University of Antwerp

Artificial Neural Networks

Prof. Jose Oramas

TA: Arian Sabaghi, Thomas Dooms, Fabian Denoodt, Benjamin Vandersmissen

Who assist you in the Practical Sessions?

- Arian Sabaghi
- Thomas Dooms
- Fabian Denoodt
- Benjamin Vandersmissen

If you have any questions regarding the Practical Sessions

- Blackboard
- Email

Where you can find us?

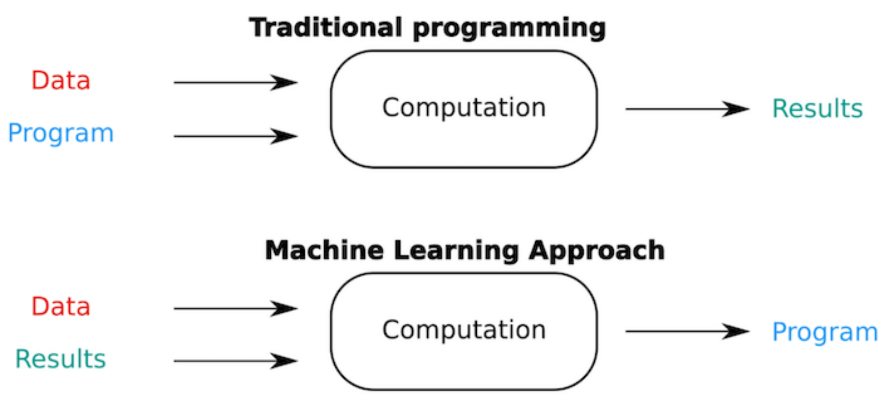
- The Beacon (Sint-Pietersvliet 7, 2000 Antwerp)
- Campus Middelheim

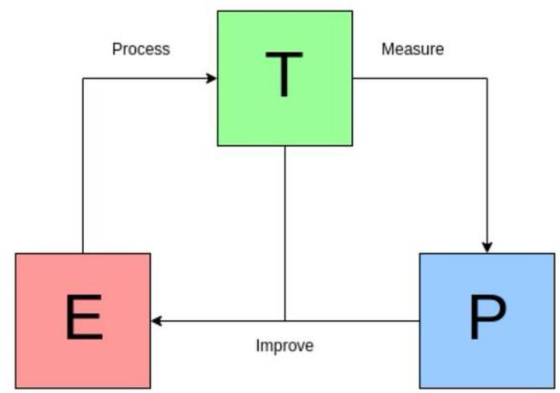
Machine Learning Introduction

University
of Antwerp

Machine Learning Definition

- A part of artificial intelligence concerns the construction and study of systems that can learn from data
- Tom Mitchell: “A computer program is said to learn from **experience E** with respect to some **class of task T** and **performance measure P**, if its performance at tasks in T, as measured by P, improves with experience E”.
- Good Fitting:
  - The model that suitably learns the training dataset and generalizes well to the test data.

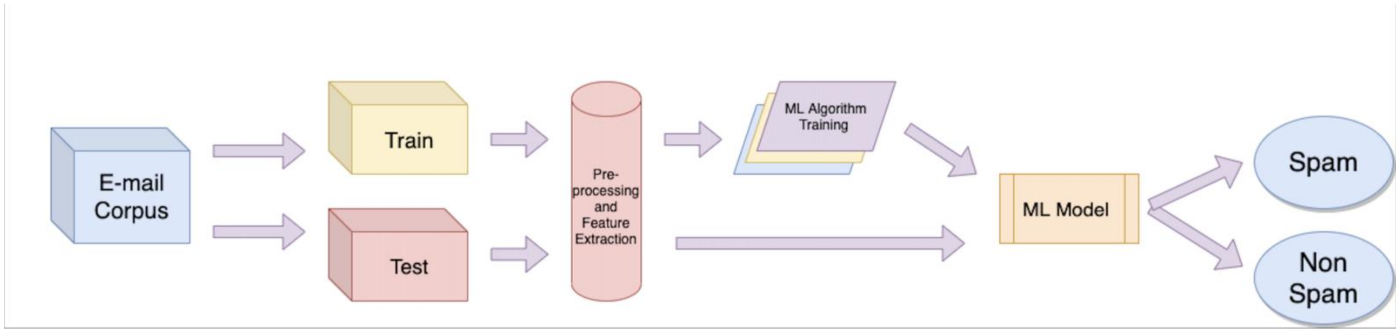




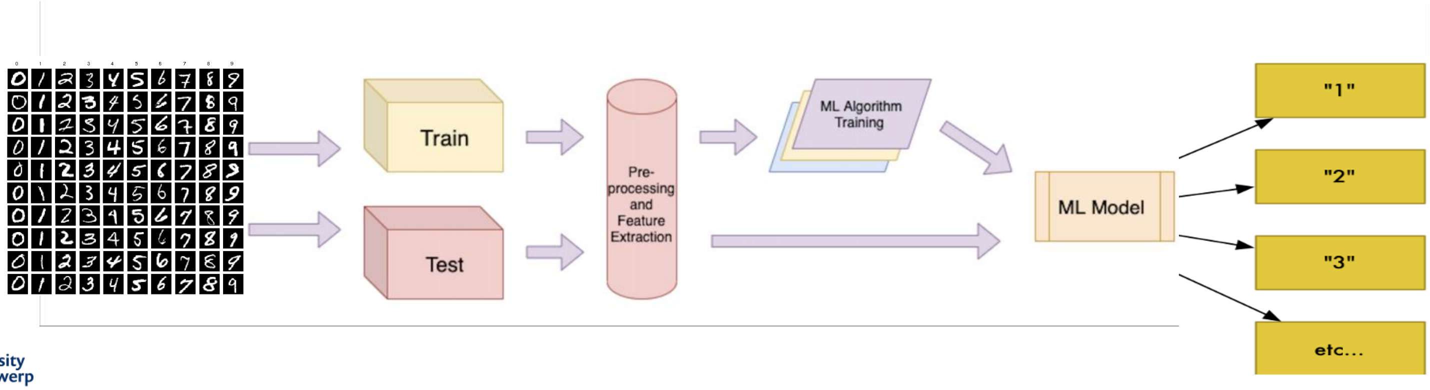
University of Antwerp

Machine Learning Definition: Example

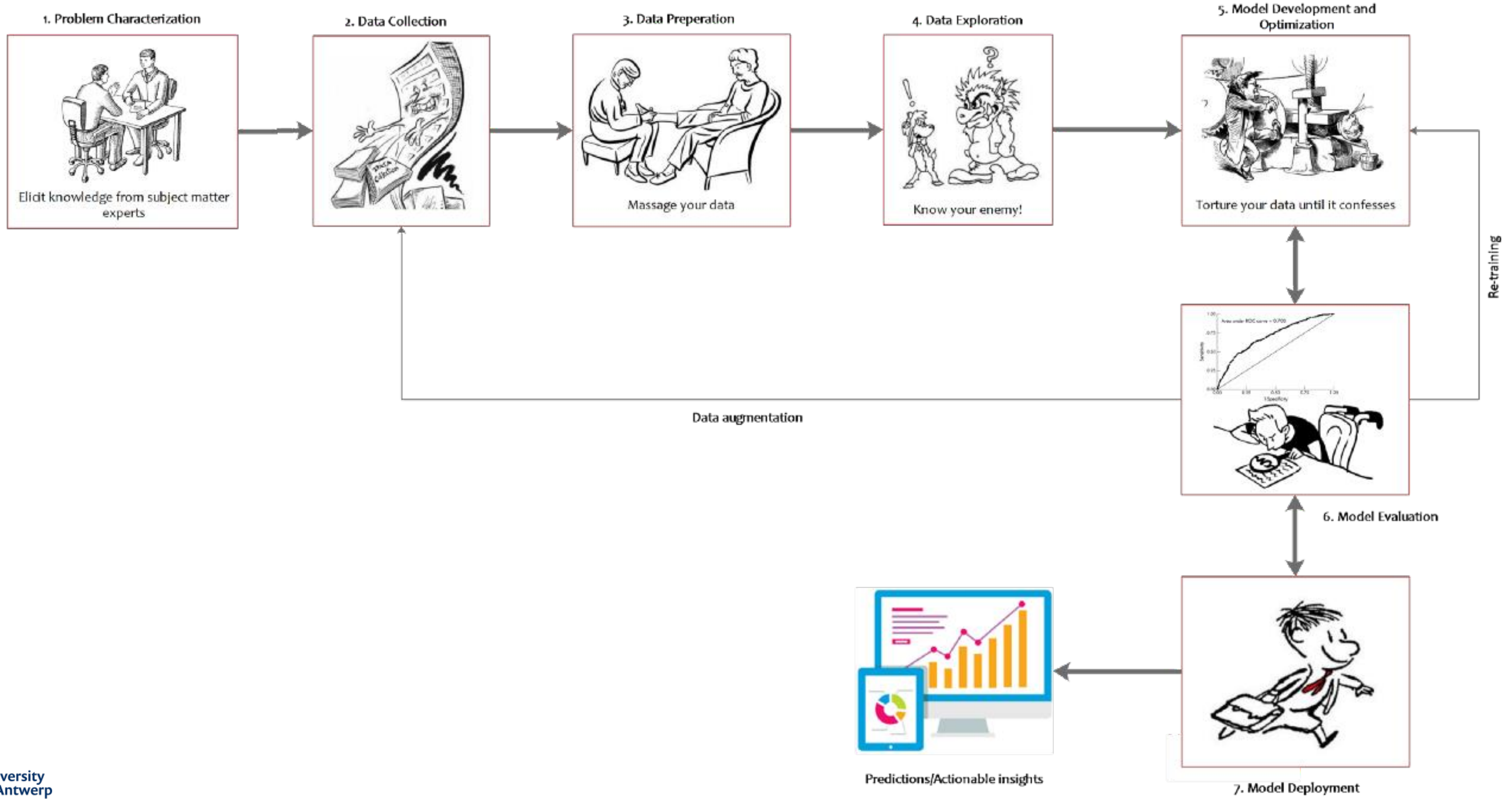
- Example 1: Email Type Recognition System



- Example 2: Handwritten Digit Recognition System

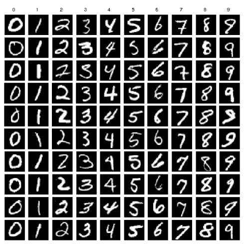


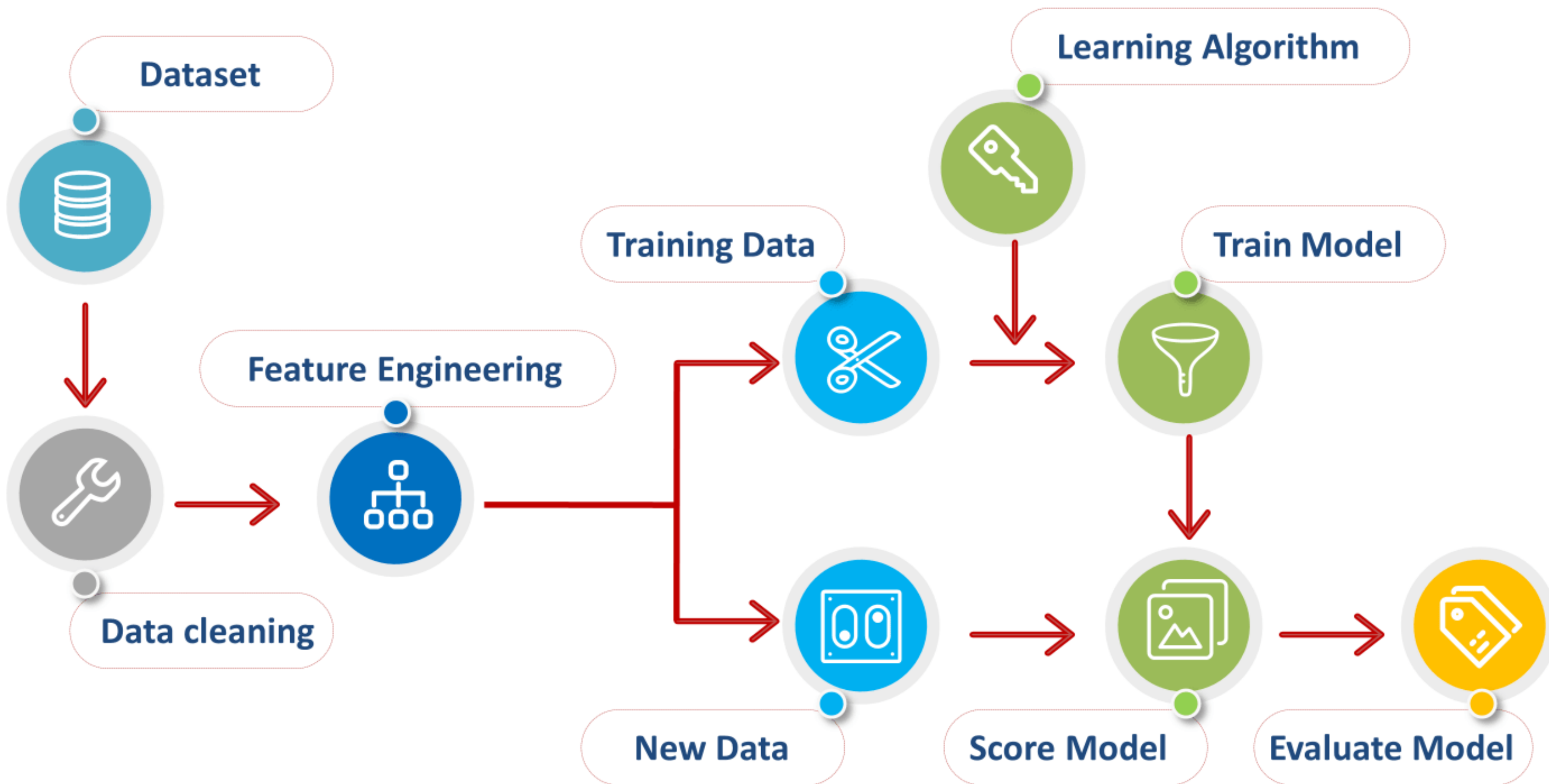
Systematic Procedure of Machine Learning Models



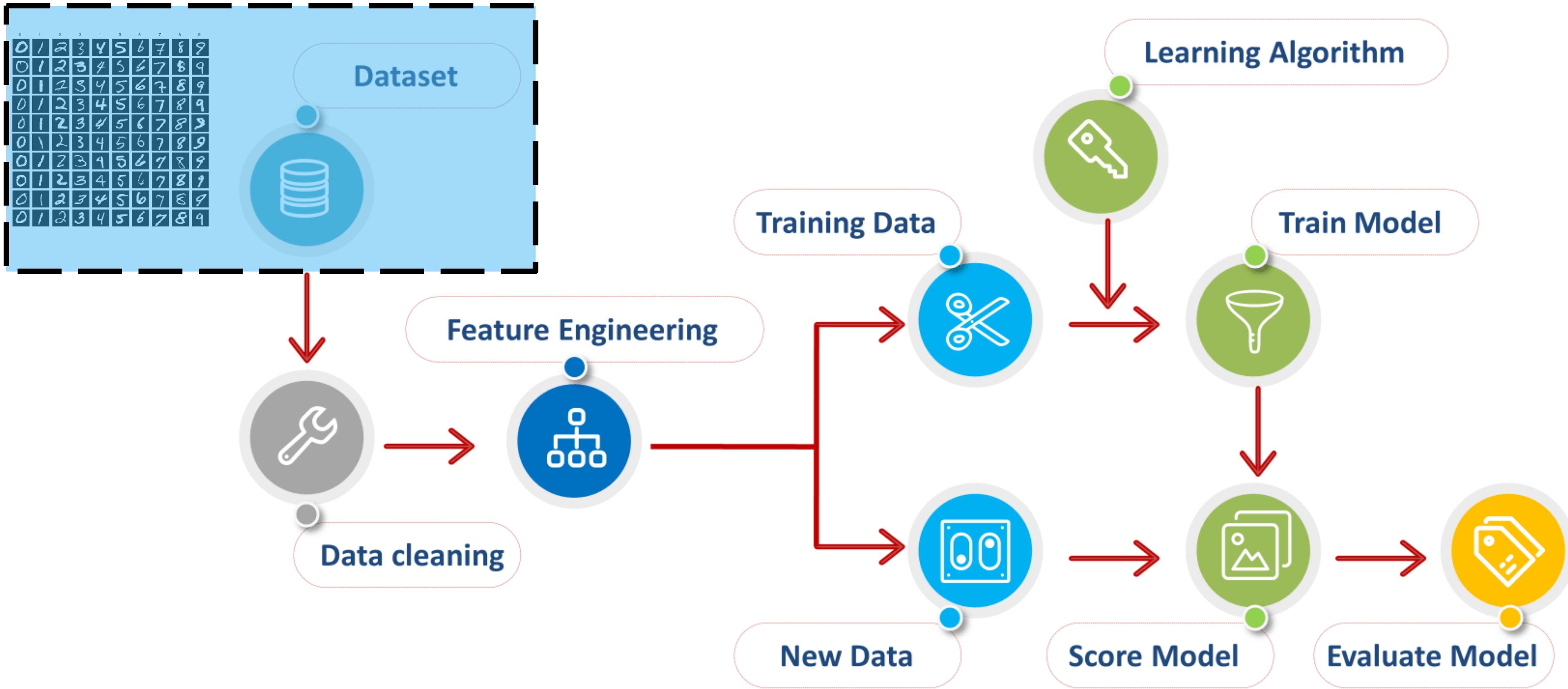
University of Antwerp

Machine Learning Approach





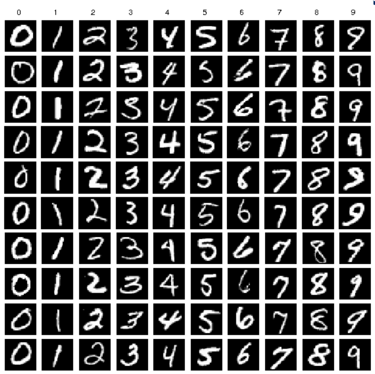
Machine Learning Approach



Data Definition

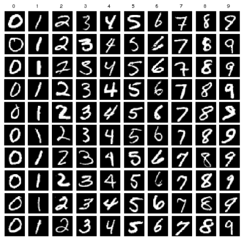
- It can be any unprocessed fact, value, text, sound, or picture that is not being interpreted and analyzed. Data is the most important part of all Data Analytics, Machine Learning, Artificial Intelligence. Without data, we can't train any model and all modern research and automation will go in vain.

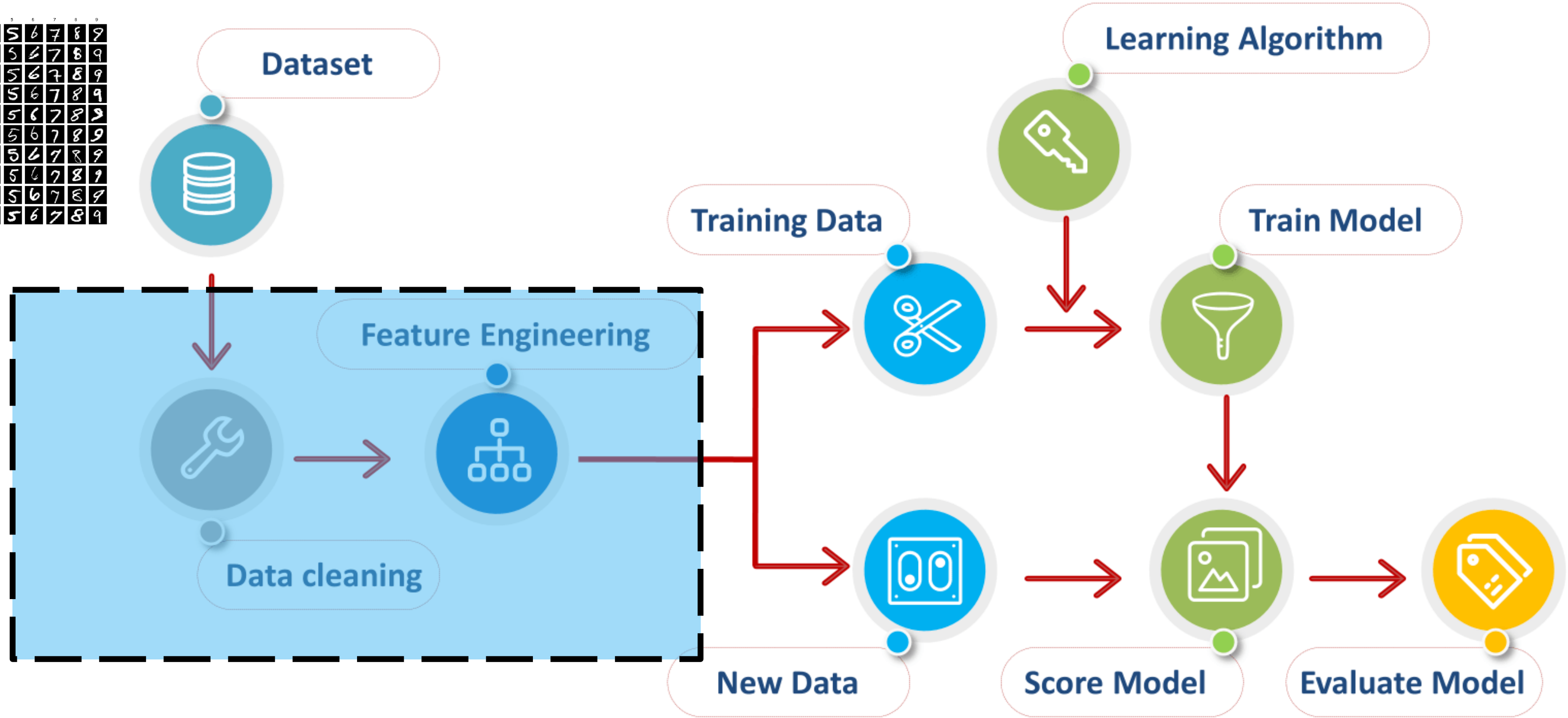
	A	B	C	D
1		Column 1	Column 2	Column 3
2	Row 1	2.2	2.3	1
3	Row 2	2.3	2.6	0
4	Row 3	2.1	2	1



- **Column:** A column describes data of a single type. For example, you could have a column of weights or heights or prices. All the data in one column will have the same scale and have meaning relative to each other.
- **Row:** A row describes a single entity or observation or data point or sample, and the columns describe properties or features about that entity or observation. The more rows you have, the more examples from the problem domain that you have.
- **Cell:** A cell is a single value in a row and column. It may be a real value (1.5) an integer (2) or a category ("red").

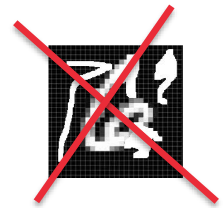
Machine Learning Approach

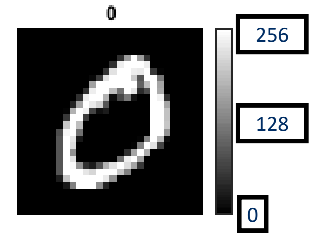


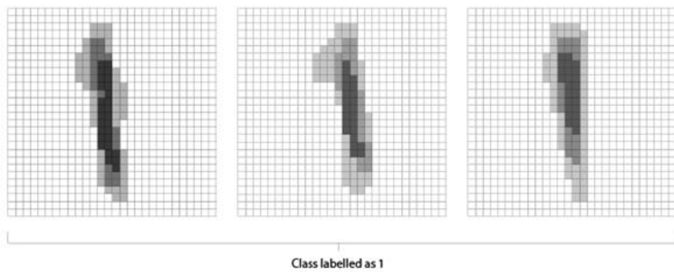


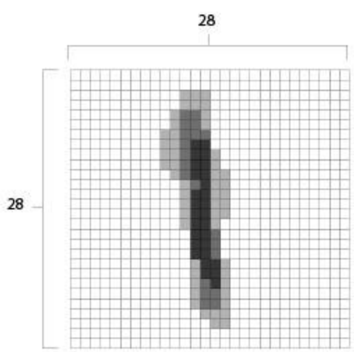
Feature Definition

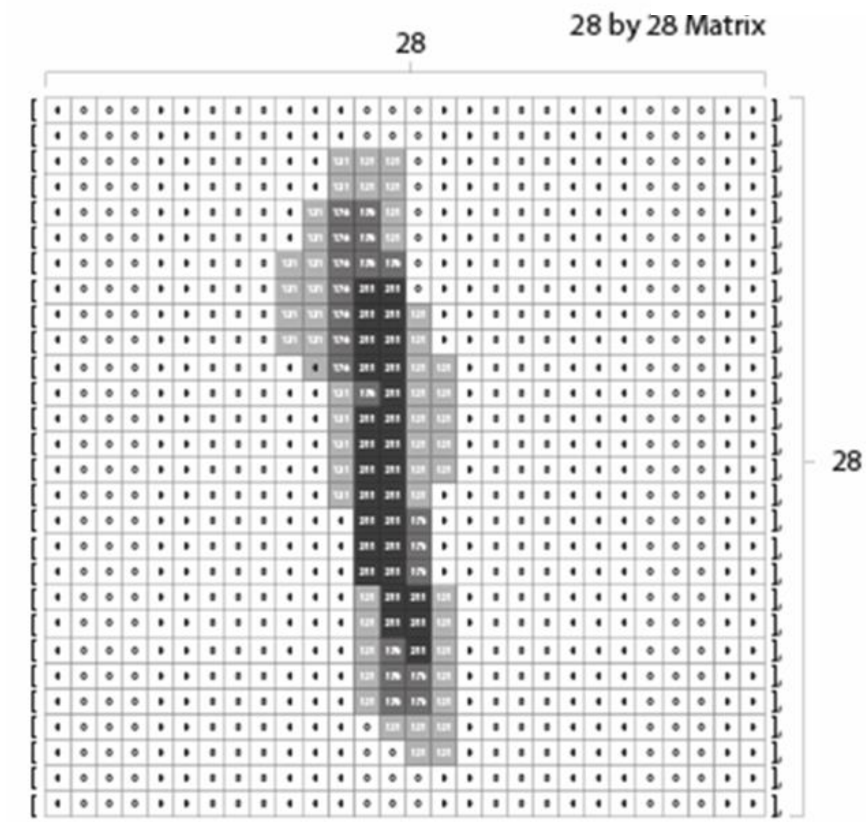
- Feature: is any distinctive aspect, quality or characteristic.









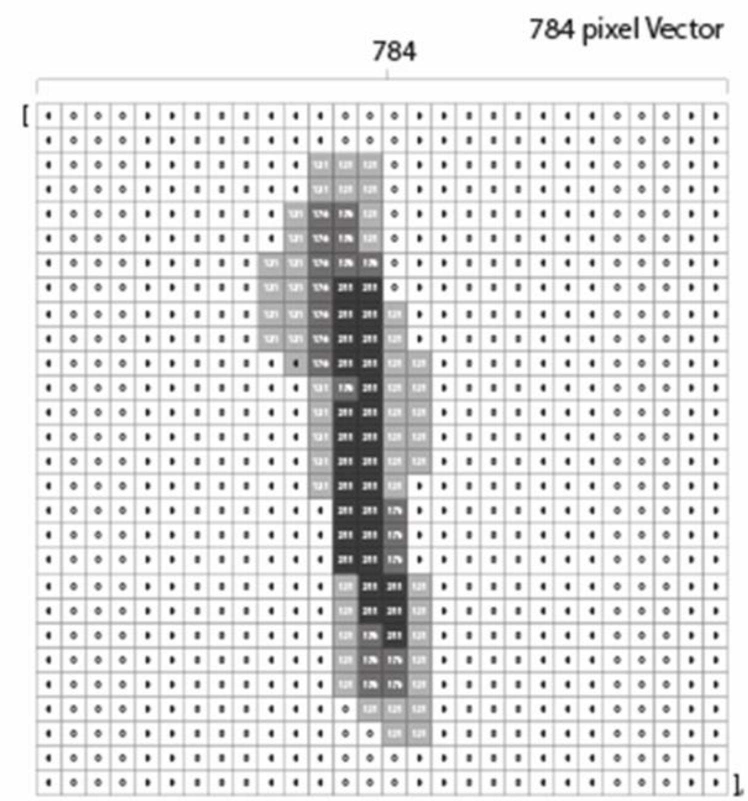


Feature Vector, Feature Space, and Feature Engineering

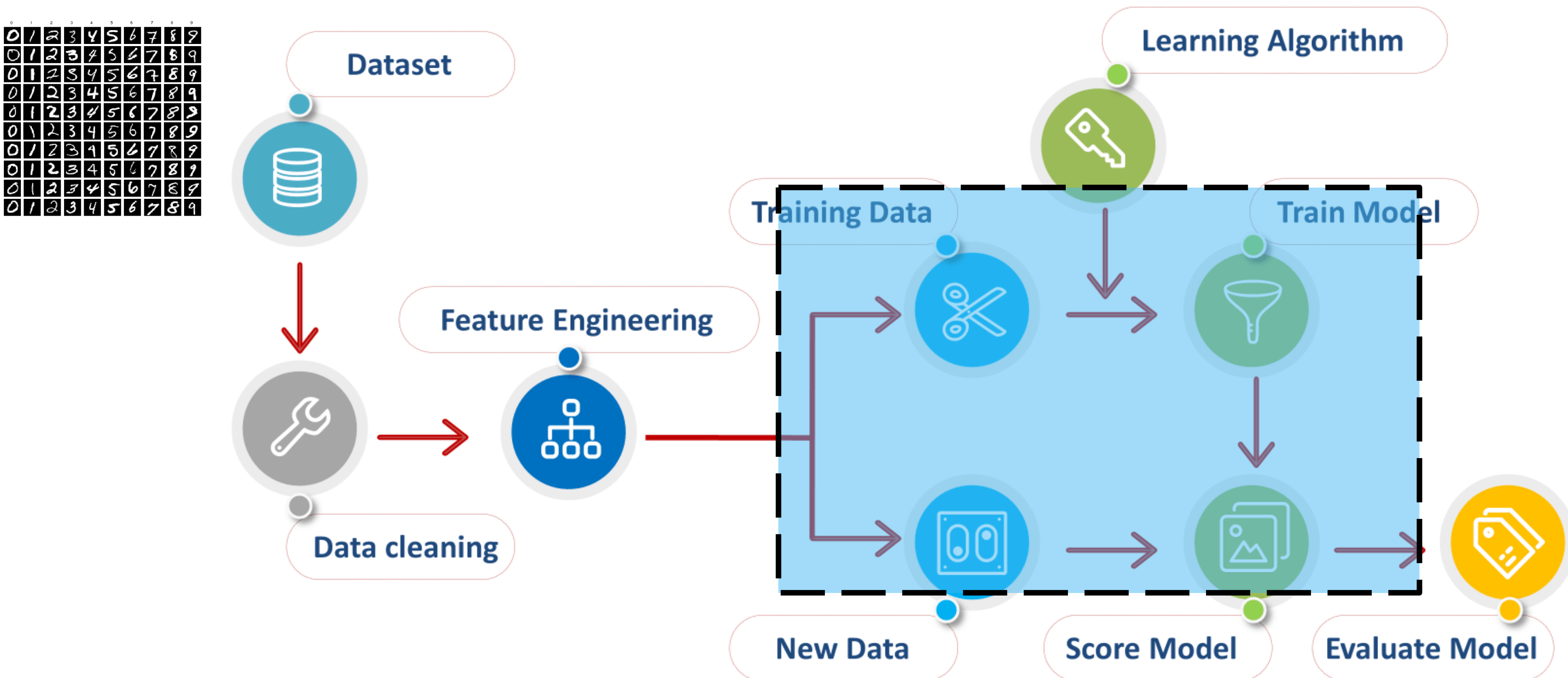
- Feature Vector: The combination of the features represented as a d-dimensional column vector.

$$
\begin{bmatrix}
0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&.6&.8&0&0&0&0&0&0\\
0&0&0&0&0&0&.7&1&0&0&0&0&0&0\\
0&0&0&0&0&0&.7&1&0&0&0&0&0&0\\
0&0&0&0&0&0&.5&1&.4&0&0&0&0&0\\
0&0&0&0&0&0&0&1&.4&0&0&0&0&0\\
0&0&0&0&0&0&0&1&.4&0&0&0&0&0\\
0&0&0&0&0&0&0&1&.7&0&0&0&0&0\\
0&0&0&0&0&0&0&1&1&0&0&0&0&0\\
0&0&0&0&0&0&0&.9&1&0&0&0&0&0\\
0&0&0&0&0&0&0&.3&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0
\end{bmatrix}
$$

	X							y
Index	Feature 1	Feature 2	Feature 3	...	....	...	Feature 784	Class/Label
0	0.00	0.00	0.41	...	...	...	0.00	1
1	0.00	0.00	0.05	...	...	...	0.00	5
2	0.00	0.00	0.71	...	...	...	0.00	7
3	0.00	0.00	0.39	...	...	...	0.00	1
...	...	...	..	...	...	...	...	...
60000	0.00	0.00	0.16	...	...	...	0.00	6

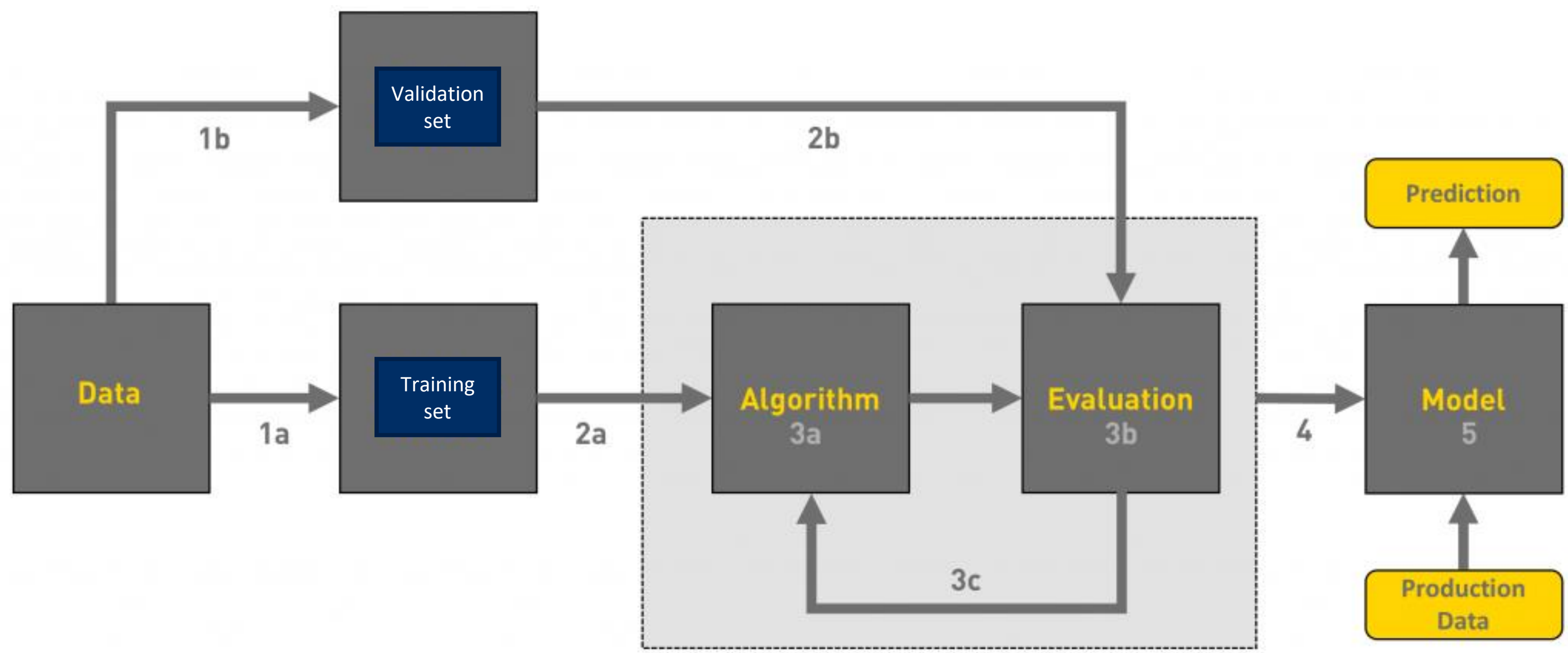


Machine Learning Approach

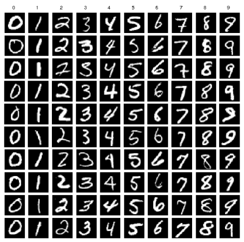


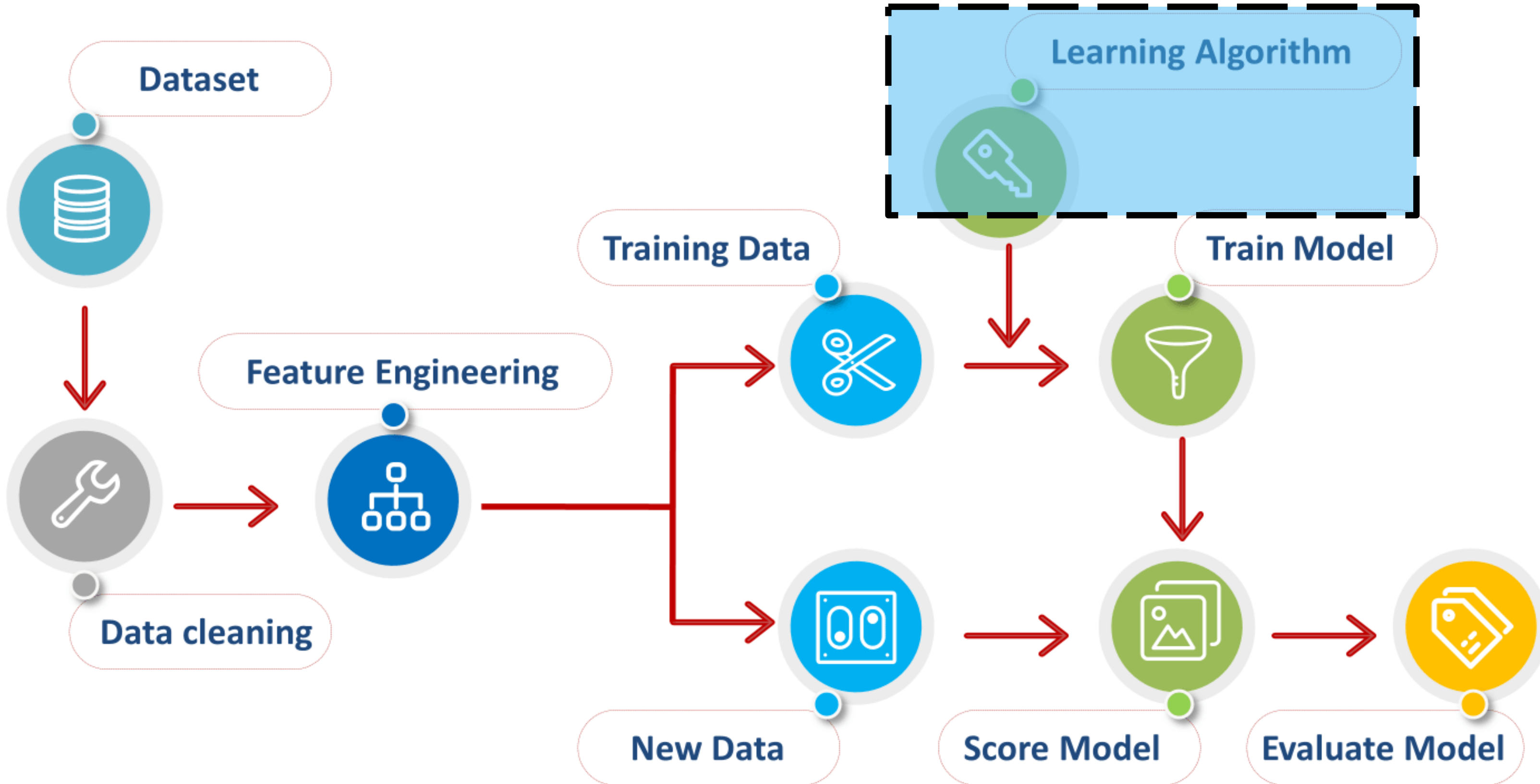
Data Splitting and Learning Loop

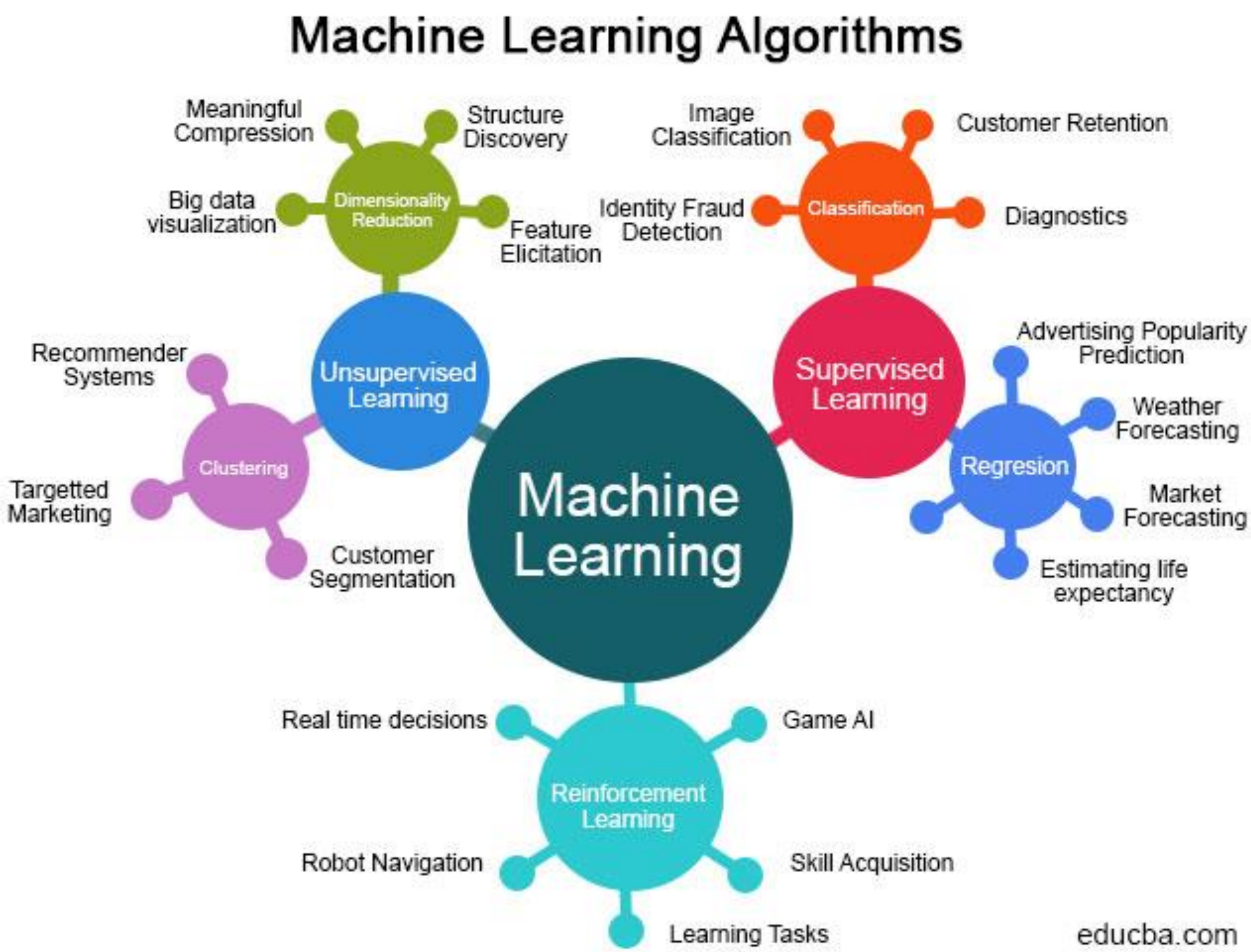
- Train/Validation/Test procedure
  - Split the training data into three sets:
    - Training set (60%)
    - Cross validation set (20%)
    - Test set (20%)



Machine Learning Approach







University of Antwerp

From an Annotation Point of View

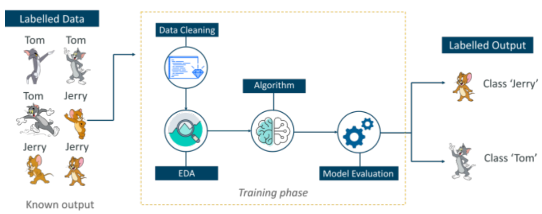
- Supervised: The true labels are available in supervised learning.
- Unsupervised: The true labels are not available
- Semi-supervised: Some samples have true label and some samples are unlabeled.

- Common Algorithms:

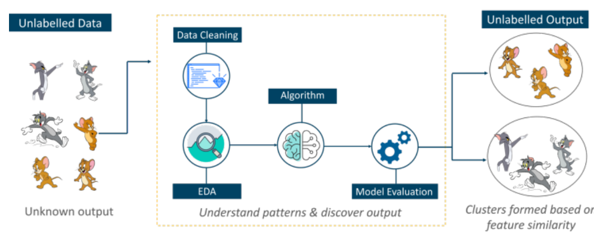
Linear Regression
Logistic Regression
Decision Tree
SVM
Naive Bayes
kNN
K-Means
Random Forest
Dimensionality Reduction Algorithms
Gradient Boosting algorithms
    GBM
    XGBoost
    LightGBM
    CatBoost

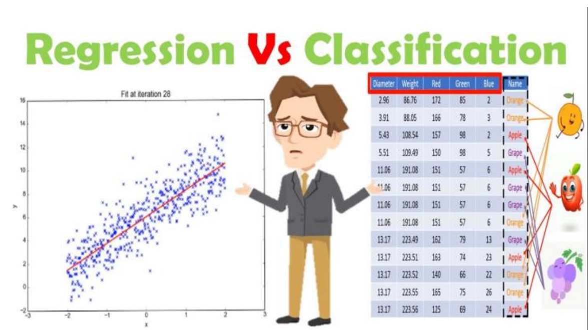
Machine Learning problems

- Binary Classification

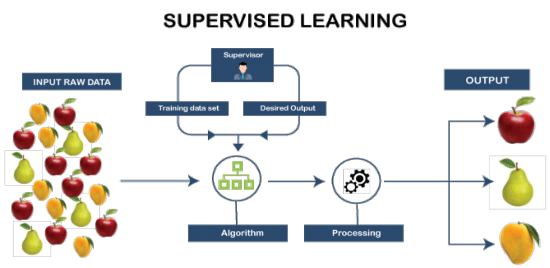


- Clustering

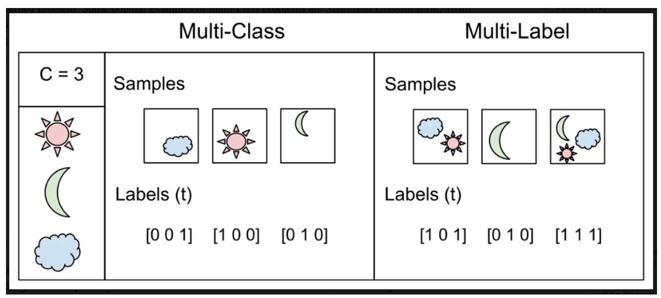


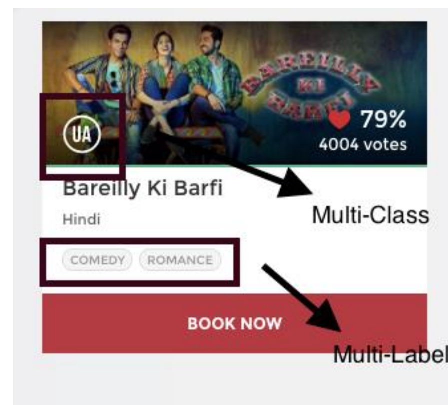


- Multi-classes classification



- Multi-label classification(Two Ex.)

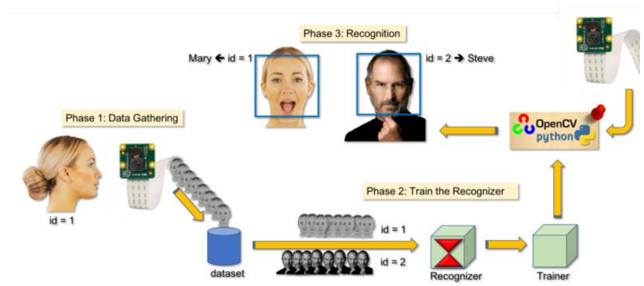




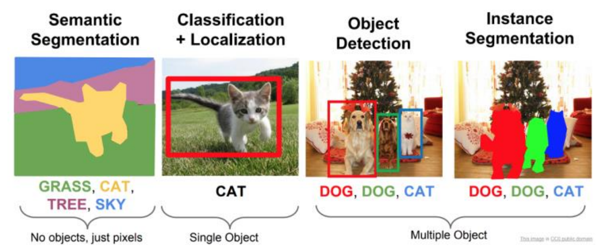
University of Antwerp

Machine Learning Applications

- Face Recognition

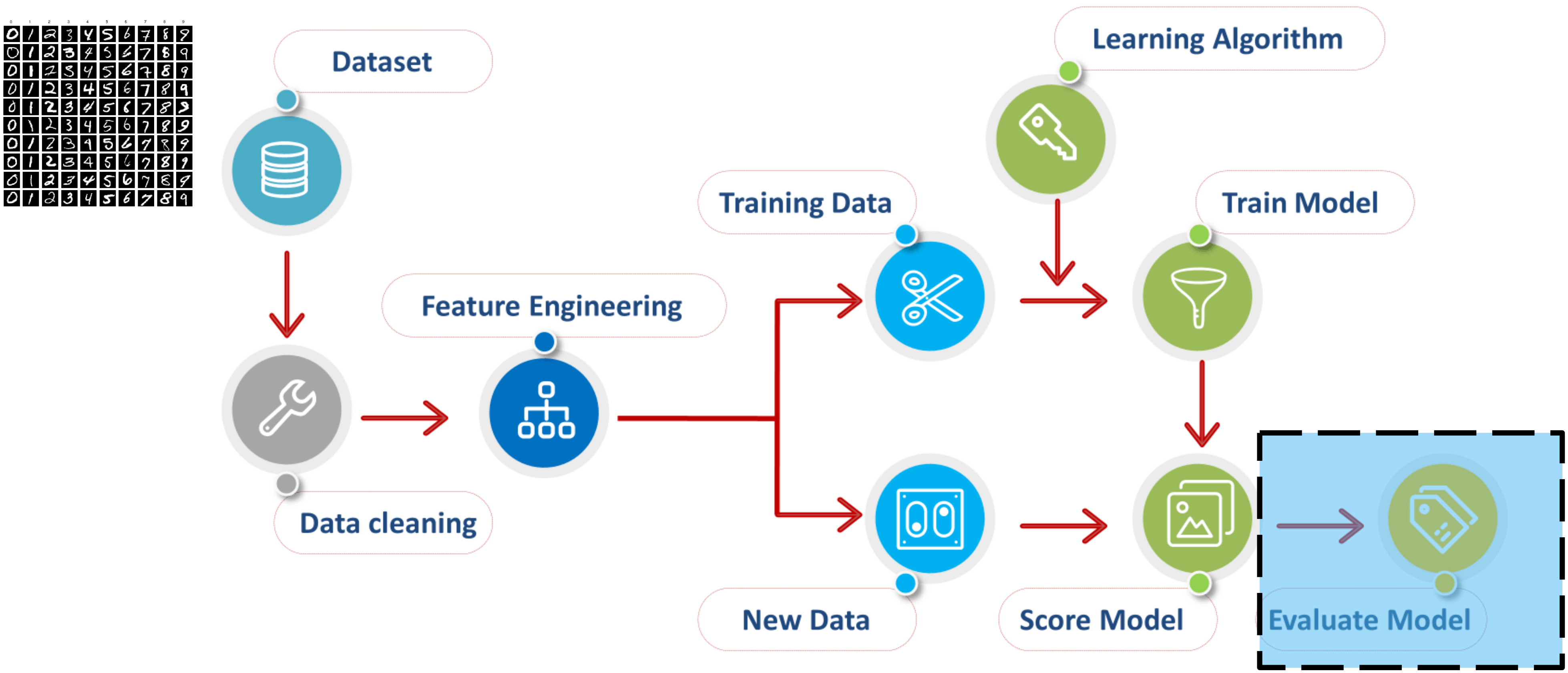


- Segmentation



- Sequence Prediction
- Action Recognition
- Speech Recognition
- Medical Applications
- Time Series Prediction

Machine Learning Approach



Performance Measures

Terminology and derivations from a confusion matrix

condition positive (P)
the number of real positive cases in the data

condition negative (N)
the number of real negative cases in the data

true positive (TP)
A test result that correctly indicates the presence of a condition or characteristic

true negative (TN)
A test result that correctly indicates the absence of a condition or characteristic

false positive (FP)
A test result which wrongly indicates that a particular condition or attribute is present

false negative (FN)
A test result which wrongly indicates that a particular condition or attribute is absent

- Confusion Matrix.

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

sensitivity, recall, hit rate, or true positive rate (TPR)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} = 1 - \mathrm{FNR}$$

specificity, selectivity or true negative rate (TNR)

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} = 1 - \mathrm{FPR}$$

precision or positive predictive value (PPV)

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} = 1 - \mathrm{FDR}$$

negative predictive value (NPV)

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}} = 1 - \mathrm{FOR}$$

accuracy (ACC)

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{P} + \mathrm{N}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

F1 score

is the harmonic mean of precision and sensitivity:

$$\mathrm{F}_1 = 2 \times \frac{\mathrm{PPV} \times \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

Performance Measures

sensitivity, recall, hit rate, or true positive rate (TPR)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} = 1 - \mathrm{FNR}$$

specificity, selectivity or true negative rate (TNR)

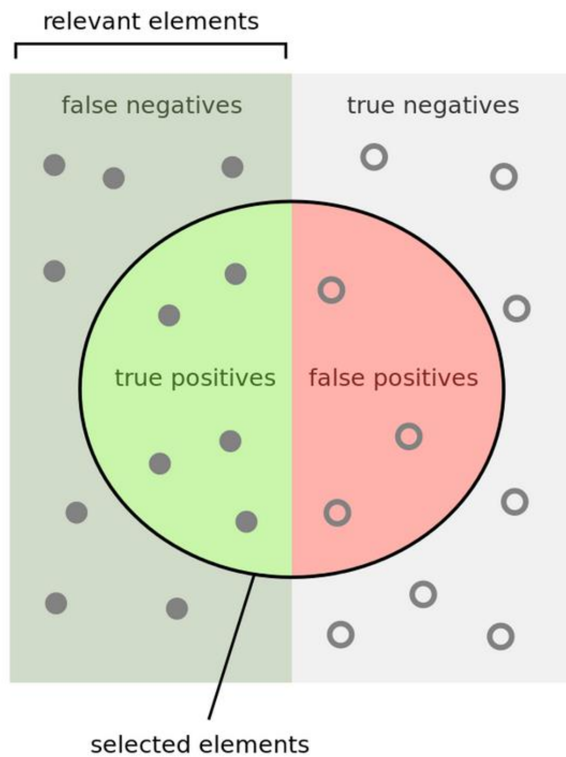
$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} = 1 - \mathrm{FPR}$$

precision or positive predictive value (PPV)

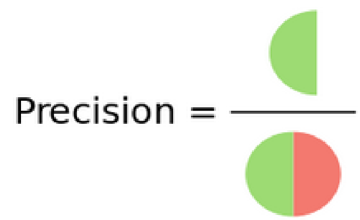
$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} = 1 - \mathrm{FDR}$$

negative predictive value (NPV)

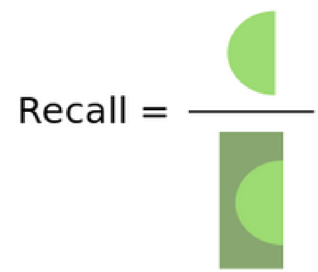
$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}} = 1 - \mathrm{FOR}$$



How many selected items are relevant



How many relevant items are selected



Confusion Matrix for more than 2 classes

Unlike binary classification, there are no positive or negative classes here. At first, it might be a little difficult to find TP, TN, FP and FN since there are no positive or negative classes, but it's actually pretty easy. What we have to do here is to find TP, TN, FP and FN for each individual class. For example, if we take class Apple, then let's see what are the values of the metrics from the confusion matrix.

		True Class		
		Apple	Orange	Mango
Predicted Class	Apple	7	8	9
	Orange	1	2	3
	Mango	3	2	1

- ***TP = 7***
- ***TN = (2+3+2+1) = 8***
- ***FP = (8+9) = 17***
- ***FN = (1+3) = 4***

- ***Precision = 7/(7+17) = 0.29***
- ***Recall = 7/(7+4) = 0.64***
- ***F1-score = 0.40***

Accuracy Paradox

	Actual - Cancer	Actual - NOT Cancer	Total
Predicted - Cancer	TP = 0	FP = 0	0
Predicted - NOT Cancer	FN = 30	TN = 270	300
Total	30	270	300

Acc.

=0.9

Poor Performance

Recall =

Precision =

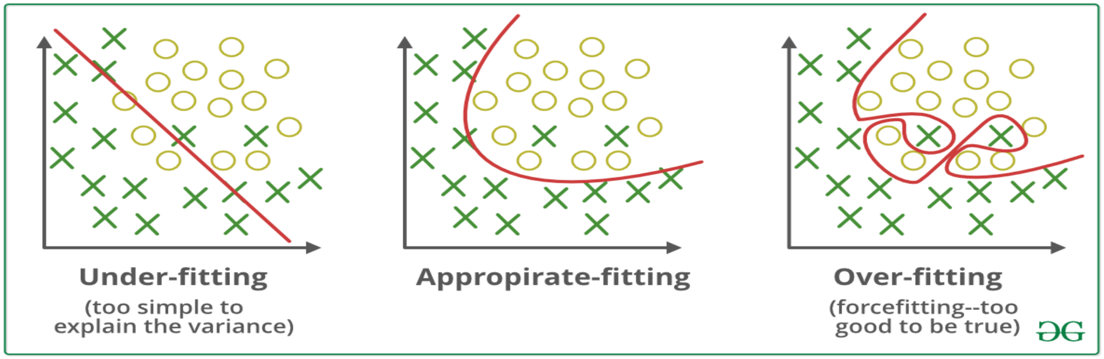
Exercise

	predicted 0	predicted 1	predicted 2	predicted 3	predicted 4	predicted 5	predicted 6	predicted 7	predicted 8	predicted 9
actual 0	954	0	0	7	1	10	6	3	7	3
actual 1	0	1031	4	3	1	4	1	2	16	2
actual 2	12	21	852	18	11	8	14	20	29	5
actual 3	2	5	9	899	1	71	0	12	23	7
actual 4	2	8	2	2	861	7	7	1	4	89
actual 5	7	5	9	24	3	833	12	8	12	2
actual 6	11	6	2	0	6	31	902	0	8	1
actual 7	3	10	5	3	7	7	1	1041	0	14
actual 8	2	28	4	29	2	31	1	9	882	21
actual 9	7	3	1	7	10	11	1	44	4	873

Problem Diagnosis

Underfitting & Overfitting

- Underfitting:
  - The model performs poorly on the training data.
- Overfitting
  - The model performs well on the training data, while it can not classify the test data with good accuracy.
  - It fails to generalize to the new examples
- Good Fitting:
  - The model that suitably learns the training dataset and generalizes well to the test data.



Data and Training Issues

- Train/Validation Test procedure might have some issues
  - If we have a sparse training dataset, we may not be able to select a portion of data
  - The test error may be an optimistic estimate of the true error due to an fortunate/unfortunate split
  - Imbalanced data: Imbalanced data typically refers to a problem with classification problems where the classes are not represented equally.
- To address first two limitations, use:
  - Random Subsampling Cross-Validation
  - K-Fold Cross-Validation
  - Leave-One-Out Cross-Validation

Random Subsampling Cross Validation

- Random Subsampling performs K data splits of entire dataset
  - Each data split randomly selects a fixed number of examples without replacement.
  - For each data split we retrain the classifier from scratch with the training examples and then estimate with the test examples

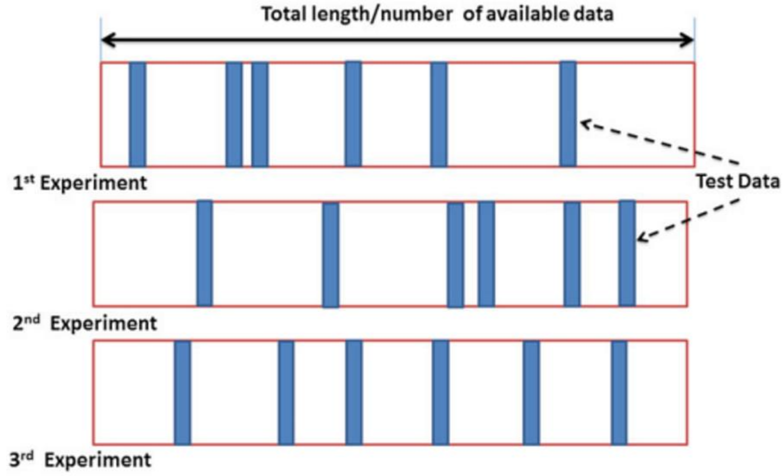
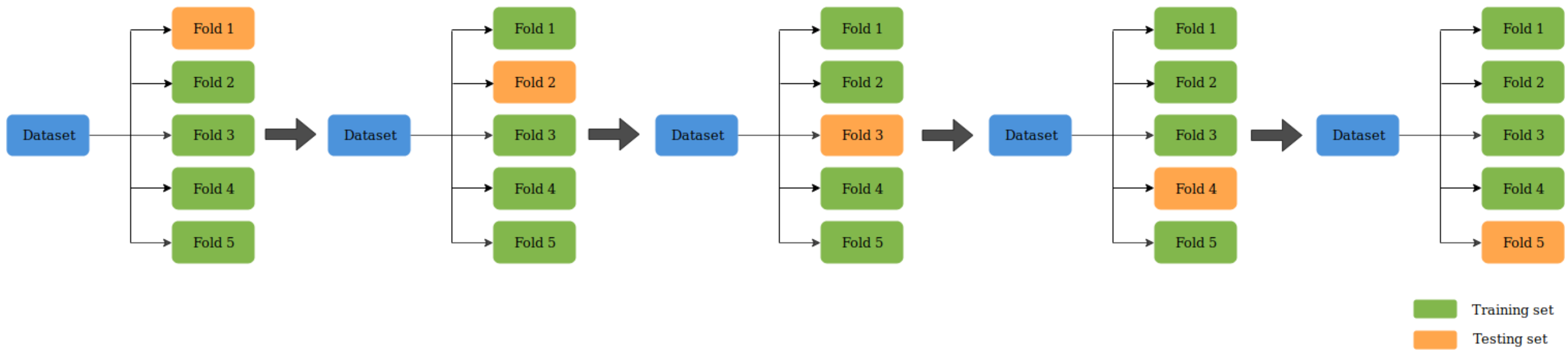


Fig. 3.7 Data splitting in the random sub-sampling approach

  - The true error estimate is obtained as the average of the separate estimations.

K-Fold Cross-Validation

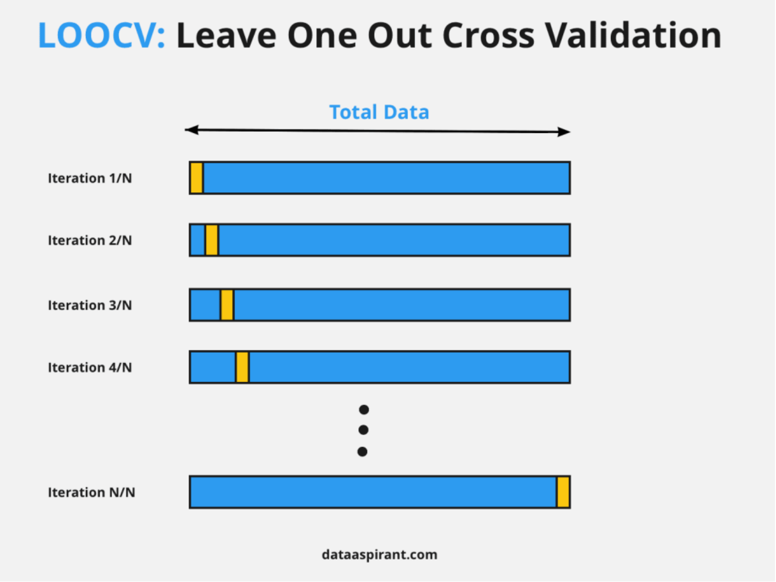
- Create a K-fold partition of the dataset
  - For each of the K experiments, use K-1 folds for the training procedure and a different fold for testing procedure



  - The true error estimate is obtained as the average of the separate estimations.

Leave-OneOut Cross-Validation

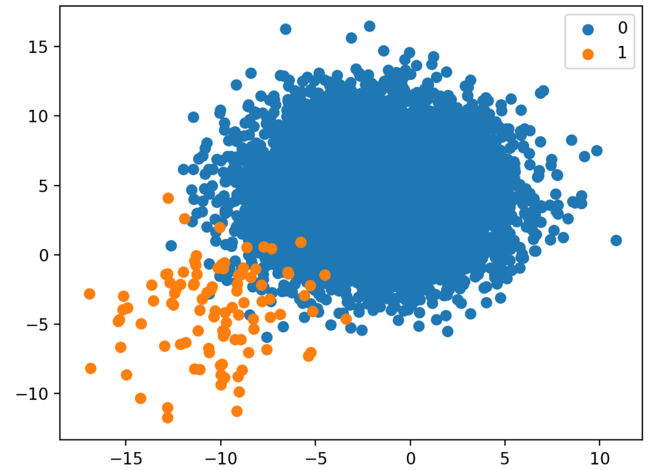
- Leave-One-Out is a K-fold Cross Validation, where K is chosen as the total number of examples
  - For each of the K experiments, use K-1 folds for the training procedure and a different fold for testing procedure



  - The true error estimate is obtained as the average of the separate estimations.

Imbalanced Data

- Imbalanced data typically refers to a problem with classification problems where the classes are not represented equally.
- For example: the ratio of Class-1 to Class-2 instances is 80:20 or more concisely 4:1
- What Happen? Accuracy Paradox



	Actual - Cancer	Actual - NOT Cancer	Total
Predicted - Cancer	TP = 10	FP = 0	10
Predicted - NOT Cancer	FN = 990	TN = 999000	999990
Total	1000	999000	1000000

Acc.

$$\frac{10 + 999000}{1000000} \approx 99.9\%$$

$$\frac{10}{10 + 0} * 100 = 100$$

Poor Performance

$$\frac{10}{10 + 990} = 1\%$$

One better metric in this situation

$$\text{F1 score} = \frac{2 \times 0.01 \times 1}{0.01 + 1} \approx 2\%$$

Some Tactics To Combat Imbalanced Training Data

- **Can You Collect More Data?**
- **Try Changing Your Performance Metric**
  - **We talked in the previous slide**

- **Try Resampling Your Dataset**
  - Down Sampling
  - Up Sampling

- **Try Generate Synthetic Samples**

- **Try Different Algorithms**
  - Decision Trees

- **Try Penalized Models**

Addressing Underfitting

- Underfitting may cause because:
  - The volume of data is not adequate for training the model.
  - The number of epochs or iterations to observe the examples is not adequate for the model.
    - What is the difference between terms epochs and iterations in machine learning?
  - More often, <u>the model is too simples</u>. What does it mean?

- Addressing underfitting might achieved by:
  - Using more training data/increasing the number of epochs.
  - Using more complex model. What does it mean?

Addressing Overfitting

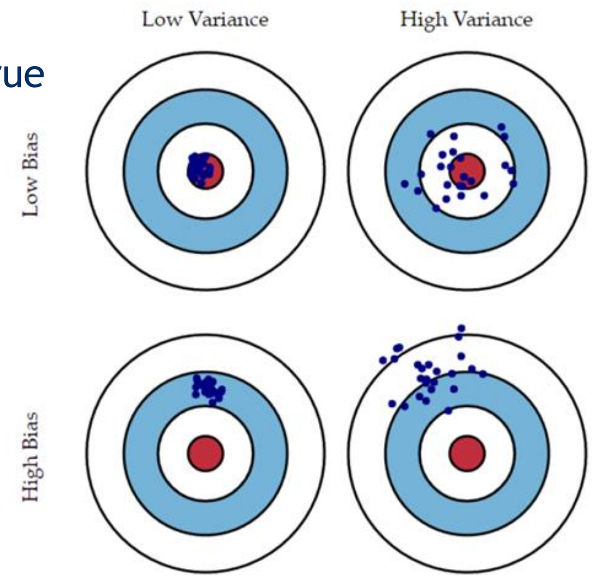
- Overfitting may cause because the model is too complex.

- Addressing overfitting might achieved by:
  - Training the model on more example.
    - How? Add more data using data augmentation
  - Reducing the complexity of the model. How?
    - Changing the model structure

Bias/Variance

- How good is an estimation?

- Two measures of “goodness” are used for statistical estimates.
  - BIAS: How close is the estimate to the true value?
    - Difference between the estimator’s expected value and the true value of the parameter being estimated.



  - VARIANCE: How much does the estimate change if you train on a different training set?
    - Being sensitive to small fluctuations in the training set.
    - This is due to the classifier being overfitted to a particular training set

Questions?