UNIVERSITY OF PORTSMOUTH

Python for Data Analysis

Modeling in Python – Evaluation

(TB2 - Week 4)

Atefeh Khazaei

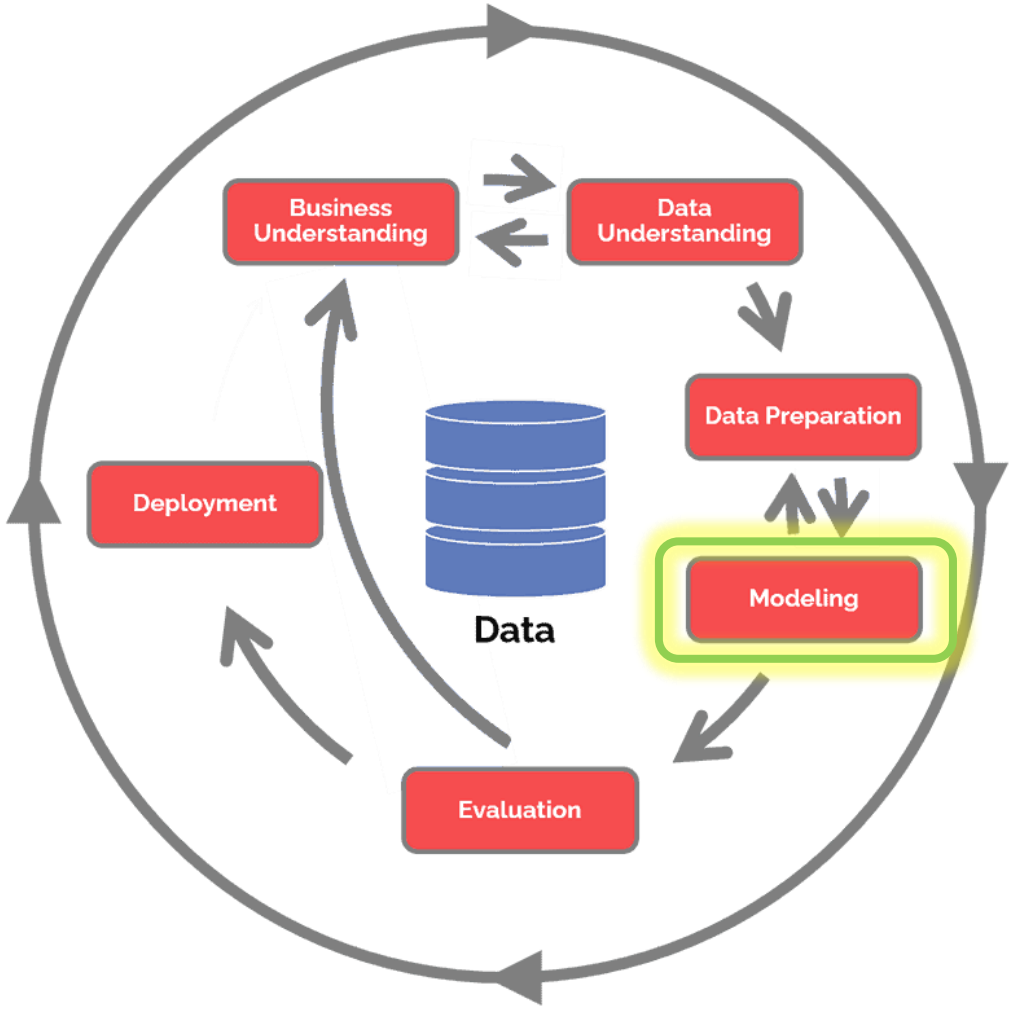
atefeh.khazaei@port.ac.uk

What we will learn this week?

- Evaluation measures
  - Confusion Matrix
  - Precision
  - Recall
  - F-measure

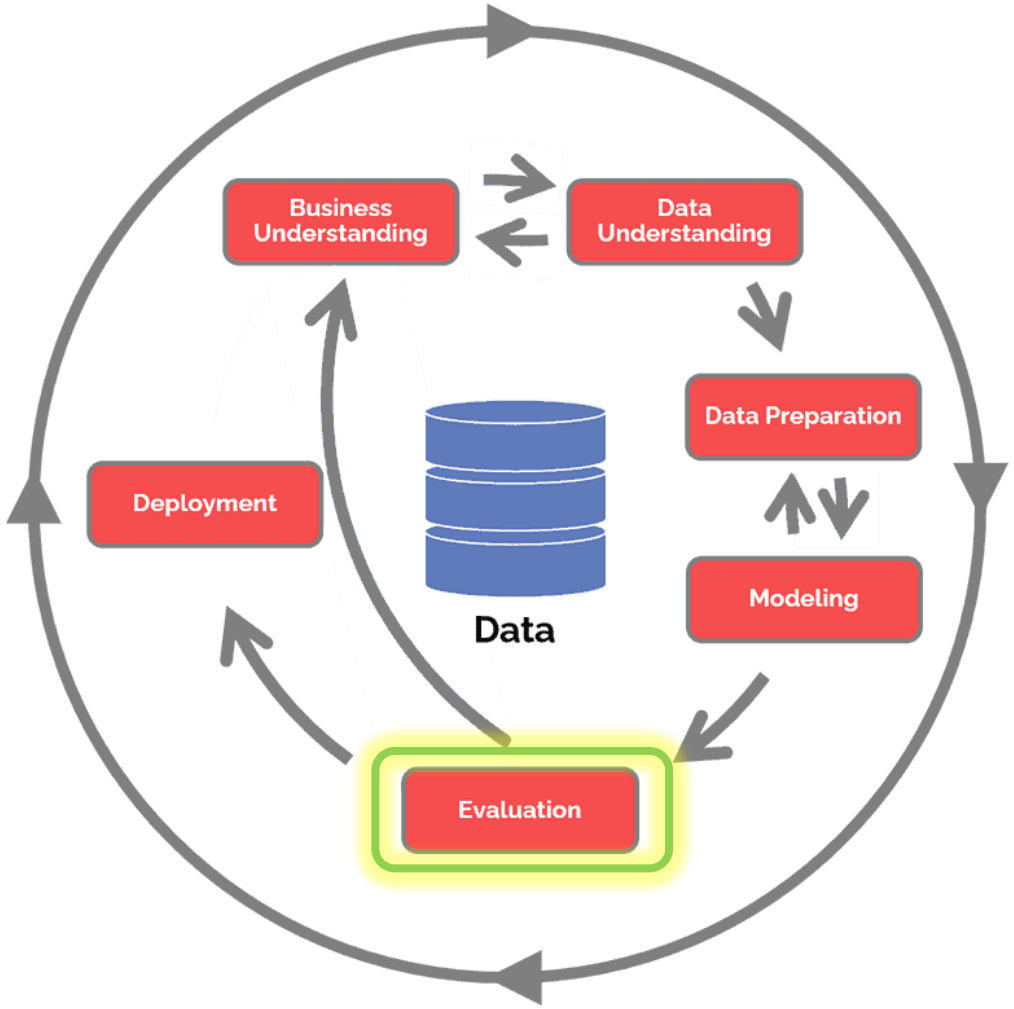
CRISP-DM
Modeling

- K-Nearest Neighbor (KNN)
- Decision Tree
- Random-Forest
- Naive Bayes
- Support Vector Machine (SVM)



CRISP-DM
Evaluation

- Classification
  - Accuracy
  - Precision
  - Recall
  - F-measure
- Regression
  - Mean Absolute Error (MAE)



Evaluation metrics
Confusion matrix

Confusion Matrix

True class \ Predict class	Positive	Negative
Positive	TP	FN
Negative	FP	TN

- **TP:** True Positive
- **TN:** True Negative
- **FP:** False Positive (type I error)
- **FN:** False Negative (Type II error)

Evaluation metrics (cont.)
Confusion matrix / Example

Confusion Matrix

Predict class / True class	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Actual Class\Predicted class	cancer = yes	cancer = no	Total
cancer = yes	90	210	300
cancer = no	140	9560	9700
Total	230	9770	10000

Evaluation metrics (cont.)
Accuracy and Error Rate

A\P	C	¬C	
C	**TP**	**FN**	**P**
¬C	**FP**	**TN**	**N**
	P'	**N'**	**All**

Actual Class\Predicted class	cancer = yes	cancer = no	Total
cancer = yes	90	210	300
cancer = no	140	9560	9700
Total	230	9770	10000

Accuracy = 0/965 ; Error-rate = 0/035

- **Accuracy**, or recognition rate: percentage of test set tuples that are correctly classified

  Accuracy = (TP + TN)/All

- **Error rate**: 1 – accuracy, or

  Error rate = (FP + FN)/All

- **Class Imbalance Problem**:
  - One class may be *rare*, e.g. fraud, or HIV-positive
  - Significant *majority of the negative class* and minority of the positive class

Evaluation metrics (cont.)
Precision, Recall, and F-measure

A\P	C	¬C	
C	**TP**	**FN**	**P**
¬C	**FP**	**TN**	**N**
	P'	**N'**	**All**

- **Precision**: exactness – what % of tuples that the classifier labeled as positive are actually positive

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

- **Recall:** completeness – what % of positive tuples did the classifier label as positive?

$$recall = \frac{TP}{TP + FN}$$

- Perfect score is 1.0
- Inverse relationship between precision & recall

Evaluation metrics (cont.)
Precision, Recall, and F-measure

- ***F* measure (*F_1* or *F*-score)**: harmonic mean of precision and recall,

$$F = \frac{2 \times precision \times recall}{precision + recall}$$

- ***$F_ß$*:** weighted measure of precision and recall
  - assigns ß times as much weight to recall as to precision

$$F_\beta = \frac{(1 + \beta^2) \times precision \times recall}{\beta^2 \times precision + recall}$$

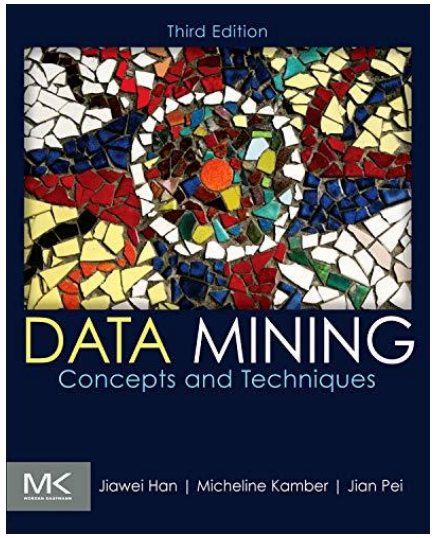
Evaluation metrics (cont.)
Precision, Recall, and F-measure / Example

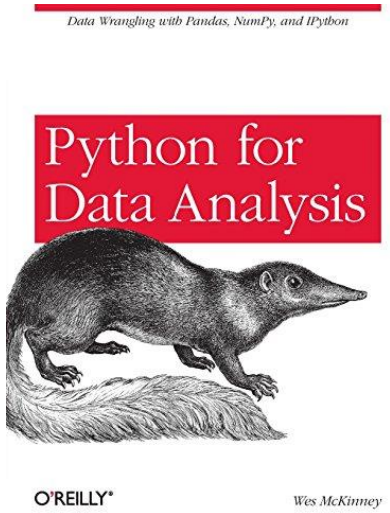
Actual Class\Predicted class	cancer = yes	cancer = no	Total
cancer = yes	90	210	300
cancer = no	140	9560	9700
Total	230	9770	10000

- Precision = 90 / 230 = 39.13 %
- Recall = 90 / 300 = 30 %
- F-measure = (2 × 0.3913 × 0/3) / (0.3913 + 0/3) = 0.3396

References & More Resources

- References:
  - McKinney, Wes. *Python for data analysis: Data wrangling with Pandas, NumPy, and Ipython*, O'Reilly Media, Inc., 2012.
  - Han, Jiawei, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.





Practical Session

- In the previous weeks, you built some models using different algorithms (KNN, Decision Tree, Random Forest, Bayesian, SVM). You only evaluated them using Accuracy.
- Evaluate your models using other evaluation measures (**Precision, Recall, F-measure, Confusion Matrix**) and compare them together.