MACHINE LEARNING

1211635

MAHASARAKHAM
UNIVERSITY

MACHINE LEARNING

Introduction to machine learning

Olarik Surinta, PhD.
Lecturer

Data Science Involves

- Data science is about using **data** to **make decision** that **drive actions**.
- Data science process involves:
  - Data selection
  - Preprocessing
  - Transformation
  - Data Mining
  - Delivering value from data: Interpretation and evaluation

Data Science

- Data Science is fat too complex
  - Cost of accessing/using efficient ML algorithms is high
  - Comprehensive knowledge required on different tools/platforms to develop a complete ML project
  - Difficult to put the developed solution into a scalable production stage

Machine Learning

- Machine learning is a subset of Artificial Intelligence (AI) where the machine is trained to learn from it’s past experience.
- The past experience is developed through the data collected. Then it combines with algorithms such as Naive Bayes, Support Vector Machine, Neural Network to deliver the final results.

Machine Learning

- Using **Known data**, develop a **model** to *predict* **unknown data**.

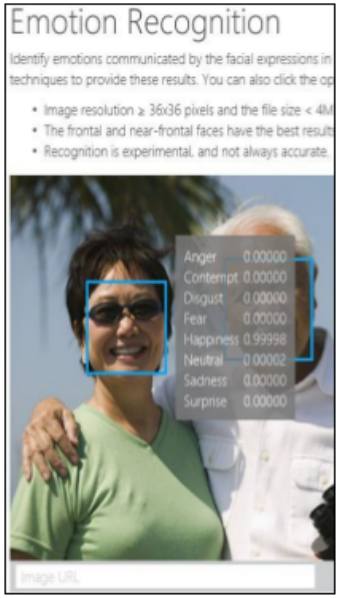
Known data: Big enough archive, previous observations, past data

Model: Known data + Algorithms (ML algorithms)

Unknown Data: Missing, Unseen, not existing, future data

Cr. https://www.slideshare.net/marafacs/azure-machine-learning-71568316

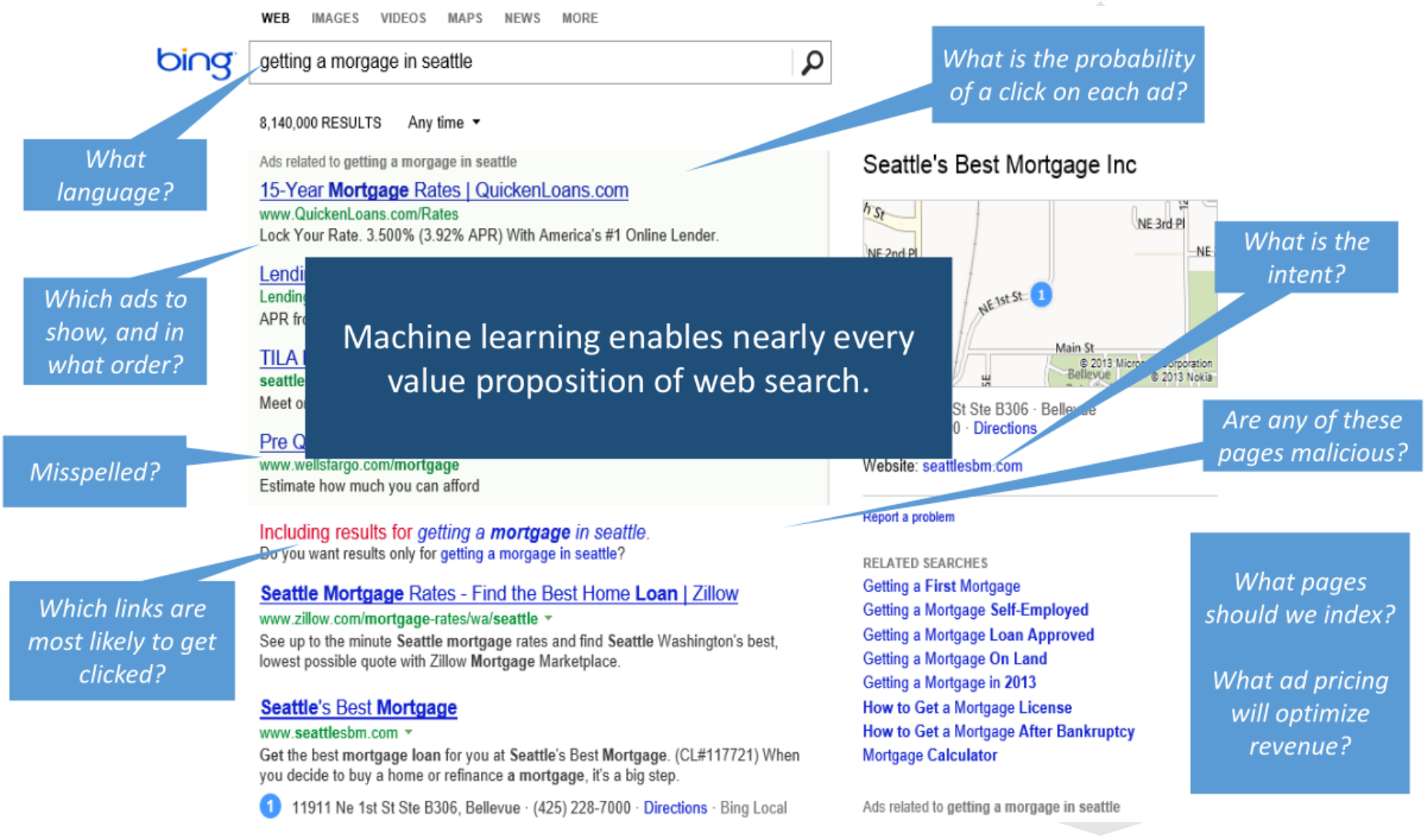
Why Machine Learning?



Applications of machine learning

- **Web search**: ranking page based on what you are most likely to click on.
- **Finance**: decide who to send what credit card offers to. Evaluation of risk on credit offers. How to decide where to invest money.
- **E-commerce**: Predicting customer. Whether or not a transaction is fraudulent.
- **Robotics:** how to handle uncertainty in new environments. Autonomous. Self-driving car.

Semantic web search



airplane - Google Search - Mozilla Firefox

airplane - Google Search

https://www.google.co.th/search?newwindow=1&d

Search

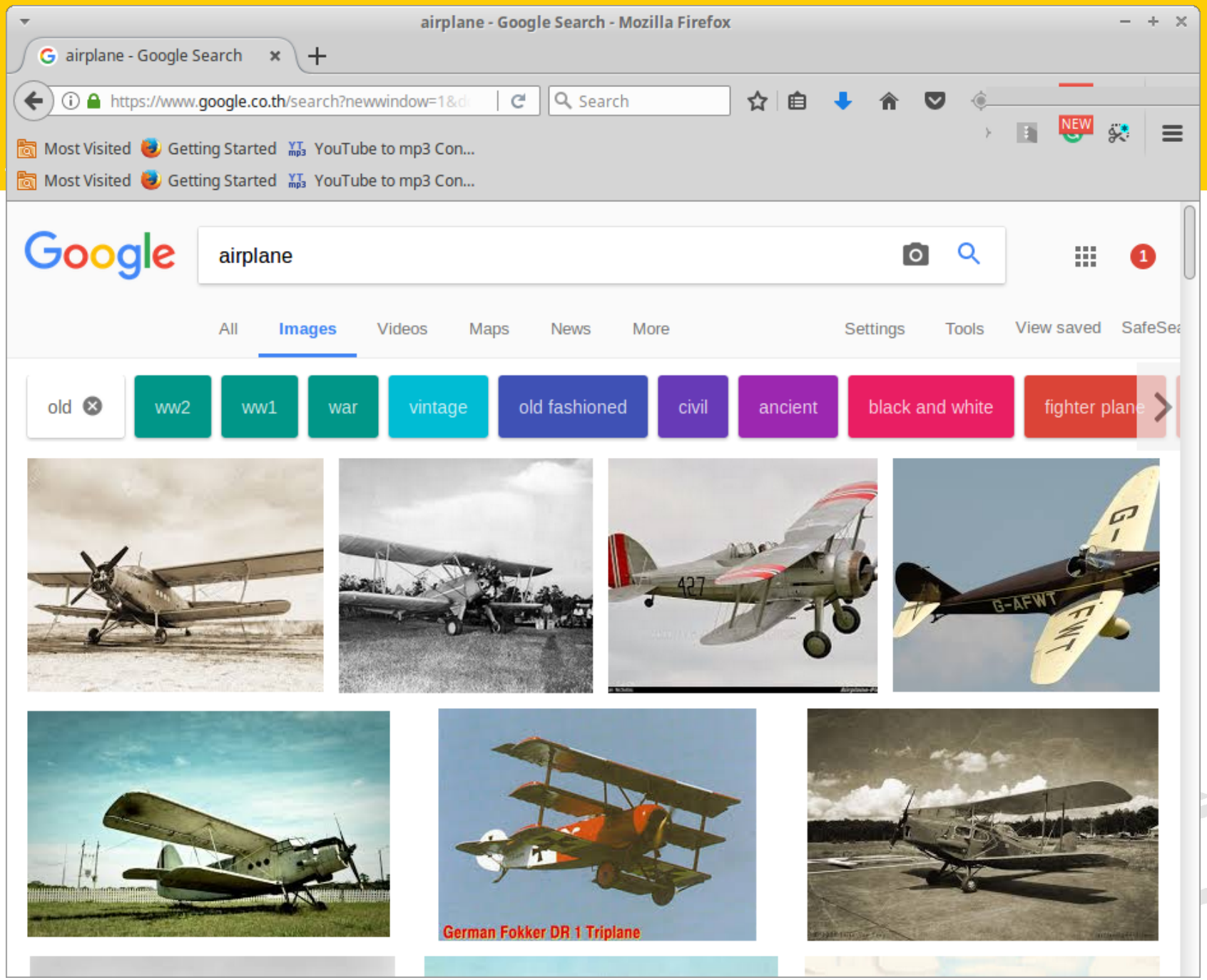
Most Visited | Getting Started | YouTube to mp3 Con...

Google

airplane

All | Images | Videos | News | Maps | More | Settings | Tools | View saved | SafeSea

1st class | the wright brothers | can | coloring | military | jet | helicopter | bomber | drone | fighter

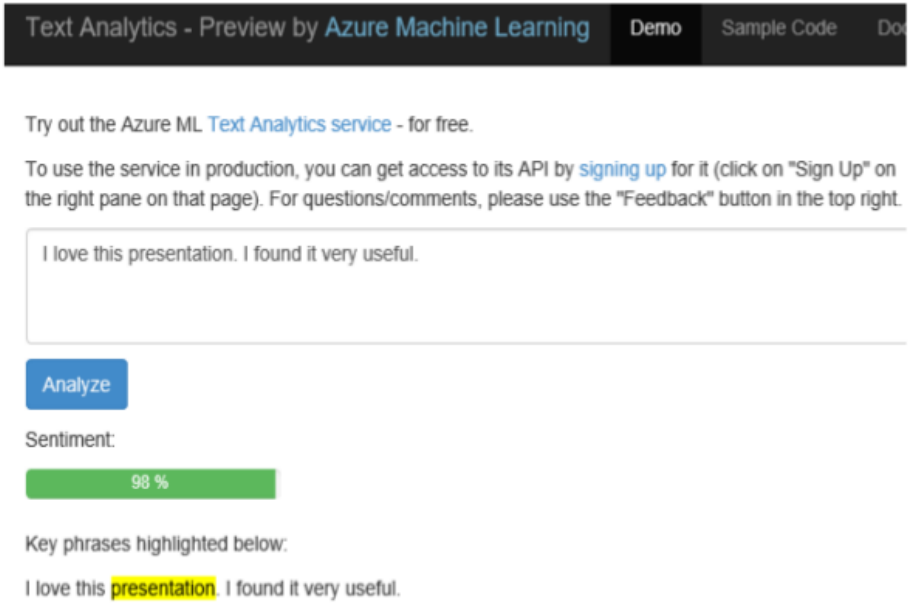

airplane - Google Search - Mozilla Firefox
airplane - Google Search
https://www.google.co.th/search?newwindow=1&
Search
Most Visited
Getting Started
YouTube to mp3 Con...
Google
airplane
All
Images
Videos
Maps
News
More
Settings
Tools
View saved
old
ww2
ww1
war
vintage
old fashioned
civil
ancient
black and white
fighter plane
G-AFWT
German Fokker DR 1 Triplane

Applications of machine learning

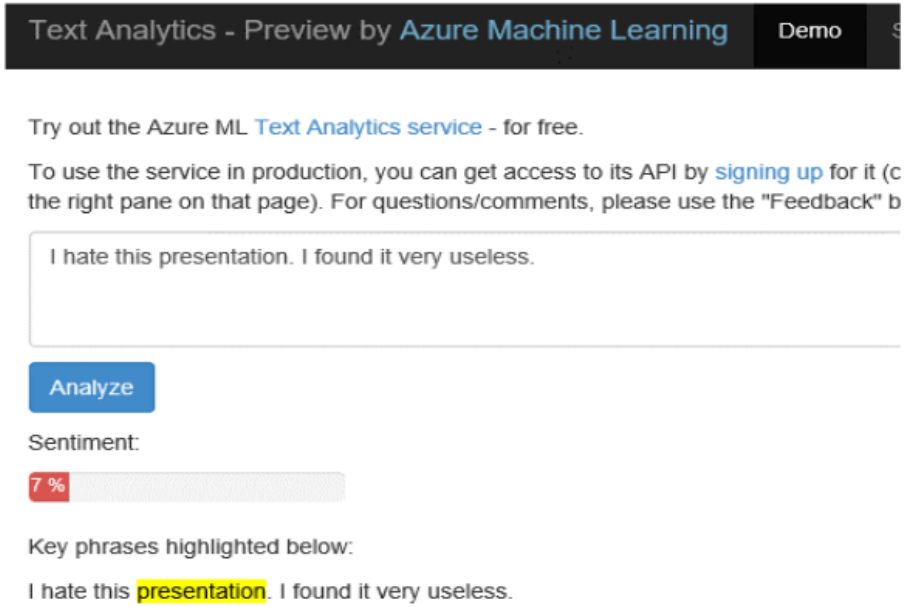
- **Web search**: ranking page based on what you are most likely to click on.
- **Finance**: decide who to send what credit card offers to. Evaluation of risk on credit offers. How to decide where to invest money.
- **E-commerce**: Predicting customer. Whether or not a transaction is fraudulent.
- **Robotics:** how to handle uncertainty in new environments. Autonomous. Self-driving car.

Text analytics: User reviews

Positive

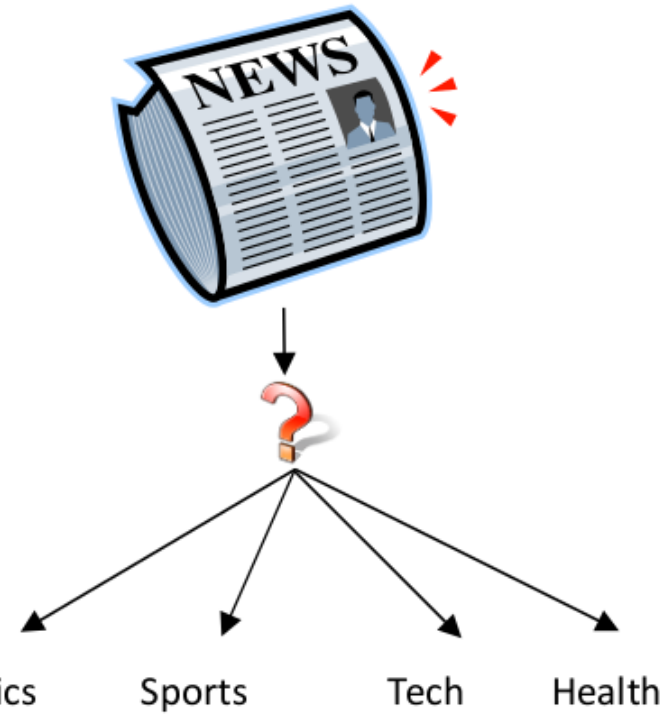


Negative



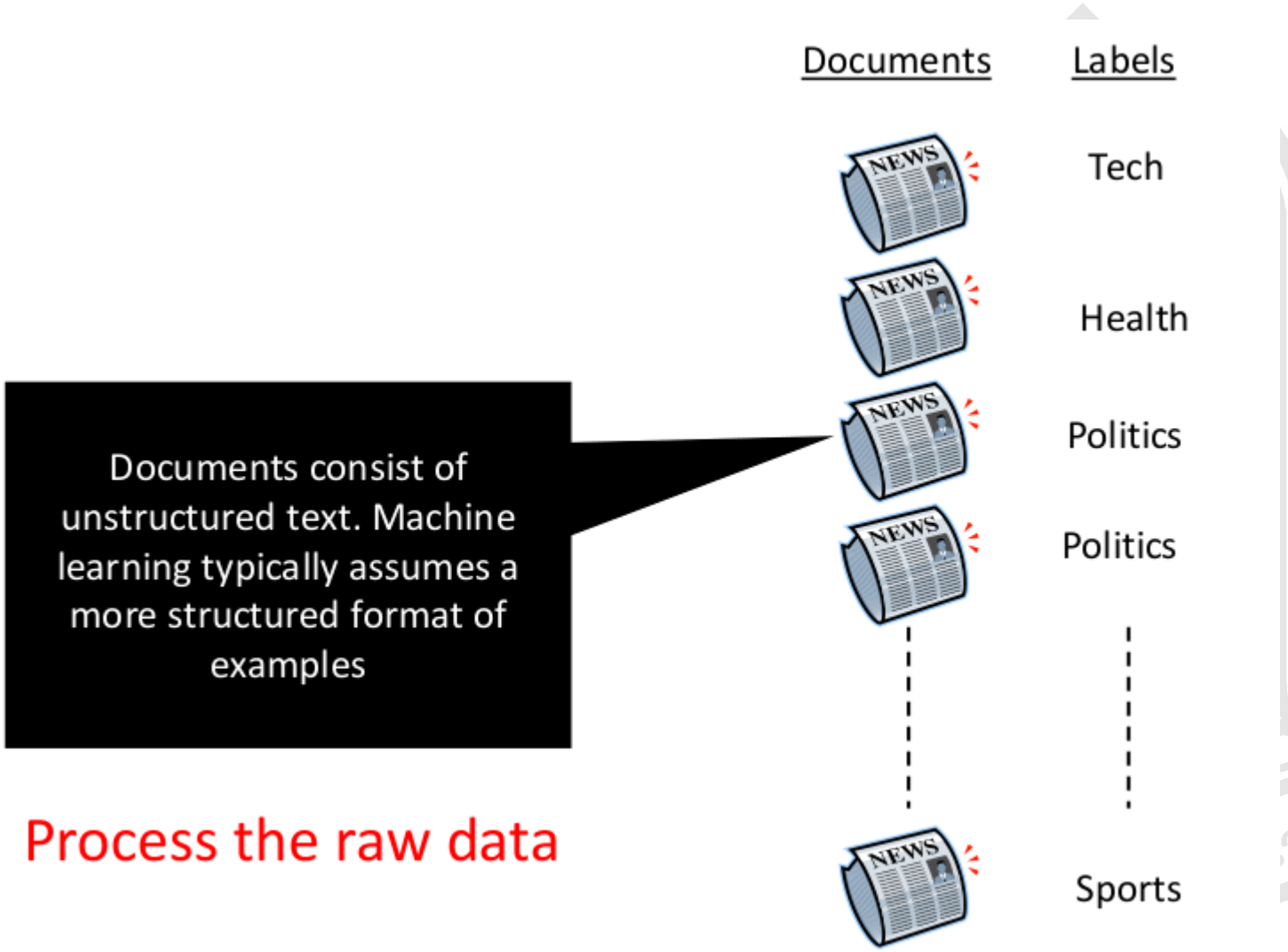
Text classification

- Classify a news article as (politics, sports, technology, health, etc.).

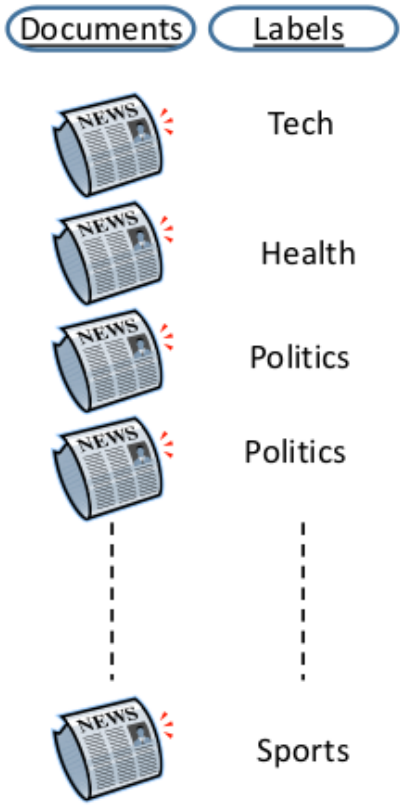


Using **known data**, develop a **model** to *predict* **unknown data**.

Known data (Training data)



Known data (Training data)

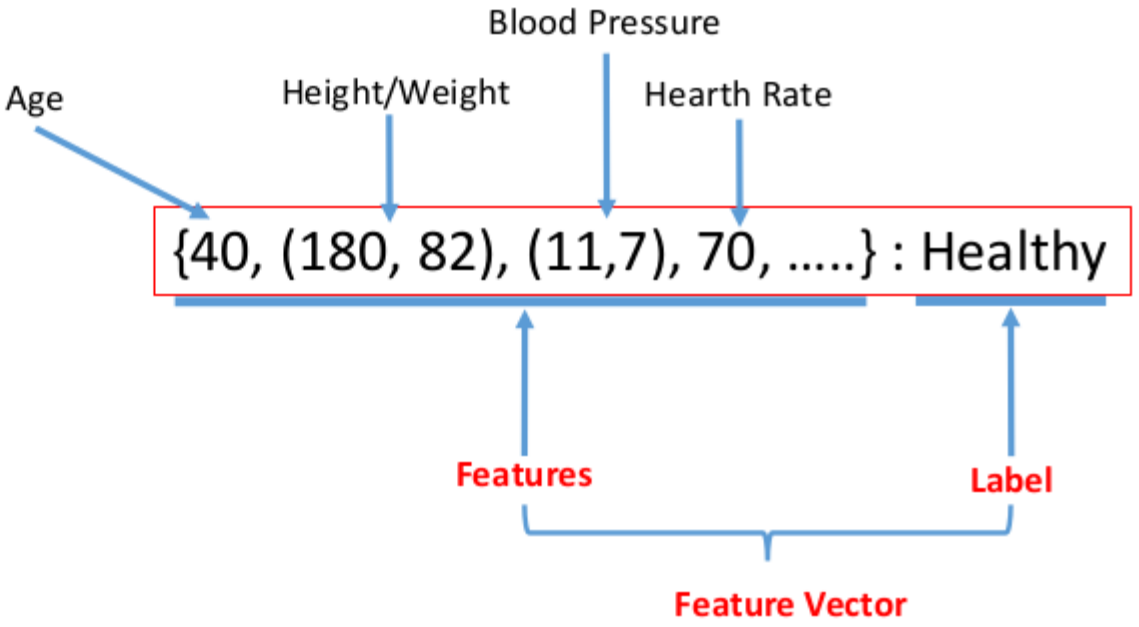


Process each data instance to represent it as a feature vector

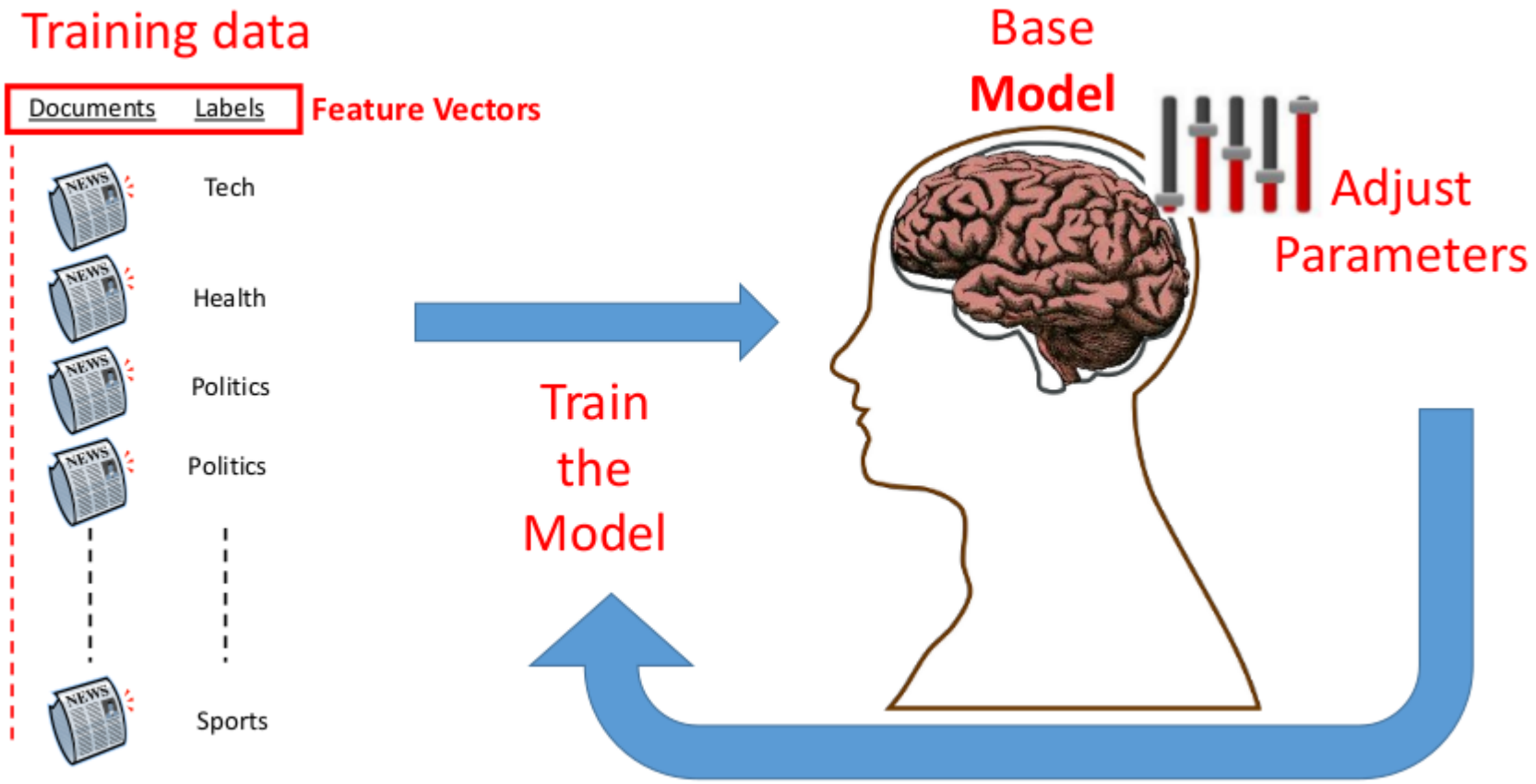
Using **known data**, develop a **model** to *predict* **unknown data**.

Feature vector

i.e.

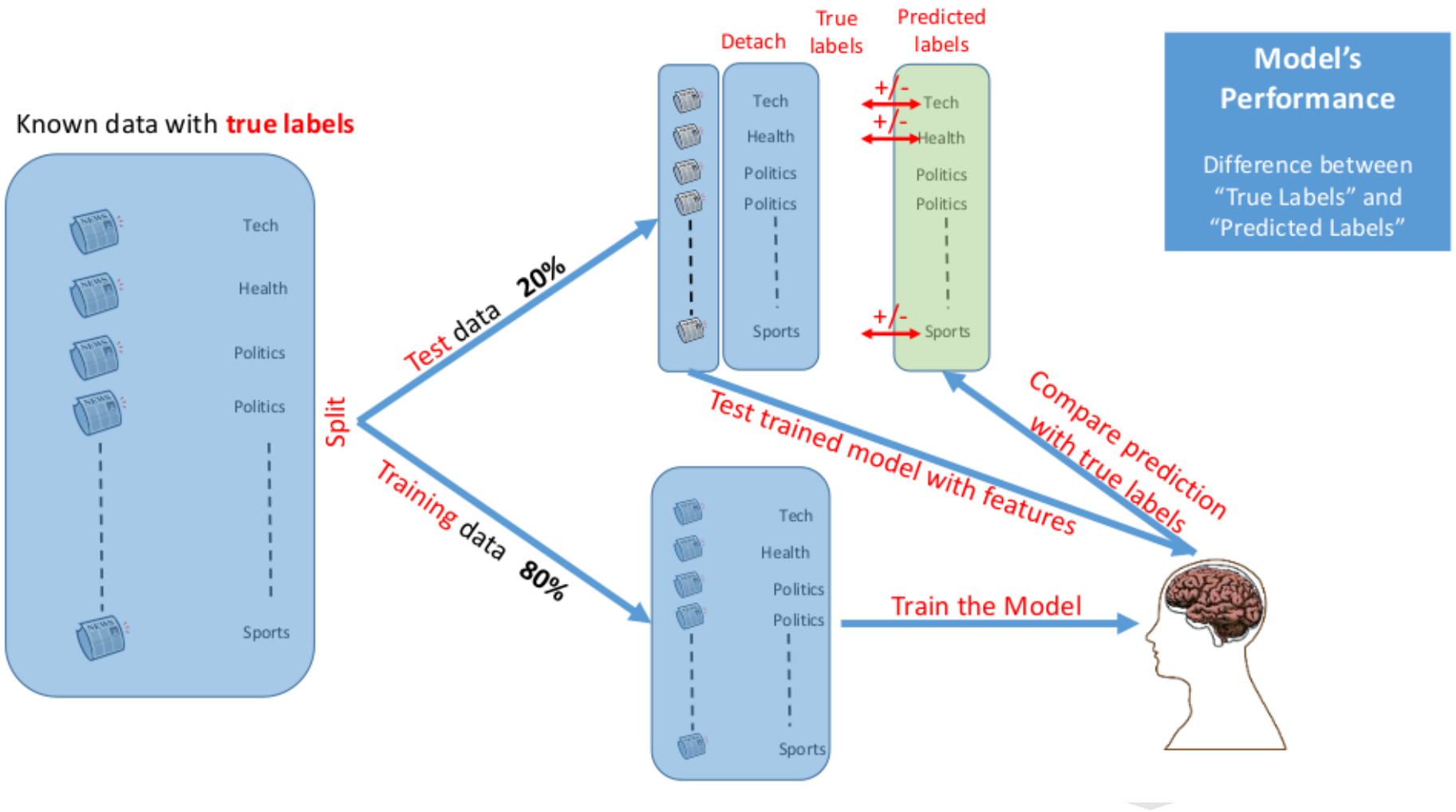


Developing a model

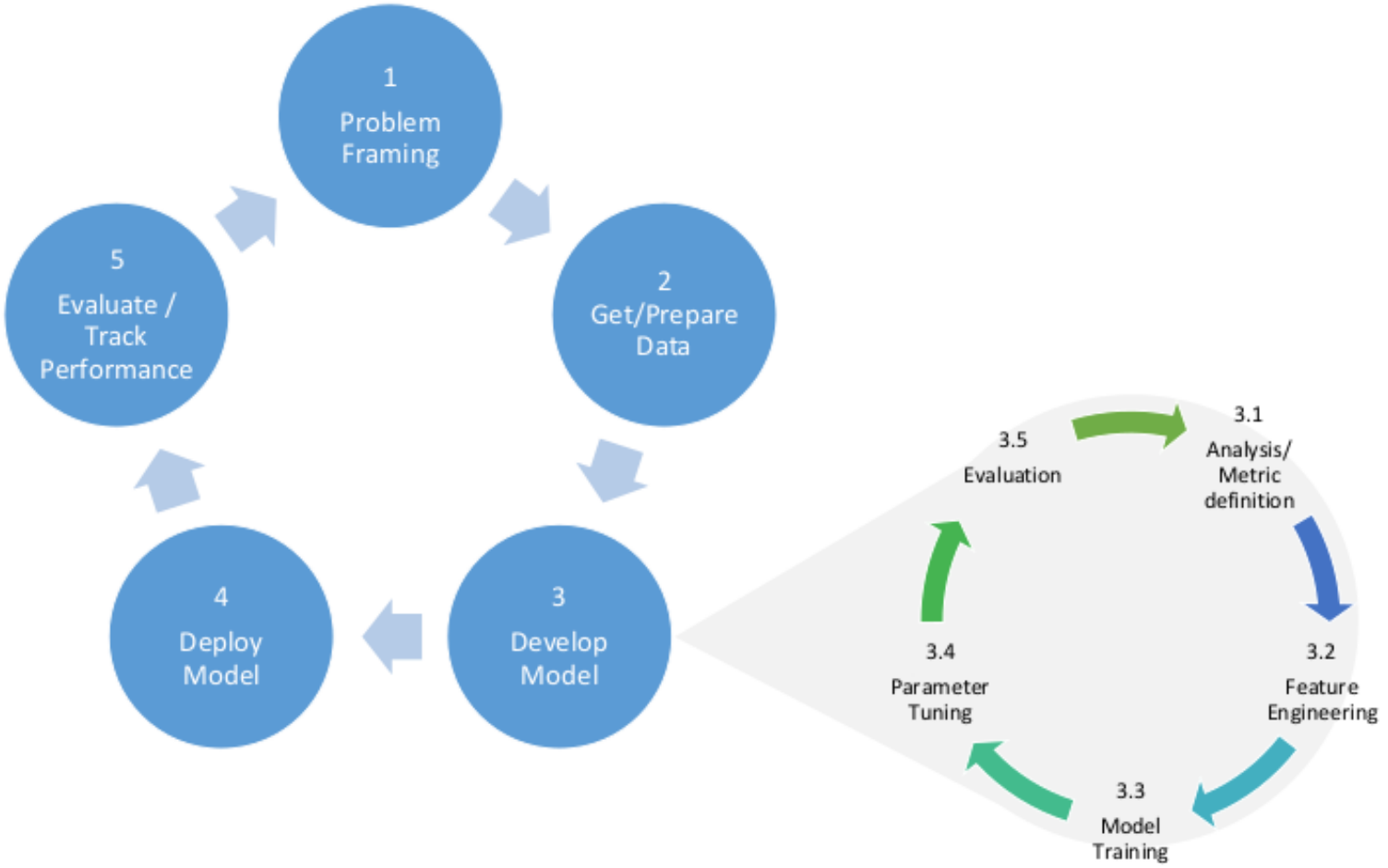


Using **known data**, develop a **model** to *predict* **unknown data**.

Model's Performance



Steps to build a machine learning solution



Applications of machine learning

- **Web search**: ranking page based on what you are most likely to click on.
- **Finance**: decide who to send what credit card offers to. Evaluation of risk on credit offers. How to decide where to invest money.
- **E-commerce**: Predicting customer. Whether or not a transaction is fraudulent.
- **Robotics:** how to handle uncertainty in new environments. Autonomous. Self-driving car.

What are the steps used in machine learning

- **Collecting data**
  - Be it the raw data from excel, access, text files etc.
- **Preparing the data**
  - Any analytical process thrives on the quality of the data used. One needs to spend time determining the quality of data and then taking steps for fixing issues such as missing data and treatment of outliers.

What are the steps used in machine learning

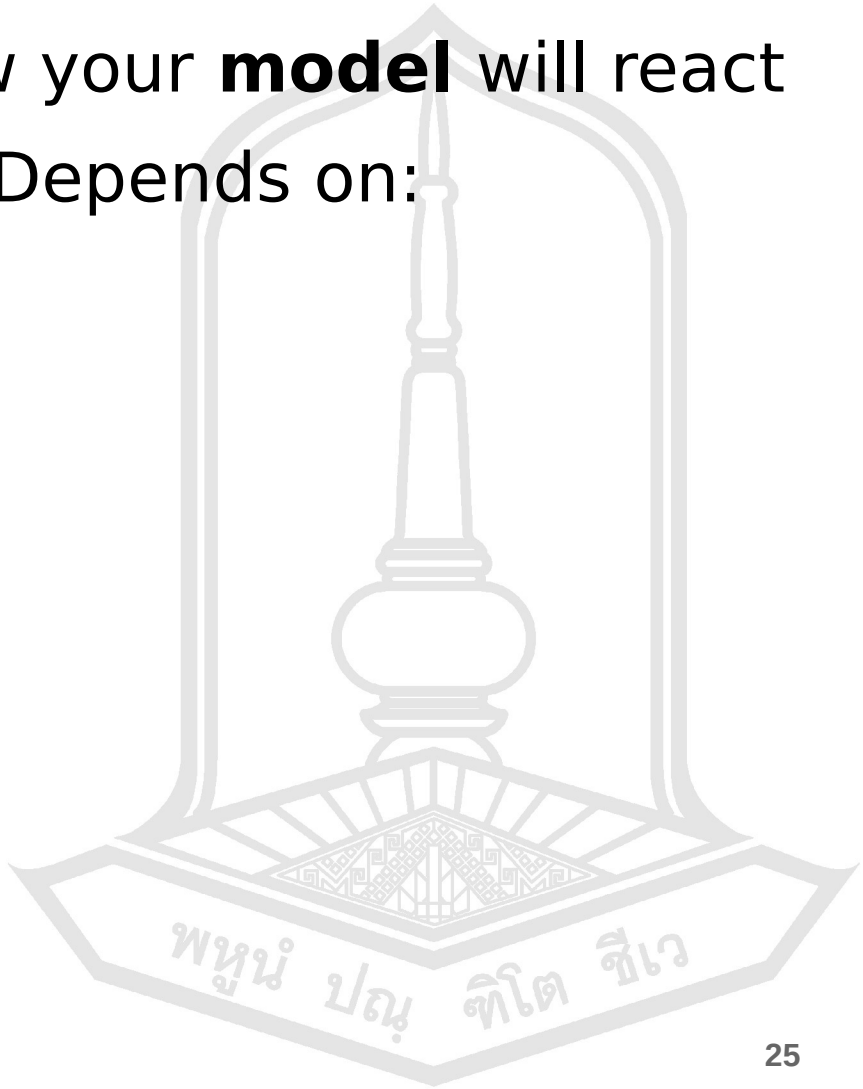
- **Training a model**
  - This step involves choosing the appropriate algorithm and representation of data in the form of the model. The cleaned data is split into two parts – train and test set. The first part (training data) is used for developing the model. The second part (test data) is used as a reference.
- **Evaluating the model**
  - To test the accuracy, the second part of the data is used. A better test to check accuracy of model is to see its performance on data which was not used at all during model build.

What are the steps used in machine learning

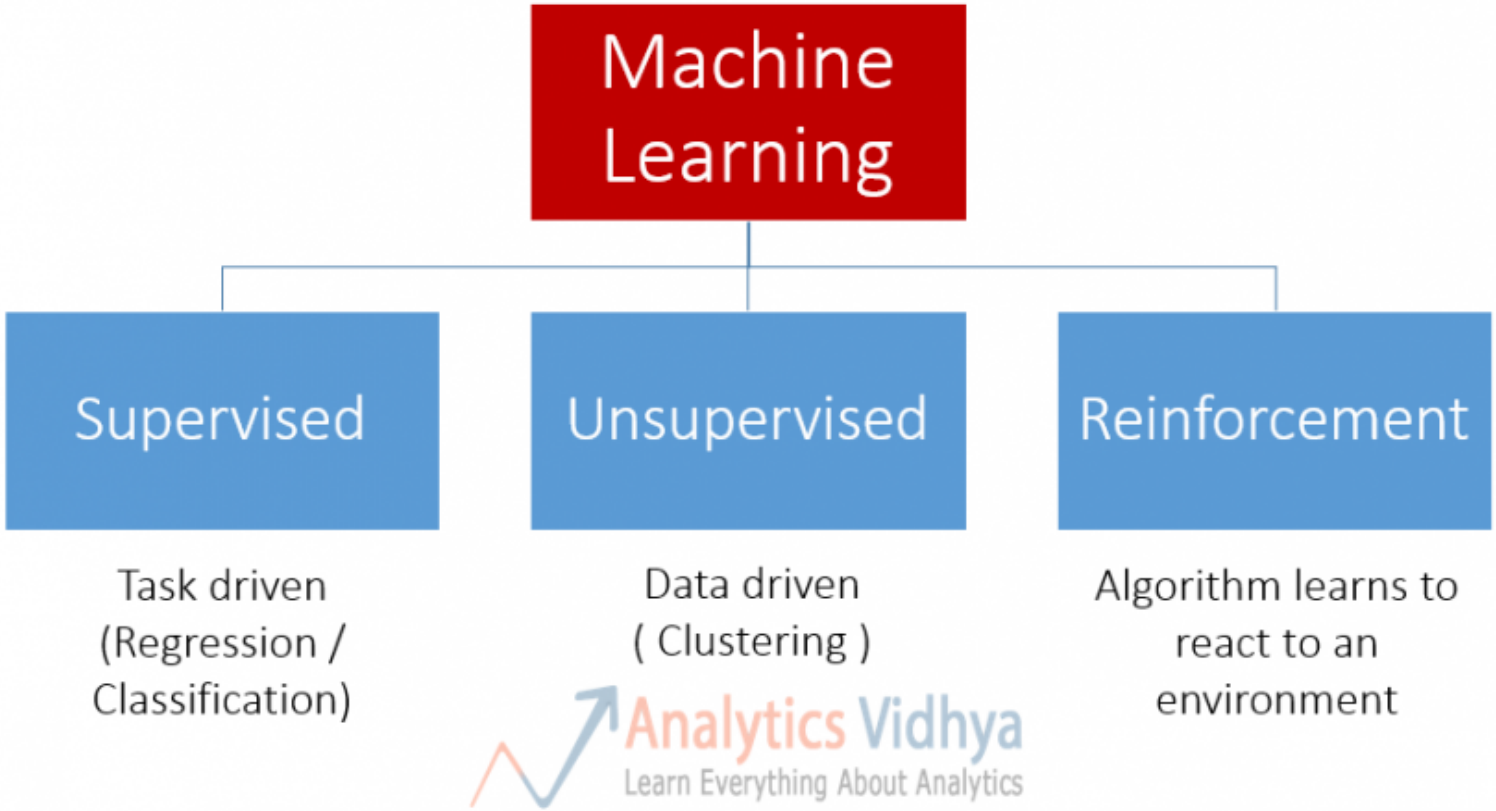
- **Improving the performance**
  - This step might involve choosing a different model altogether or introducing more variables to augment the efficiency.
  - That’s why significant amount of time needs to be spent in data collection and preparation

Machine learning algorithms

- ML algorithm defines how your **model** will react
- Which algorithm to use? Depends on:
  - Data quality
  - Data size
  - What you want to predict
  - Time constraint
  - Computation power
  - Memory limits



Types of machine learning



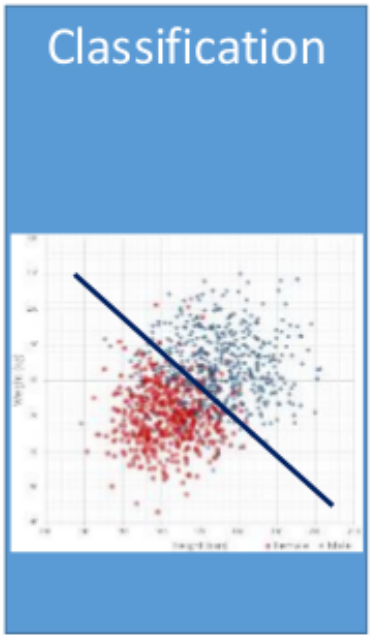
Supervised vs Unsupervised learning

- **Supervised Learning:** You give to the computer some pairs of inputs/outputs, so in the future when new inputs are presented you have an intelligent output.
- **Unsupervised Learning:** You let the computer learn from the data itself without showing what is the expected output.

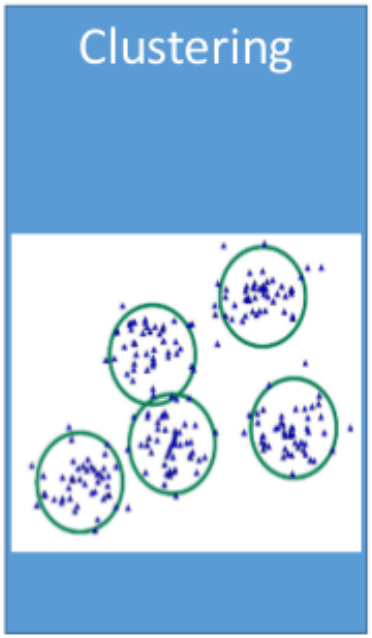
Supervised vs Unsupervised learning

- **Supervised learning** examples
  - This customer will like coffee
  - This network traffic indicates a denial of service attack
- **Unsupervised learning** examples
  - These customers are similar
  - This network traffic is unusual

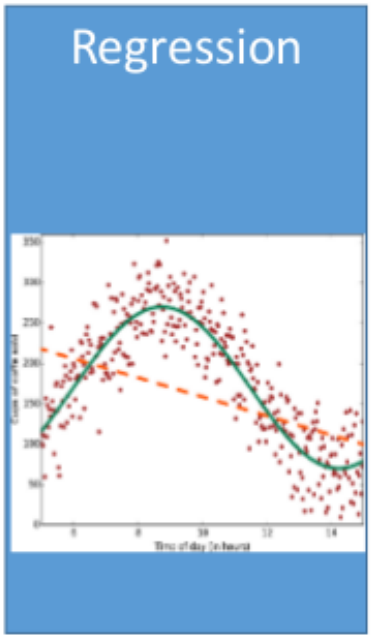
Common classes of algorithms



Supervised



UnSupervised

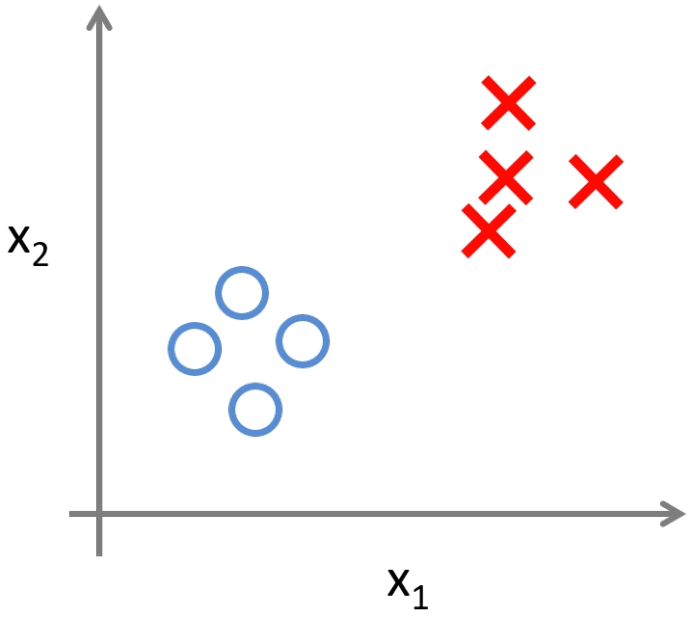


Supervised

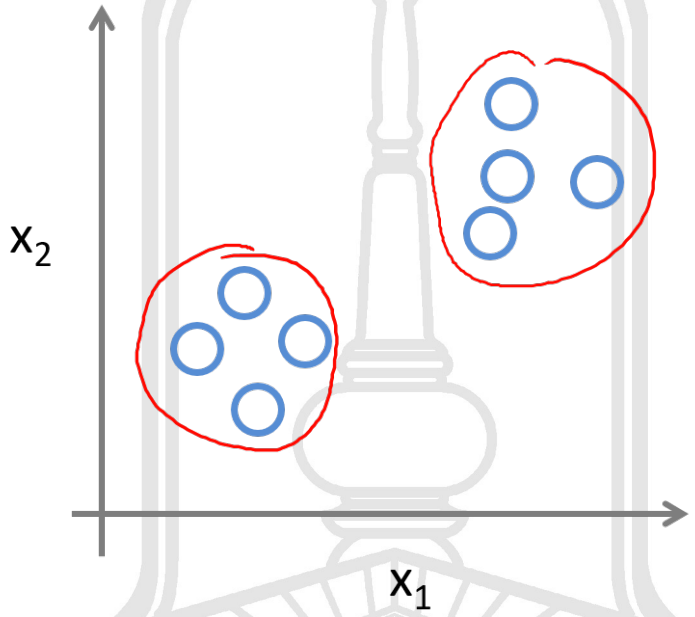
Supervised

Supervised vs Unsupervised learning

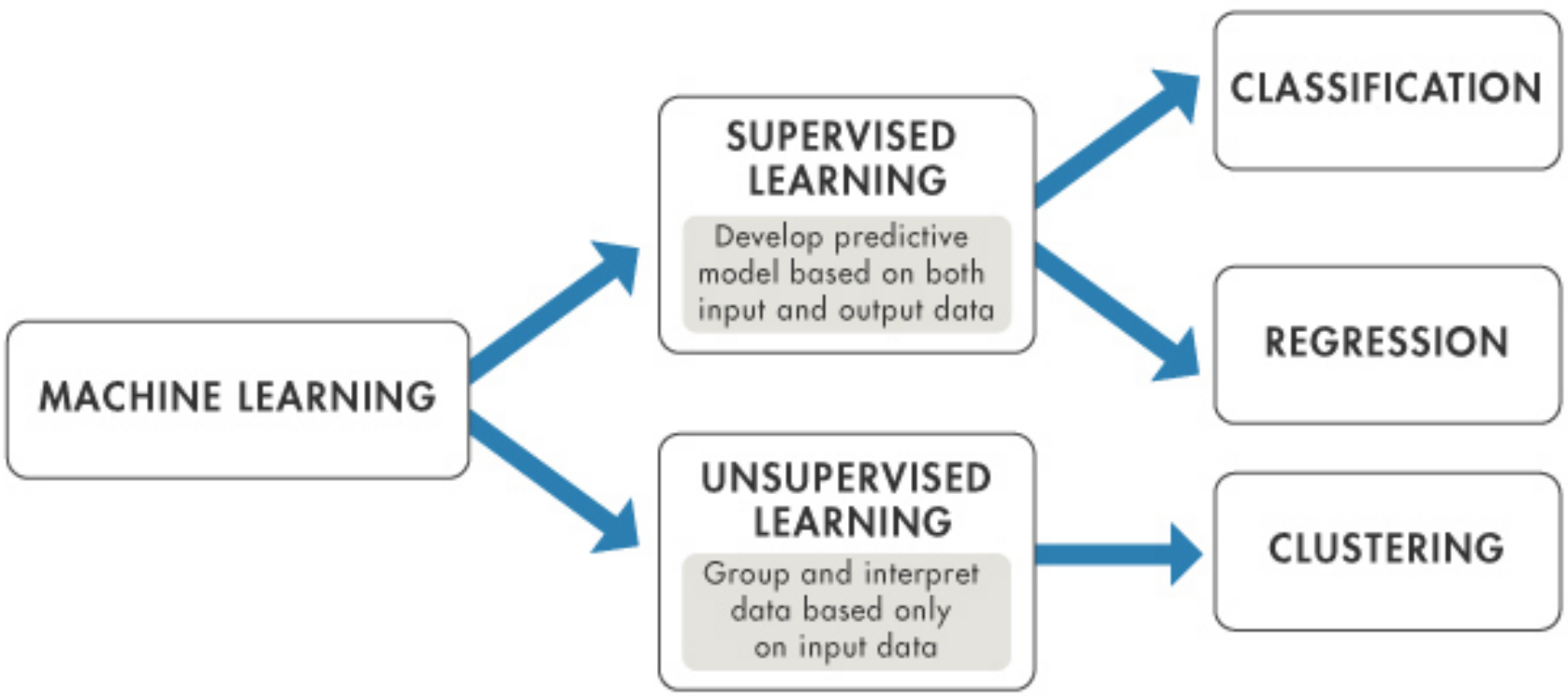
Supervised Learning

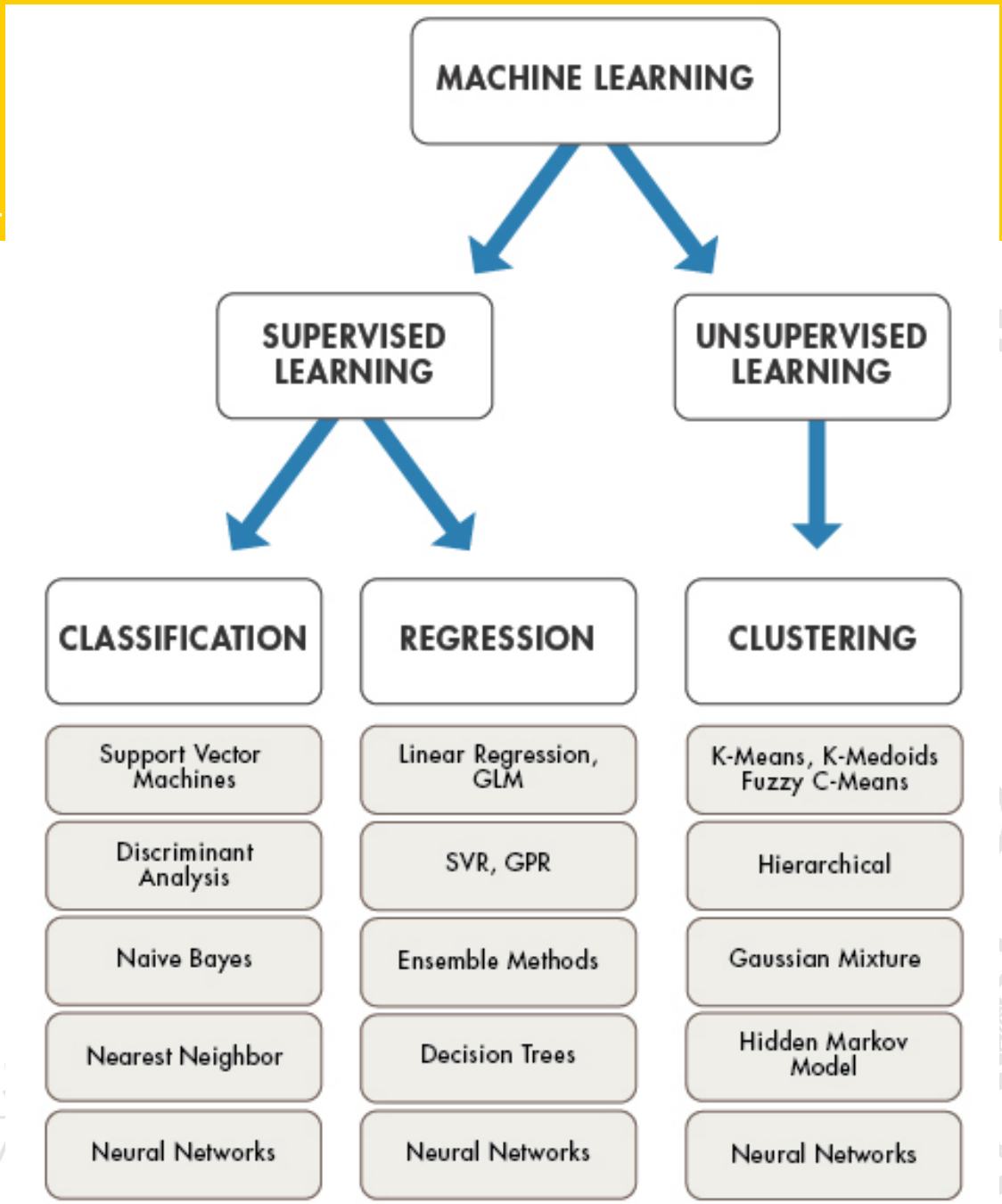


Unsupervised Learning

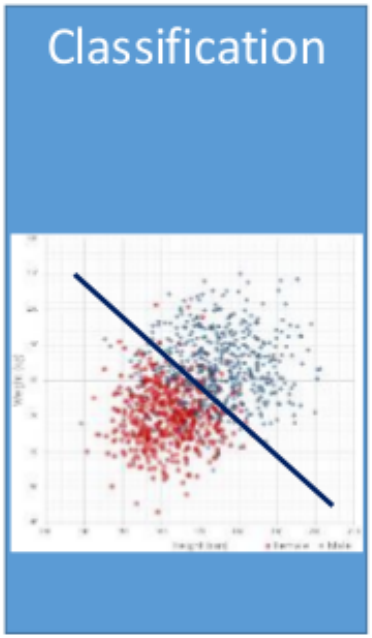


Supervised vs Unsupervised learning

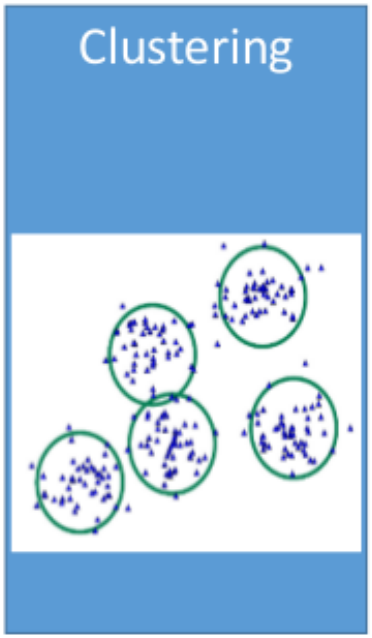



MACHINE LEARNING
SUPERVISED LEARNING
UNSUPERVISED LEARNING
CLASSIFICATION
REGRESSION
CLUSTERING
Support Vector Machines
Discriminant Analysis
Naive Bayes
Nearest Neighbor
Neural Networks
Linear Regression, GLM
SVR, GPR
Ensemble Methods
Decision Trees
Neural Networks
K-Means, K-Medoids Fuzzy C-Means
Hierarchical
Gaussian Mixture
Hidden Markov Model
Neural Networks

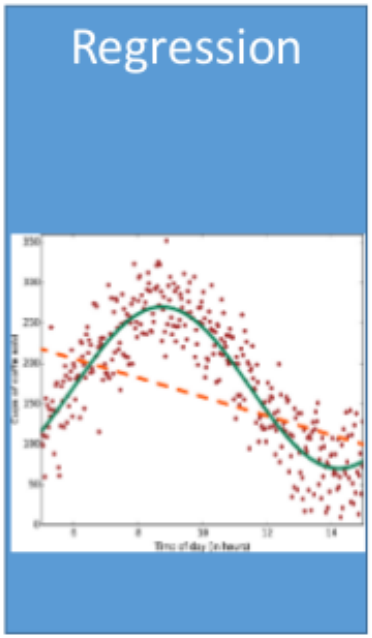
Common classes of algorithms



Supervised



UnSupervised



Supervised

Supervised

Why you need to know these algorithms?

- If you want to answer a *YES|NO* question, it is **classification**
- If you want to predict a *numerical value*, it is **regression**
- If you want to group data into similar observations, it is **clustering**

Classification

- Scenarios:
  - Which customer are more likely to buy, stay, leave
  - Which transactions|actions are fraudulent
  - Which quotes are more likely to become orders
  - Recognition of patterns: speech, speaker, image, movement, etc.
- Algorithms: Decision tree, Logistic Regression, SVM, ANN, etc.

Clustering

- Scenarios:
  - Customer segmentation: divide a customer base into groups of individuals that are similar in specific ways relevant to marketing, such as age, gender, interests, spending habits, etc.
  - Market segmentation
  - Quantization of all sorts, such as, data compression, color reduction, etc.
  - Pattern recognition
- Algorithms: k-means, Gaussian mixture model (GMM), mean shift

Regression

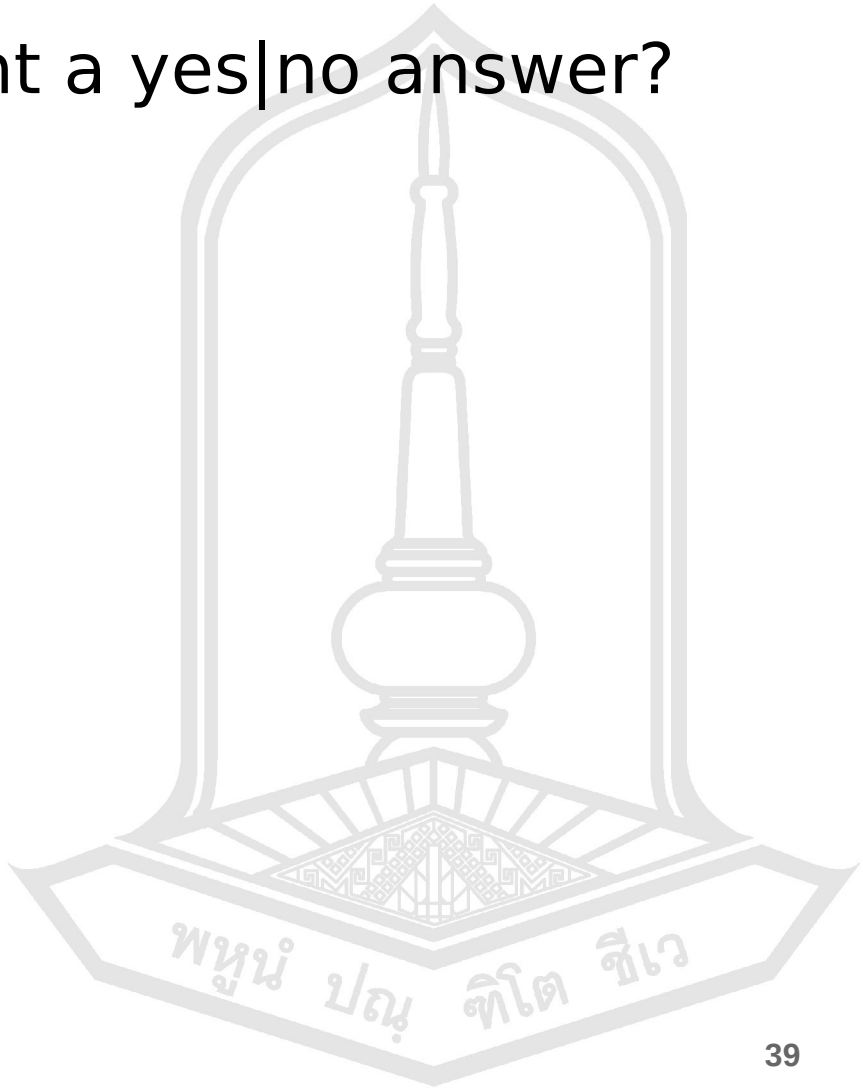
- Scenarios:
  - Stock prices prediction
  - Sales forecasts
  - Premiums on insurance based on different factors
  - Quality control: number of complaints over time based on product specs, utilization, etc.
  - Workforce prediction
  - Workload prediction
- Algorithms: bayesian linear, linear regression, ordinal regression, ANN, Boosted decision tree

Regression vs Classification

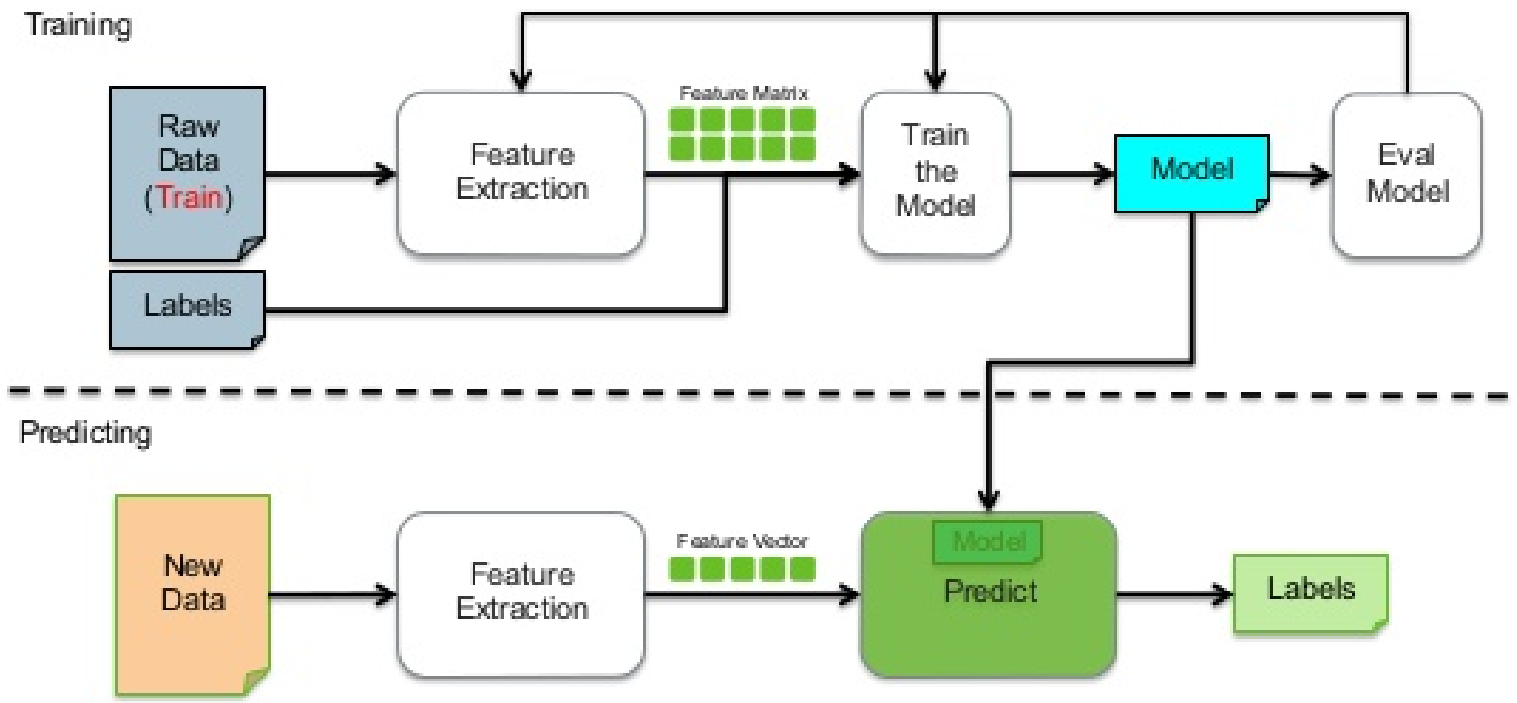
- Does your customer want to predict|estimate a number (regression) or apply a label|categorize (classification)?
  - Regression problems
    - Estimate household power consumption
    - Estimate customer's income
  - Classification problems
    - Power station will|will not meet demand
    - Customer will respond to advertising

Binary vs Multiclass classification

- Does your customer want a yes|no answer?
  - Binary examples
    - Click prediction
    - yes|no
    - over|under
    - win|loss
  - Multiclass examples
    - Kind of tree
    - Kind of network attack
    - Type of heart disease

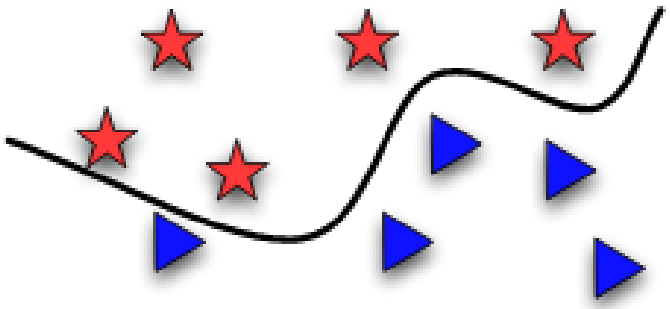


Supervised learning

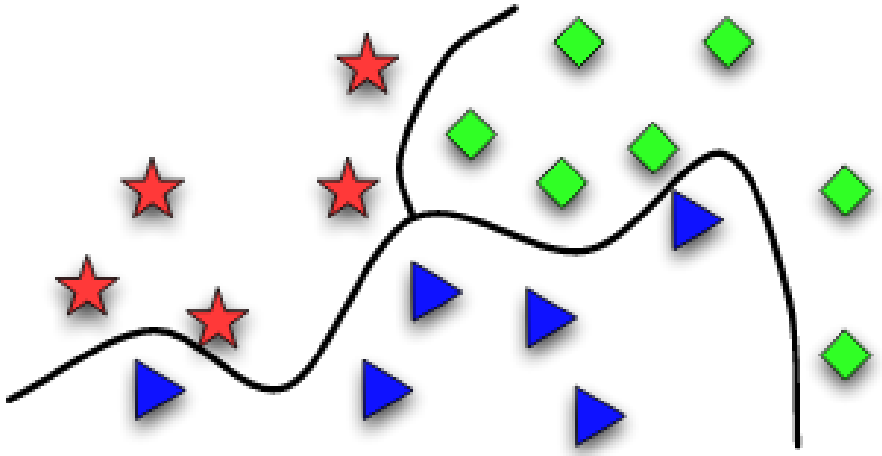


Hortonworks

Classification problems

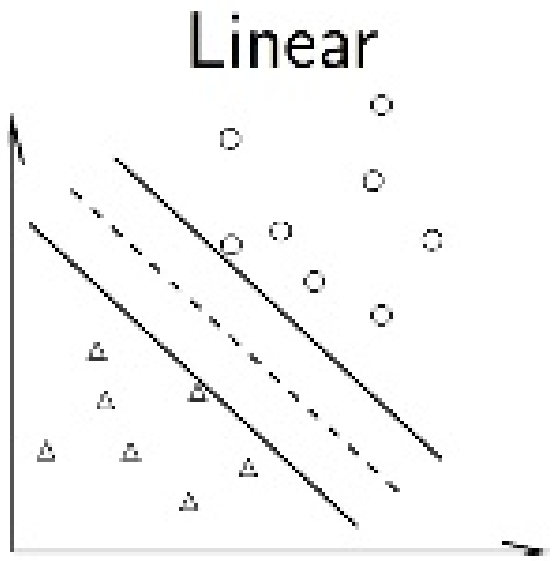


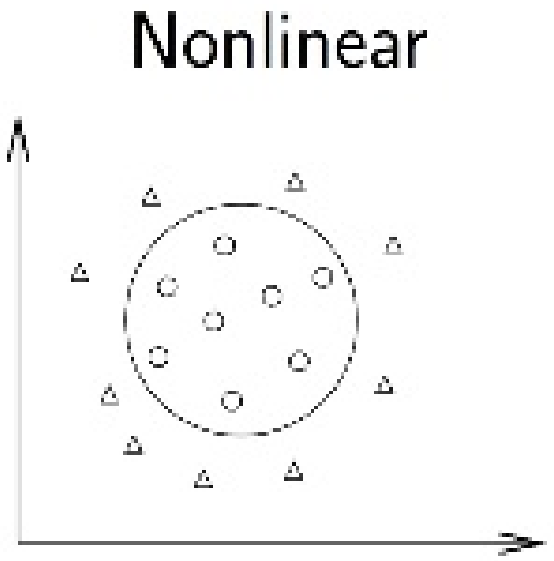
Binary classification



3-class classification

Linear / Nonlinear Classification

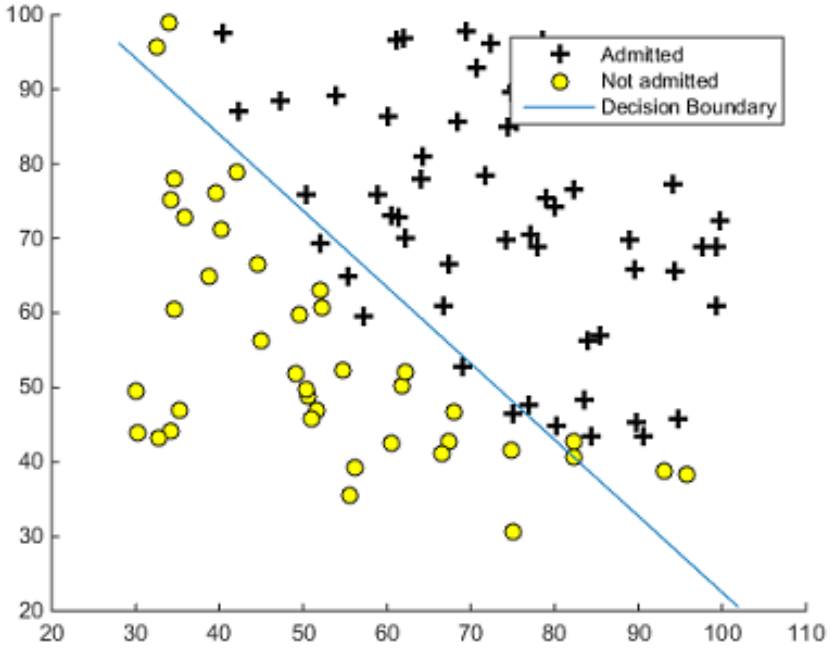


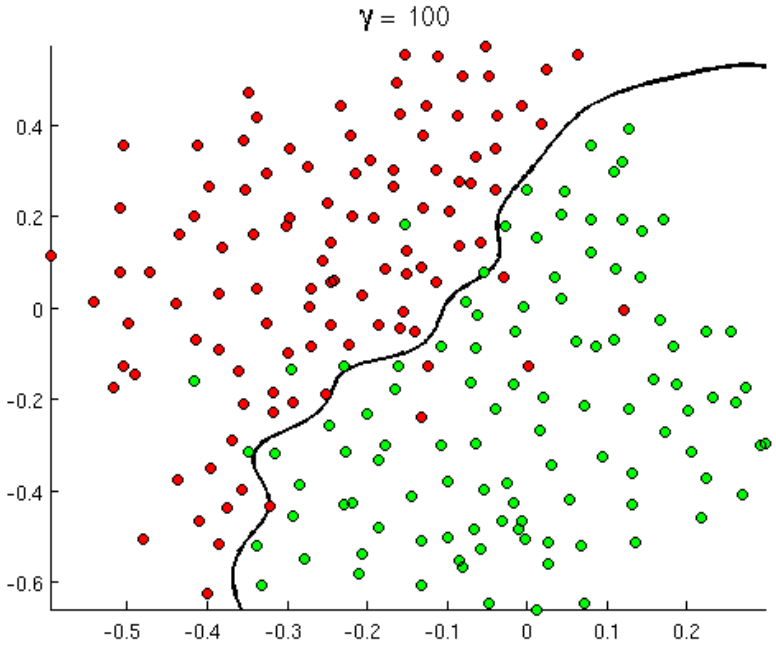


Linear – In the data's original input space, labels can be classified by a linear decision boundary.

Nonlinear – The classifiers have nonlinear, and possibly discontinuous decision boundaries.

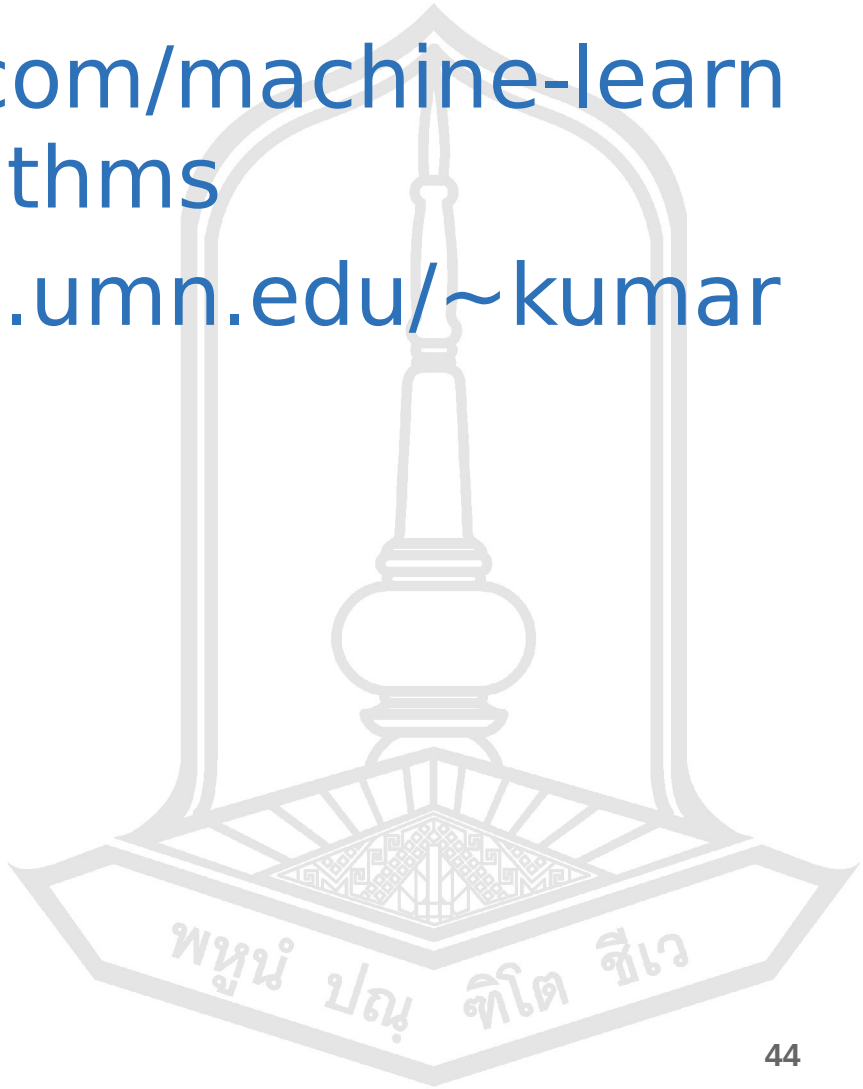
Linear vs Nonlinear





References

- https://www.toptal.com/machine-learning/clustering-algorithms
- http://www-users.cs.umn.edu/~kumar/dmbook/index.php



References

- http://www.coe.utah.edu/~cs4640/slides/Lecture0.pdf
- http://www.imageprocessingplace.com/downloads_V3/root_downloads/tutorials/Image%20Processing-Introduction-Bryan-Mac-Namee.pdf
- https://www.slideshare.net/marafacs/azure-machine-learning-71568316