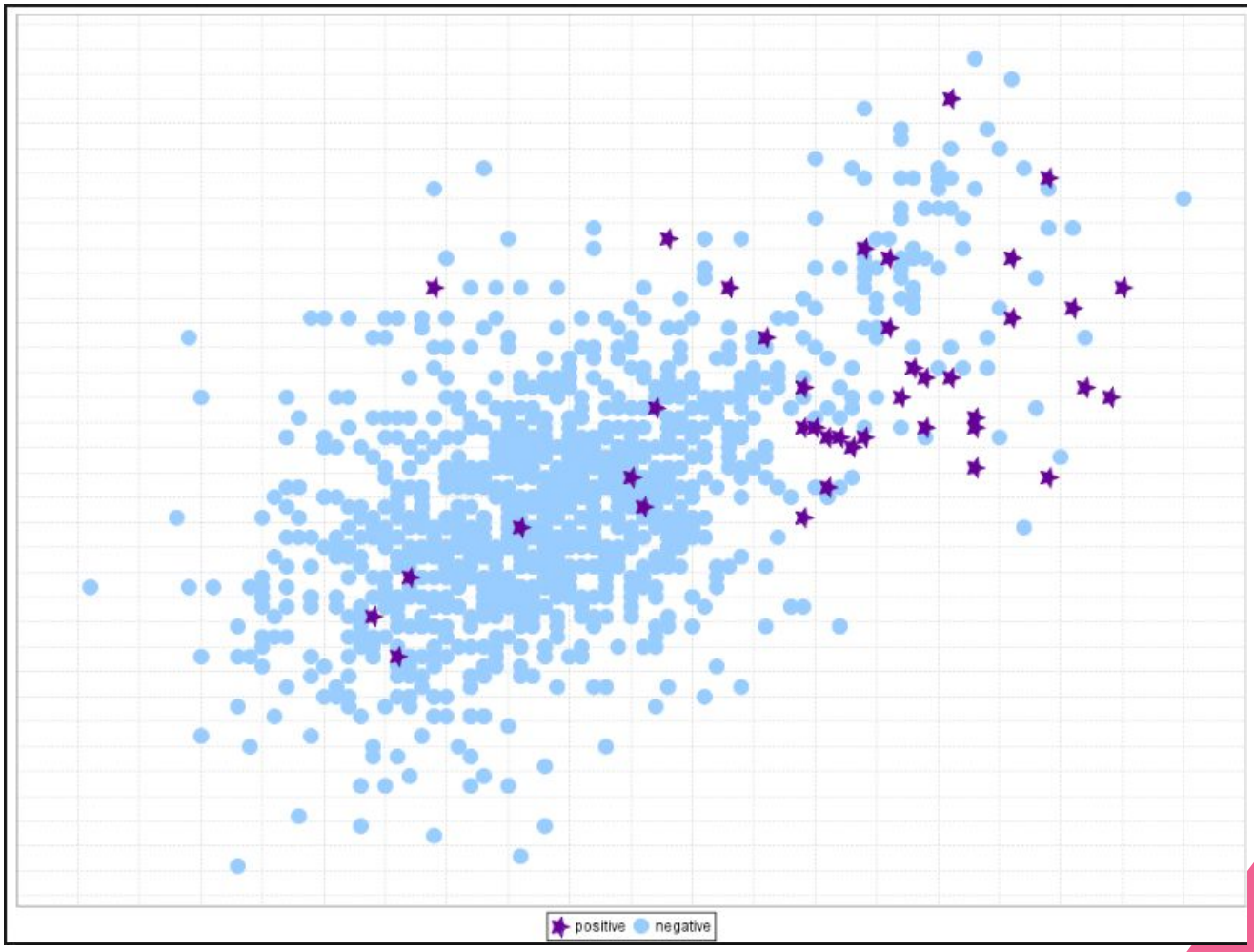


$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

5. Imbalance dataset measurement



I have dataset with ratio 90% vs 10%

After training,
I obtained accuracy 90%

Is it good result?

5. Imbalance dataset measurement

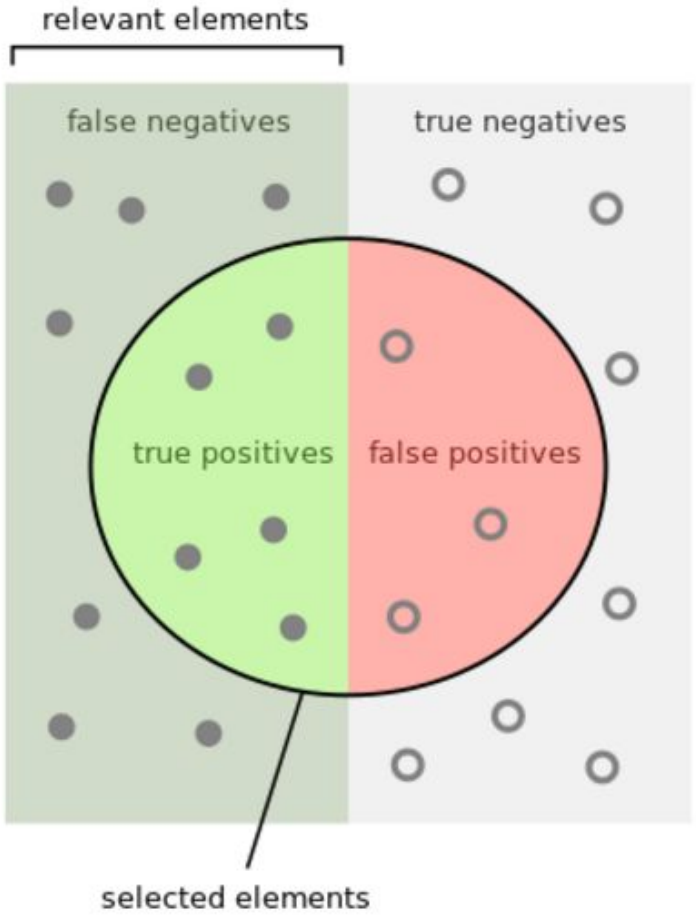
$$Precision = \frac{TP}{TP + FP}$$

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

$$Specificity = \frac{TN}{TN + FP}$$

$$F - Measure = \frac{2 * precision * recall}{precision + recall}$$

$$B - Acc = 0.5 * (Specificity + Sensitivity)$$



6. Measurement in Data Construction

$$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$$

p_o is the relative observed agreement among raters, and ⟶ $p_o = (a + d) / (a+b+c+d)$

p_e is the hypothetical probability of chance agreement,

- Class/ Reader A: $marginal_a = ((a + b)*(a+c)) / (a+b+c+d)$
- Class/ Reader B: $marginal_b = ((c + d)*(b+d)) / (a+b+c+d)$

$p_e = (marginal_a + marginal_b) / (a+b+c+d)$

		B	
		Yes	No
A	Yes	a	b
	No	c	d

6. Measurement in Data Construction

Value of Kappa	Level of Agreement	% of Data that are Reliable
0-.20	None	0-4%
.21-.39	Minimal	4-15%
.40-.59	Weak	15-35%
.60-.79	Moderate	35-63%
.80-.90	Strong	64-81%
Above .90	Almost Perfect	82-100%

7. Homework

Given prediction and actual label of binary classification in a list,

Please measure:

1. Confusion Metrics
2. Accuracy
3. Precision
4. Recall
5. F1-Measure
6. B-ACC

Prediction:

Actual: