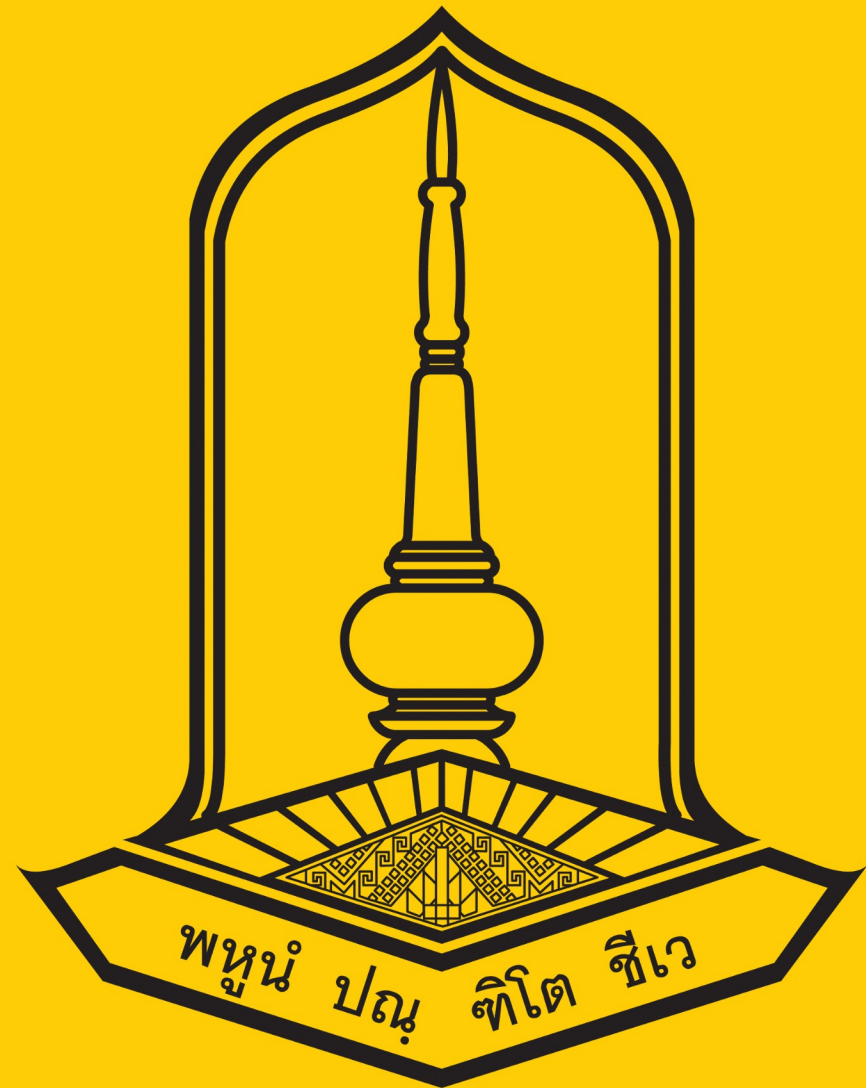


MACHINE LEARNING

1211635

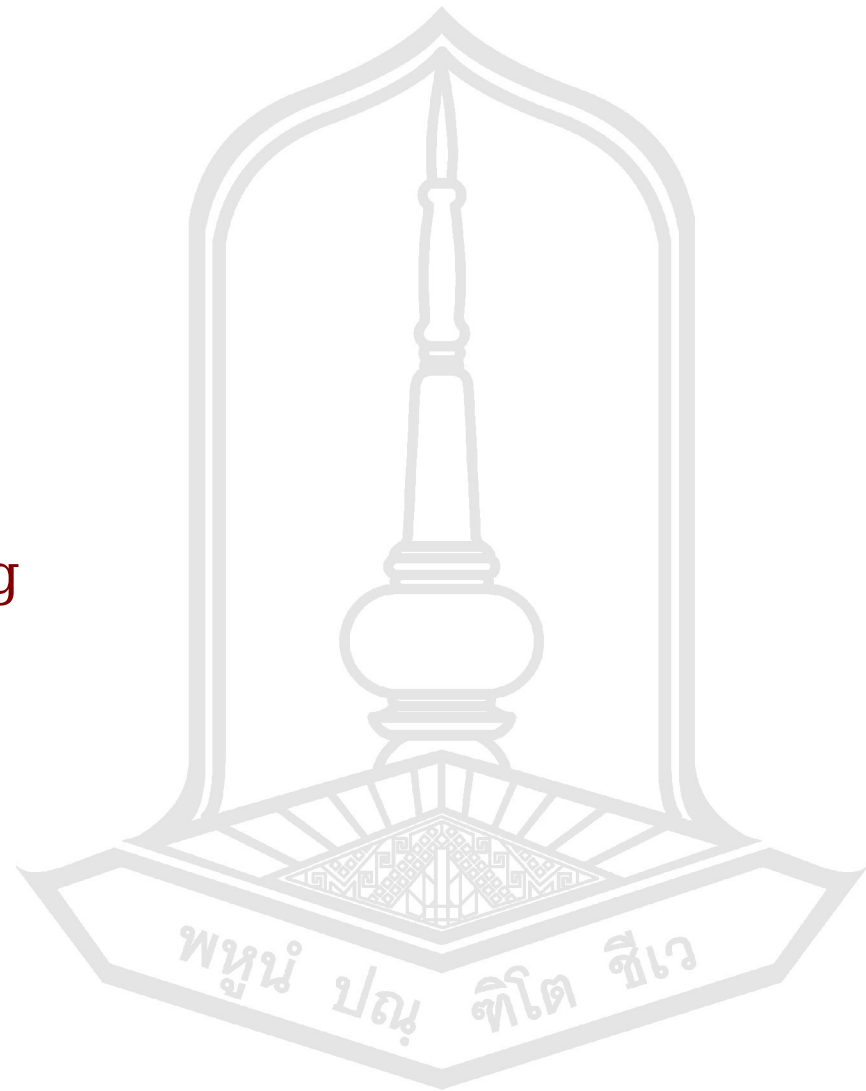
MAHASARAKHAM
UNIVERSITY



MACHINE LEARNING

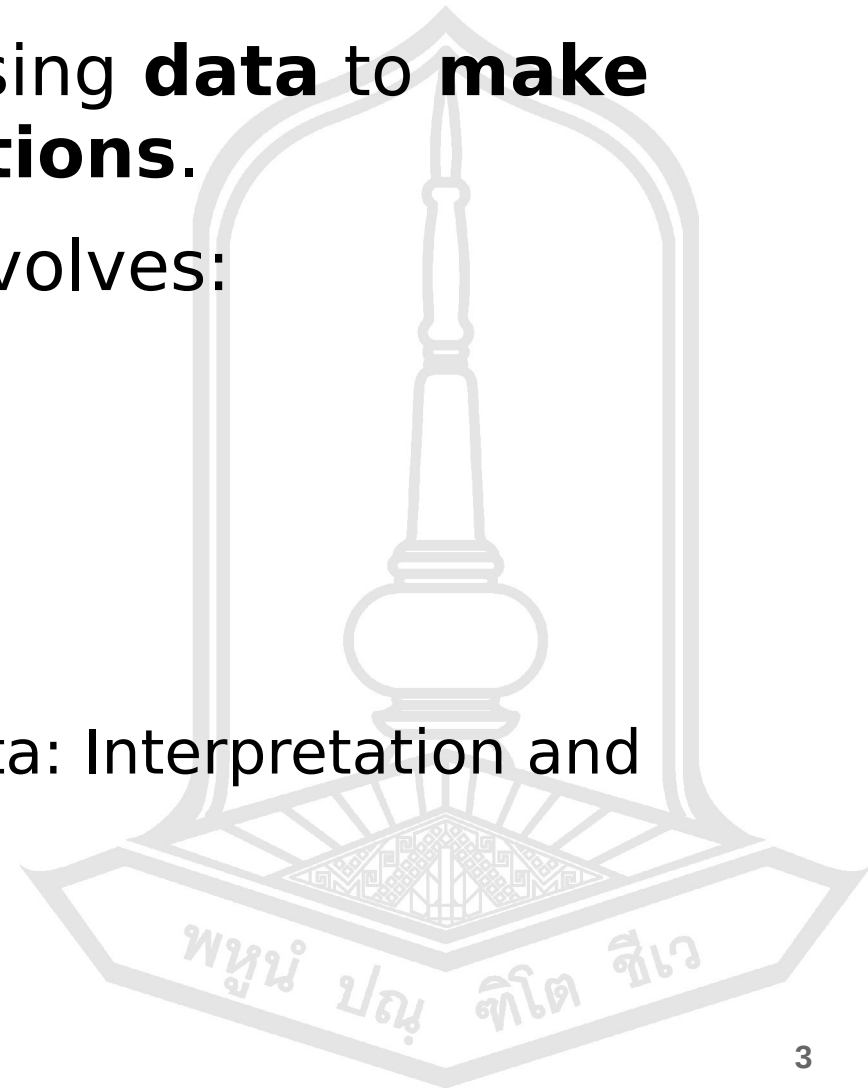
Introduction to machine learning

Olarik Surinta, PhD.
Lecturer



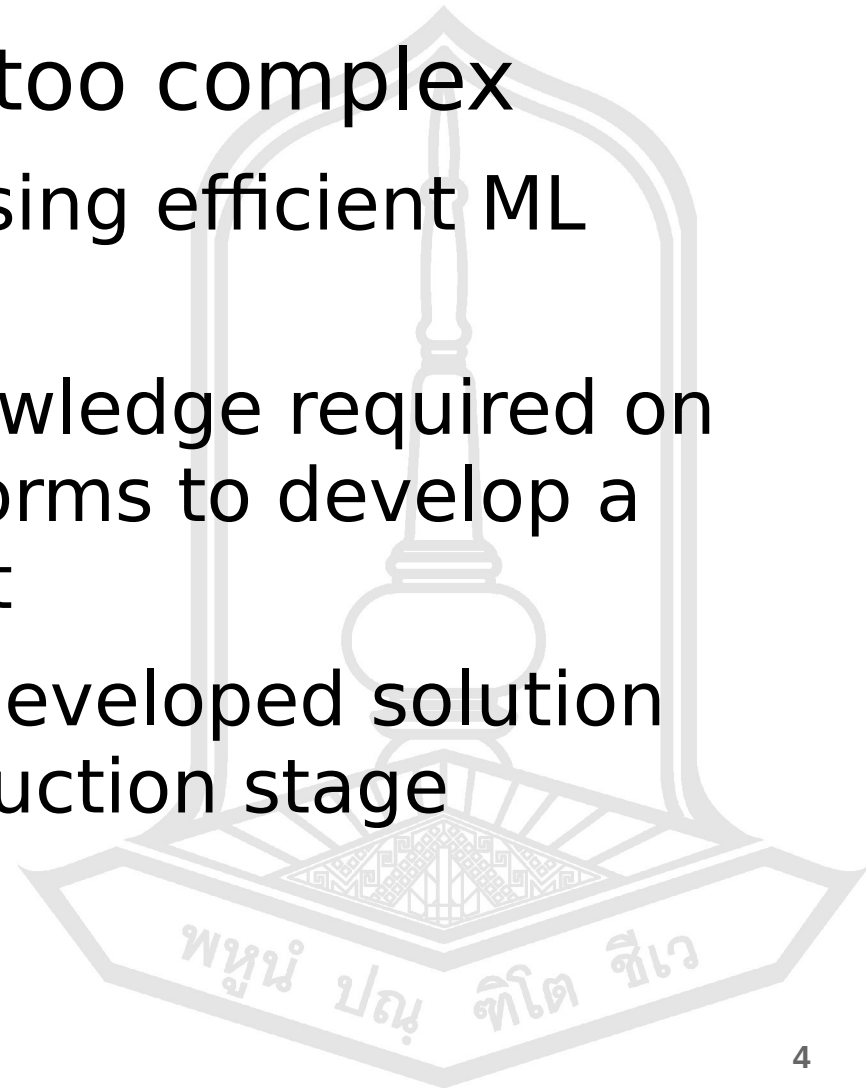
Data Science Involves

- 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



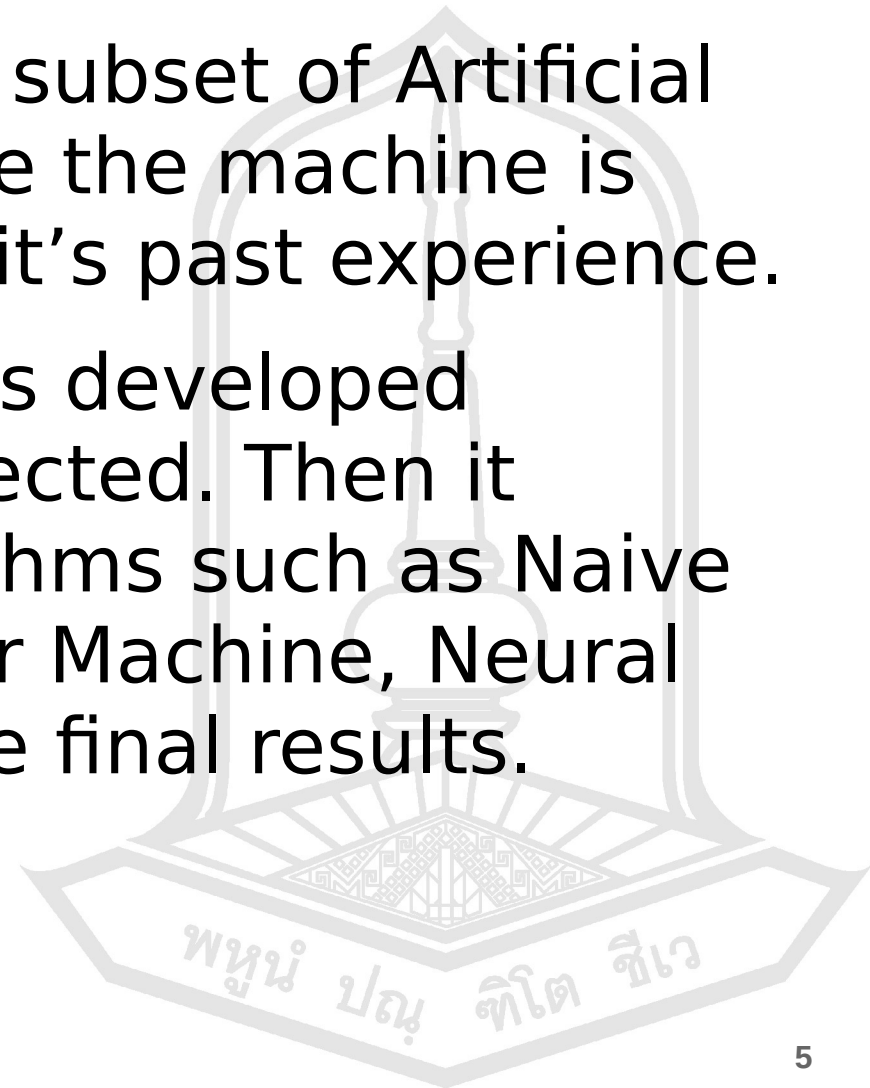
Data Science

- 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

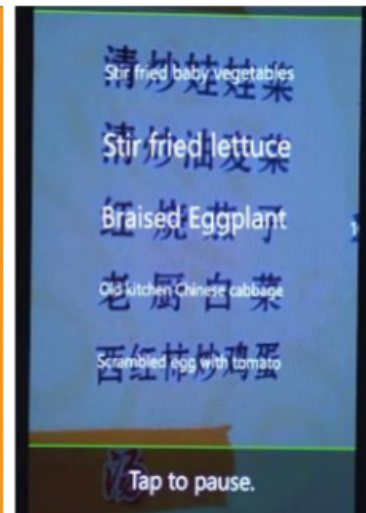
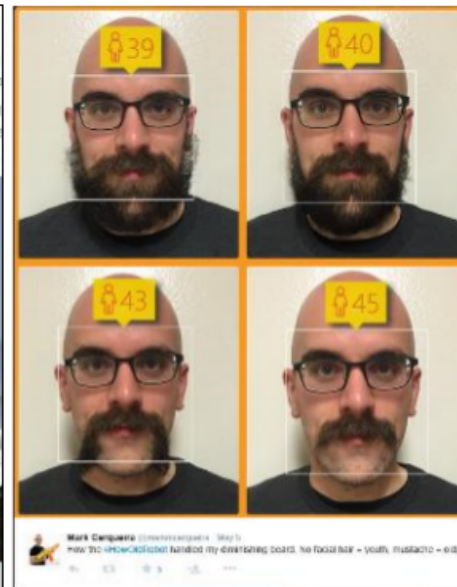
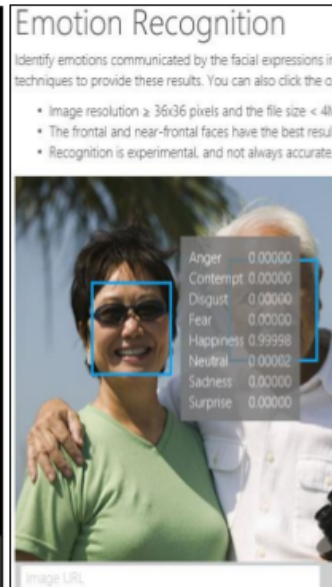
- Using **Known data**, develop a **model** to predict **unknown data**.

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

Model: Known data + Algorithms (ML algorithms)

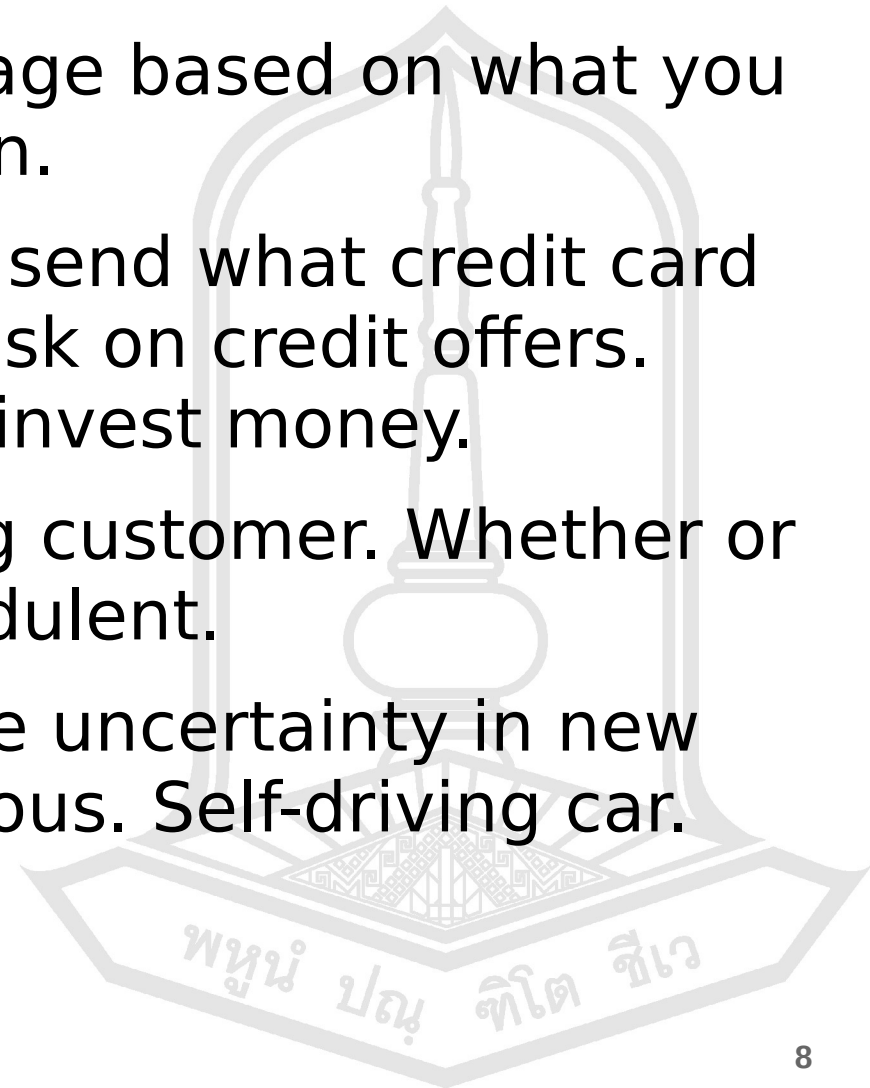
Unknown Data: Missing, Unseen, not existing, future data

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

bing

getting a mortgage in seattle

8,140,000 RESULTS Any time ▾

Ads related to getting a mortgage in seattle

15-Year Mortgage Rates | QuickenLoans.com
www.QuickenLoans.com/Rates
Lock Your Rate. 3.500% (3.92% APR) With America's #1 Online Lender.

Seattle's Best Mortgage Inc
h St
NE 2nd Pl
NE 3rd Pl
NE 1st St
Main St
© 2013 Microsoft Corporation
© 2013 Nokia
St Ste B306 · Bellevue
0 · Directions
Website: seattlesbm.com
Report a problem

Including results for getting a mortgage in seattle.
Do you want results only for getting a mortgage in seattle?

Seattle Mortgage Rates - Find the Best Home Loan | Zillow
www.zillow.com/mortgage-rates/wa/seattle
See up to the minute Seattle mortgage rates and find Seattle Washington's best, lowest possible quote with Zillow Mortgage Marketplace.

Seattle's Best Mortgage
www.seattlesbm.com
Get the best mortgage loan for you at Seattle's Best Mortgage. (CL#117721) When you decide to buy a home or refinance a mortgage, it's a big step.

1 11911 Ne 1st St Ste B306, Bellevue · (425) 228-7000 · Directions · Bing Local

Ads related to getting a mortgage in seattle

What language?

What is the probability of a click on each ad?

Which ads to show, and in what order?

Machine learning enables nearly every value proposition of web search.

What is the intent?

Are any of these pages malicious?

Misspelled?

Which links are most likely to get clicked?

What pages should we index?

What ad pricing will optimize revenue?

Google

airplane



All Images

Images

Videos

News

Maps

More

Settings

Tools

View saved

SafeSearch

1st class

the wright brothers

can

coloring

military

jet

helicopter

bomber

drone

fighter





airplane



1

All

Images

Videos

Maps

News

More

Settings

Tools

View saved

SafeSe...

old

ww2

ww1

war

vintage

old fashioned

civil

ancient

black and white

fighter plane

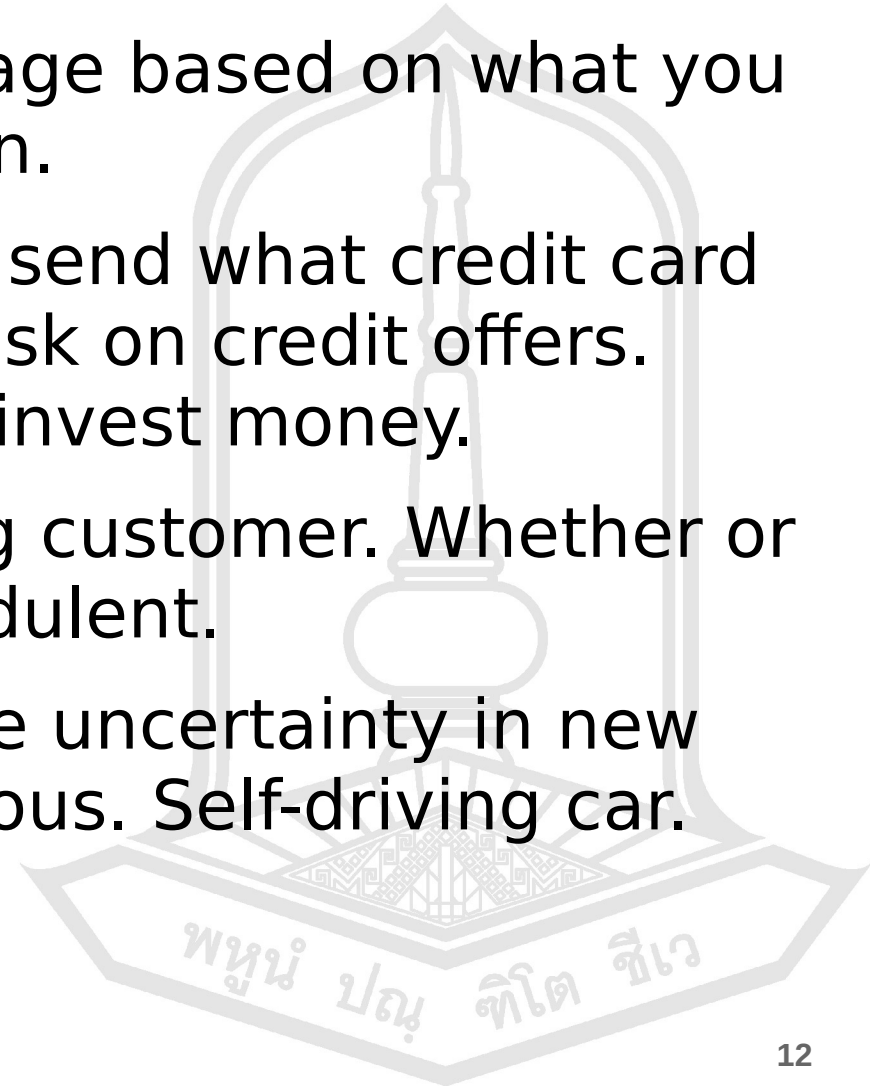


German Fokker DR 1 Triplane



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

Text Analytics - Preview by [Azure Machine Learning](#) Demo Sample Code Doc

Try out the Azure ML [Text Analytics service](#) - for free.

To use the service in production, you can get access to its API by [signing up](#) for it (click on "Sign Up" on the right pane on that page). For questions/comments, please use the "Feedback" button in the top right.

I love this presentation. I found it very useful.

Analyze

Sentiment:

98 %

Key phrases highlighted below:

I love this **presentation**. I found it very useful.

Negative

Text Analytics - Preview by [Azure Machine Learning](#) Demo S

Try out the Azure ML [Text Analytics service](#) - for free.

To use the service in production, you can get access to its API by [signing up](#) for it (c the right pane on that page). For questions/comments, please use the "Feedback" b

I hate this presentation. I found it very useless.

Analyze

Sentiment:

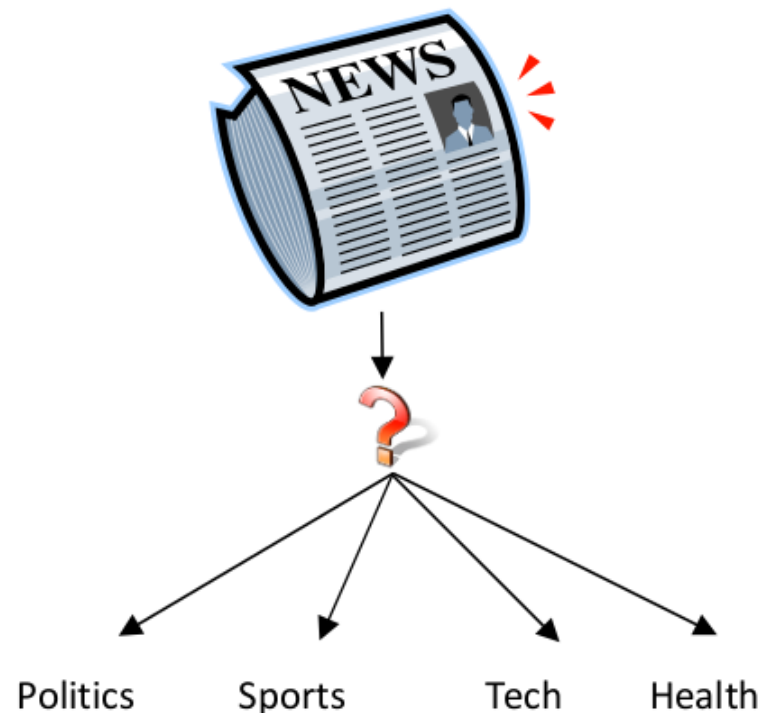
7 %

Key phrases highlighted below:

I hate this **presentation**. I found it very useless.

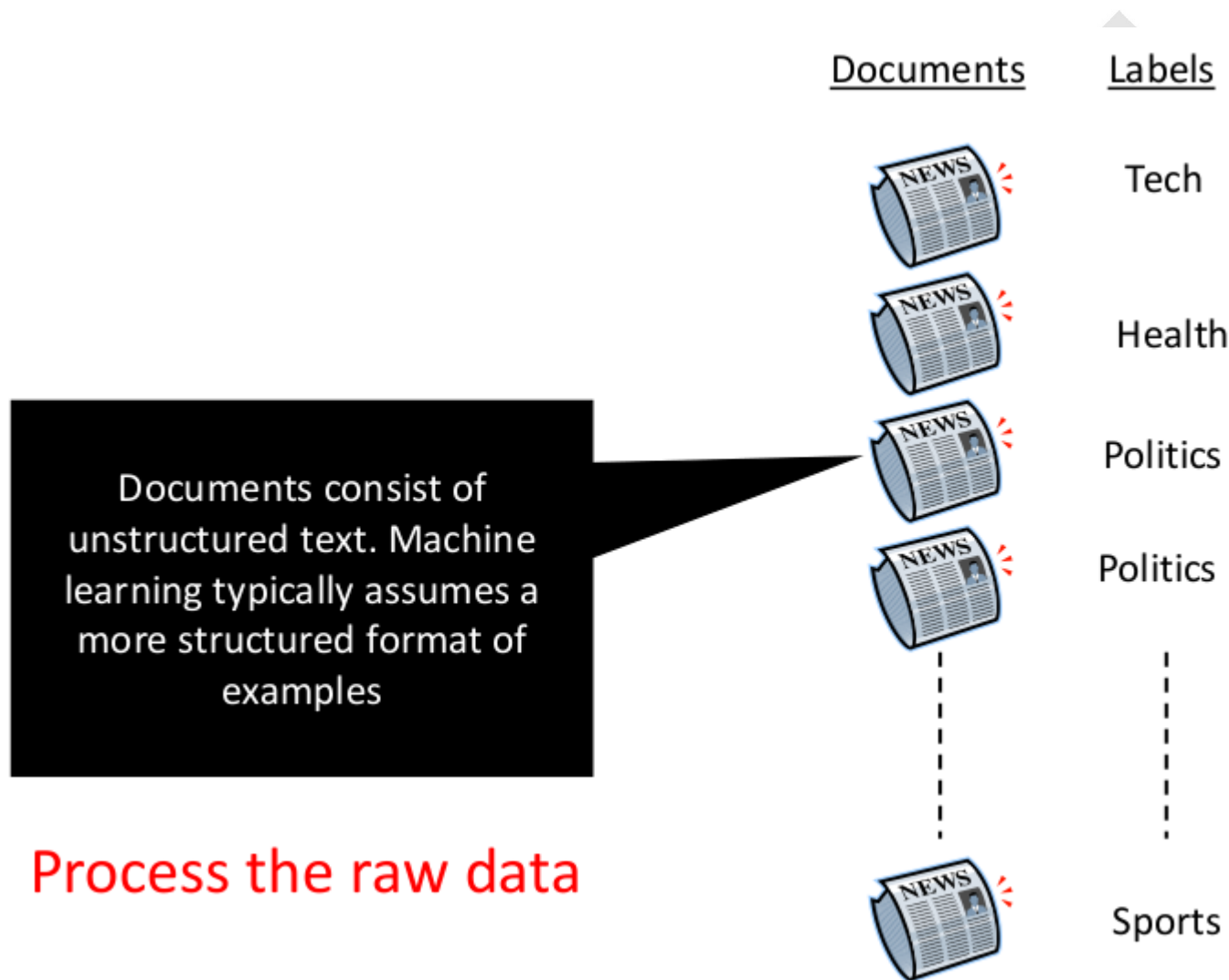
Text classification

- 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)

Documents Labels



Tech



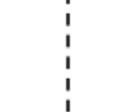
Health



Politics



Politics



Sports

Process each data instance to represent it as a feature vector

Diagram illustrating the process of converting documents into feature vectors for training data.

The diagram shows a list of documents on the left, each associated with a label (Tech, Health, Politics, Politics, Sports). A blue box labeled "Feature" points to a specific document in the list. A large blue arrow points from the "Feature" box to a detailed view of the document's feature vector representation.

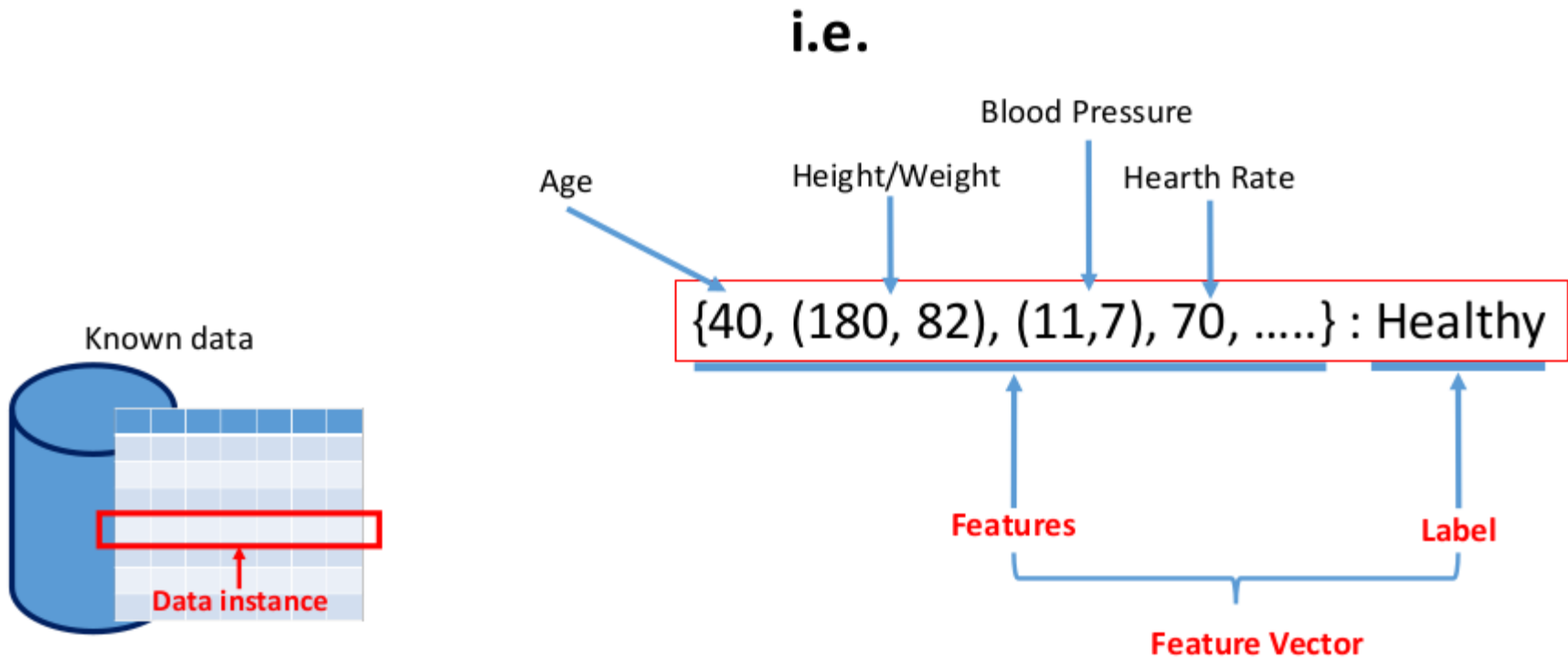
The detailed view shows a table with columns labeled "Documents" and "Labels". The "Documents" column contains a list of documents, and the "Labels" column contains the corresponding labels. A blue box labeled "Feature" points to a specific document in the "Documents" column, and a blue circle highlights the corresponding label in the "Labels" column.

Documents	Labels
Document 1	Tech
Document 2	Health
Document 3	Politics
Document 4	Politics
Document 5	Sports

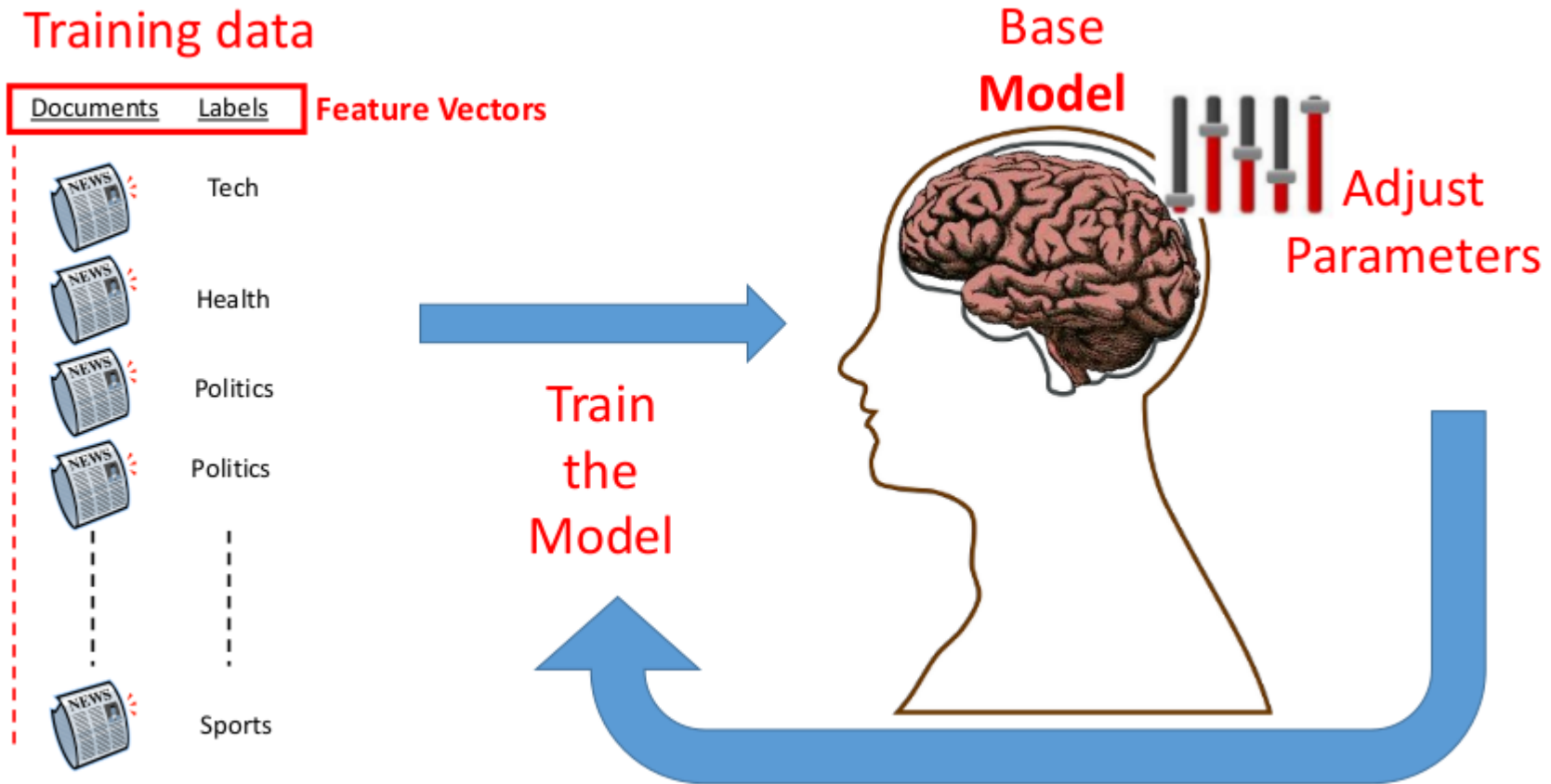
MAHASARAKHAM
UNIVERSITY

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

Feature vector

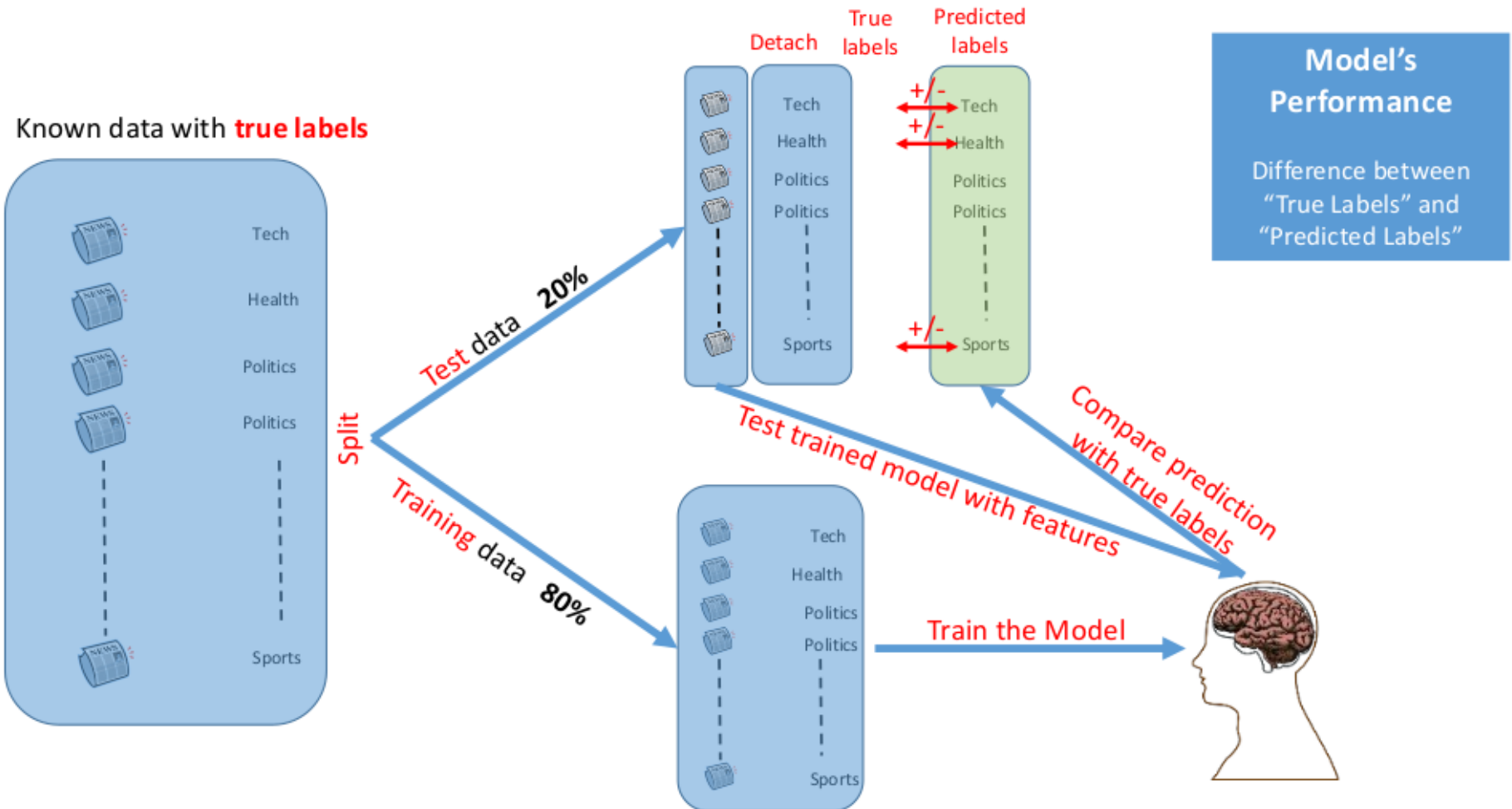


Developing a model

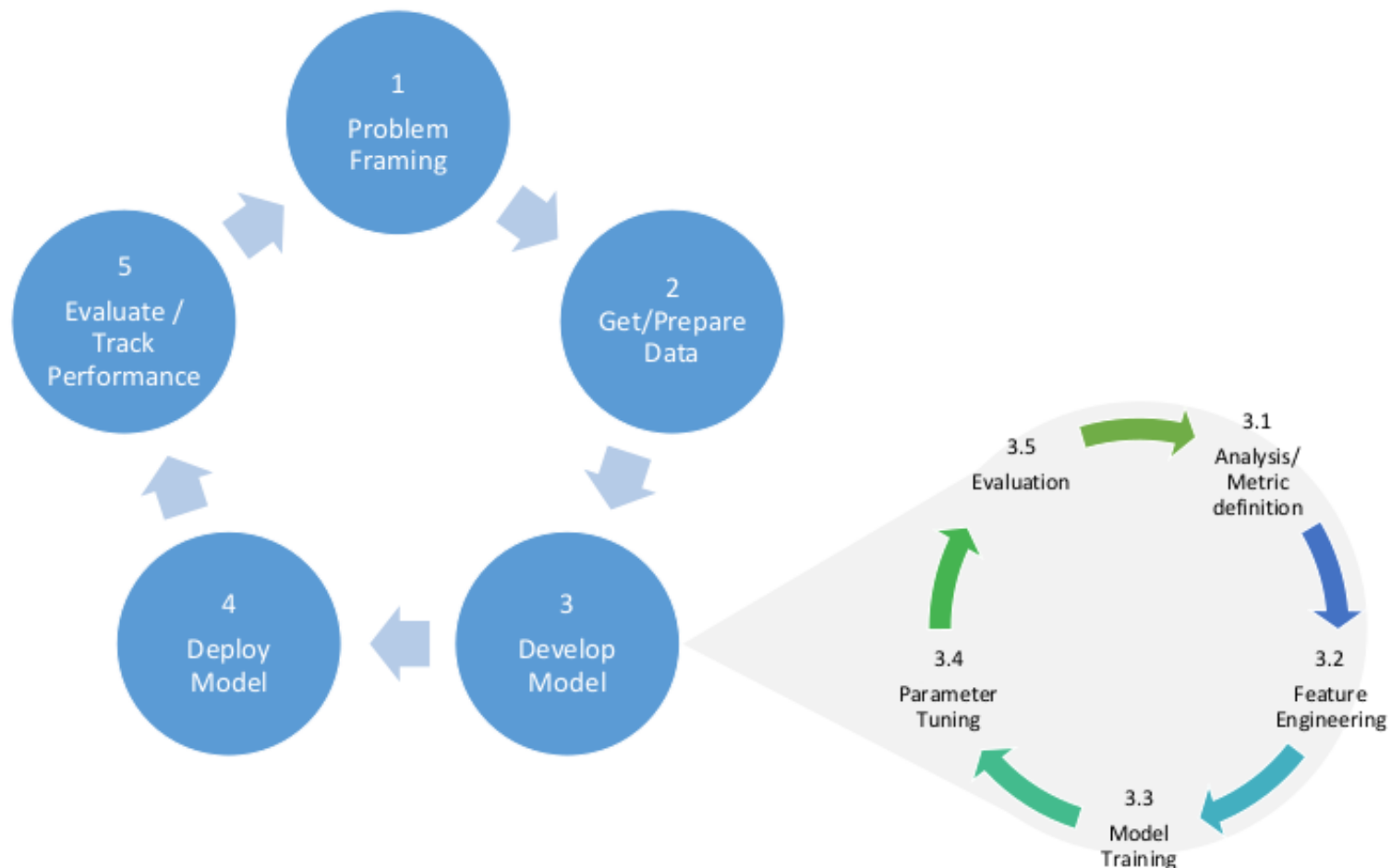


Using **known data**, develop a **model** to predict **unknown data**.

Model's Performance

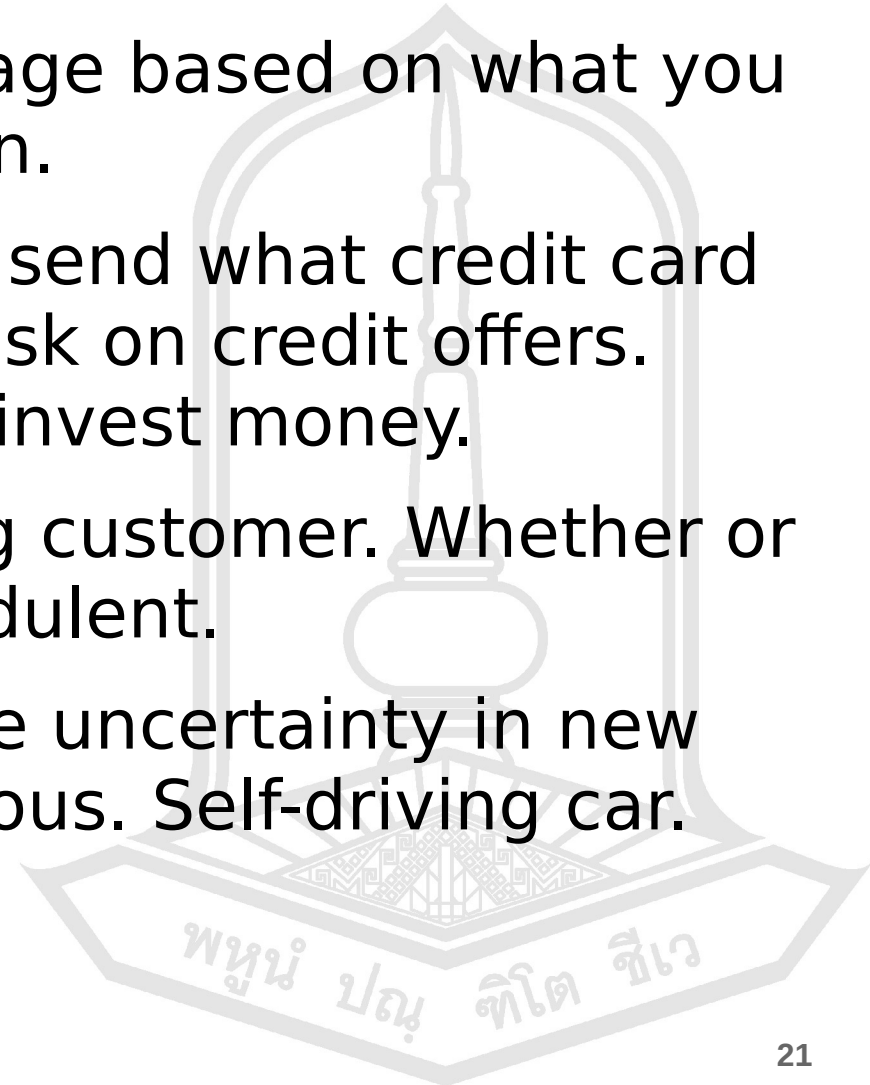


Steps to build a machine learning solution



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.



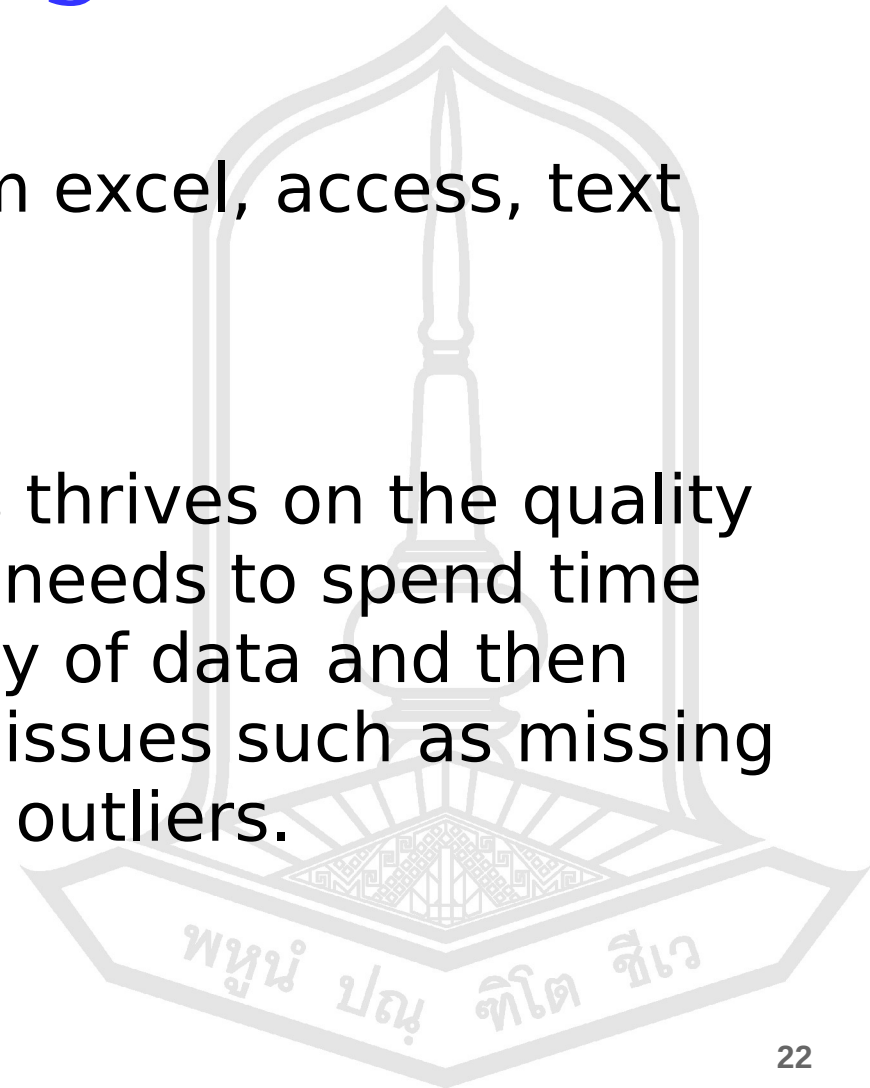
What are the steps used in machine learning

- **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.



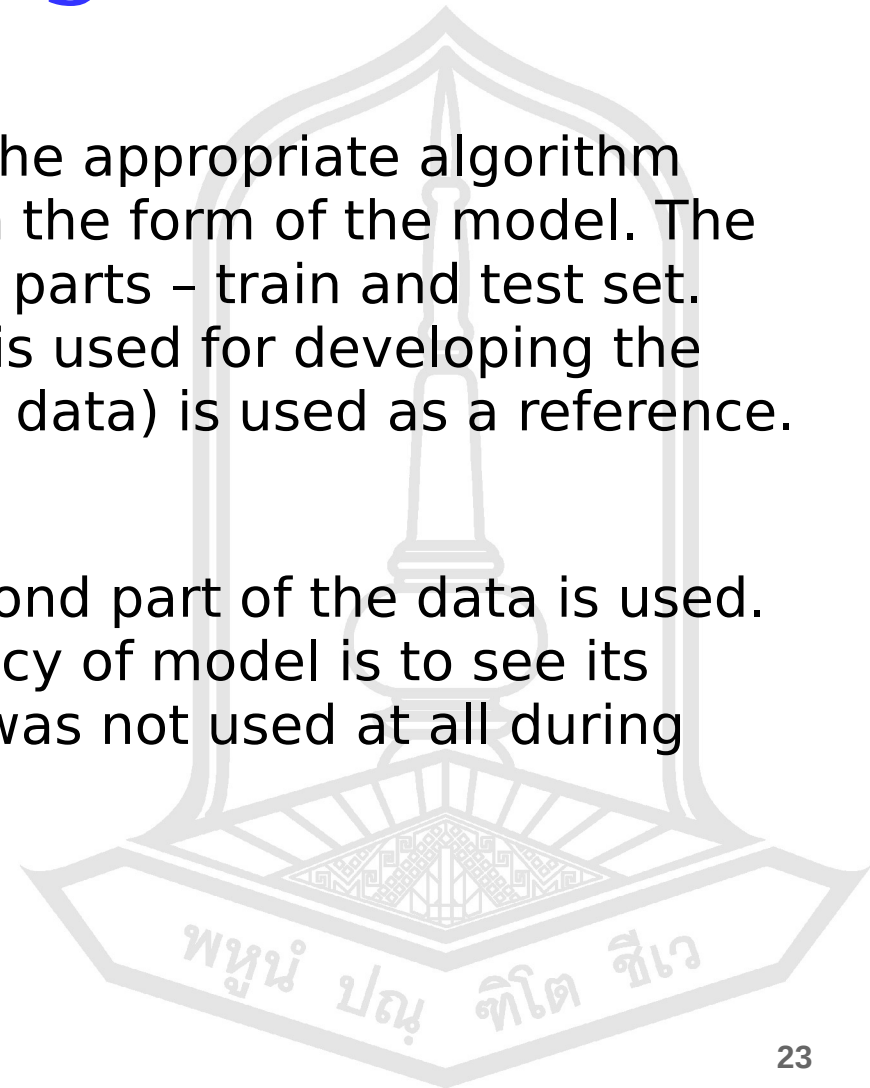
What are the steps used in machine learning

- **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.

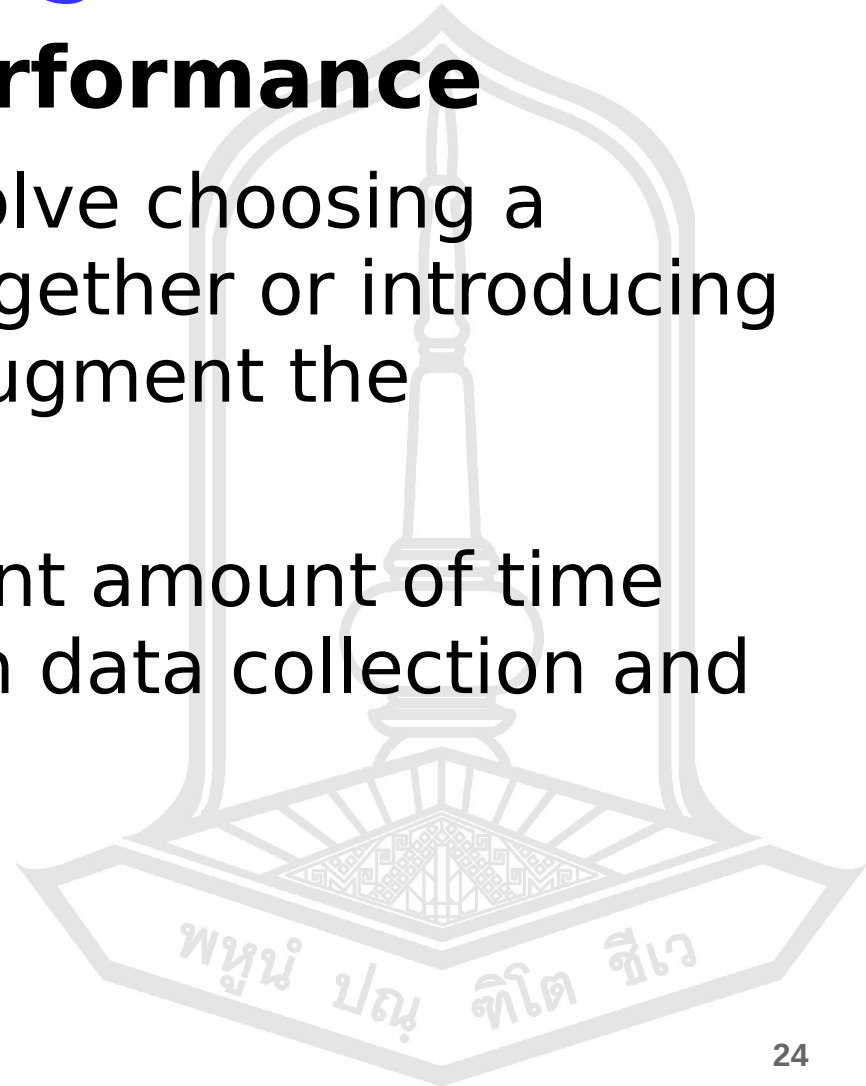
- **Evaluating the model**

- 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.



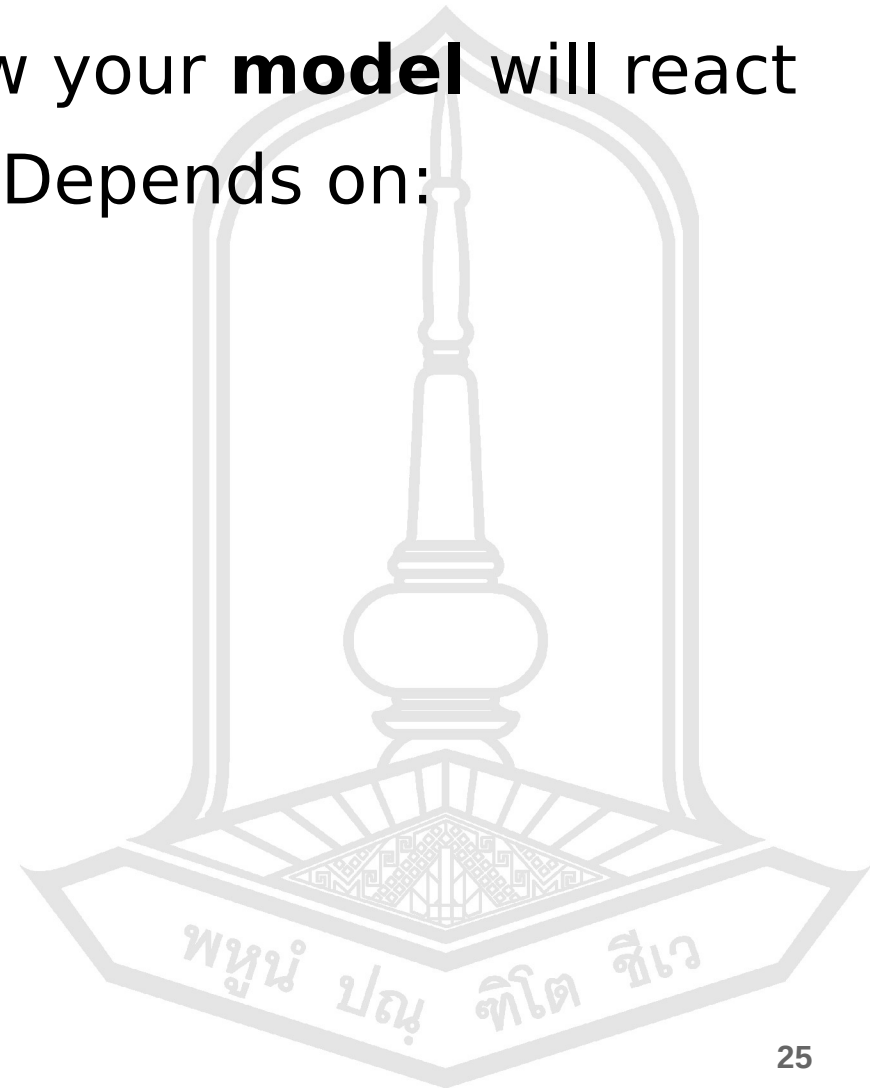
What are the steps used in machine learning

- **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

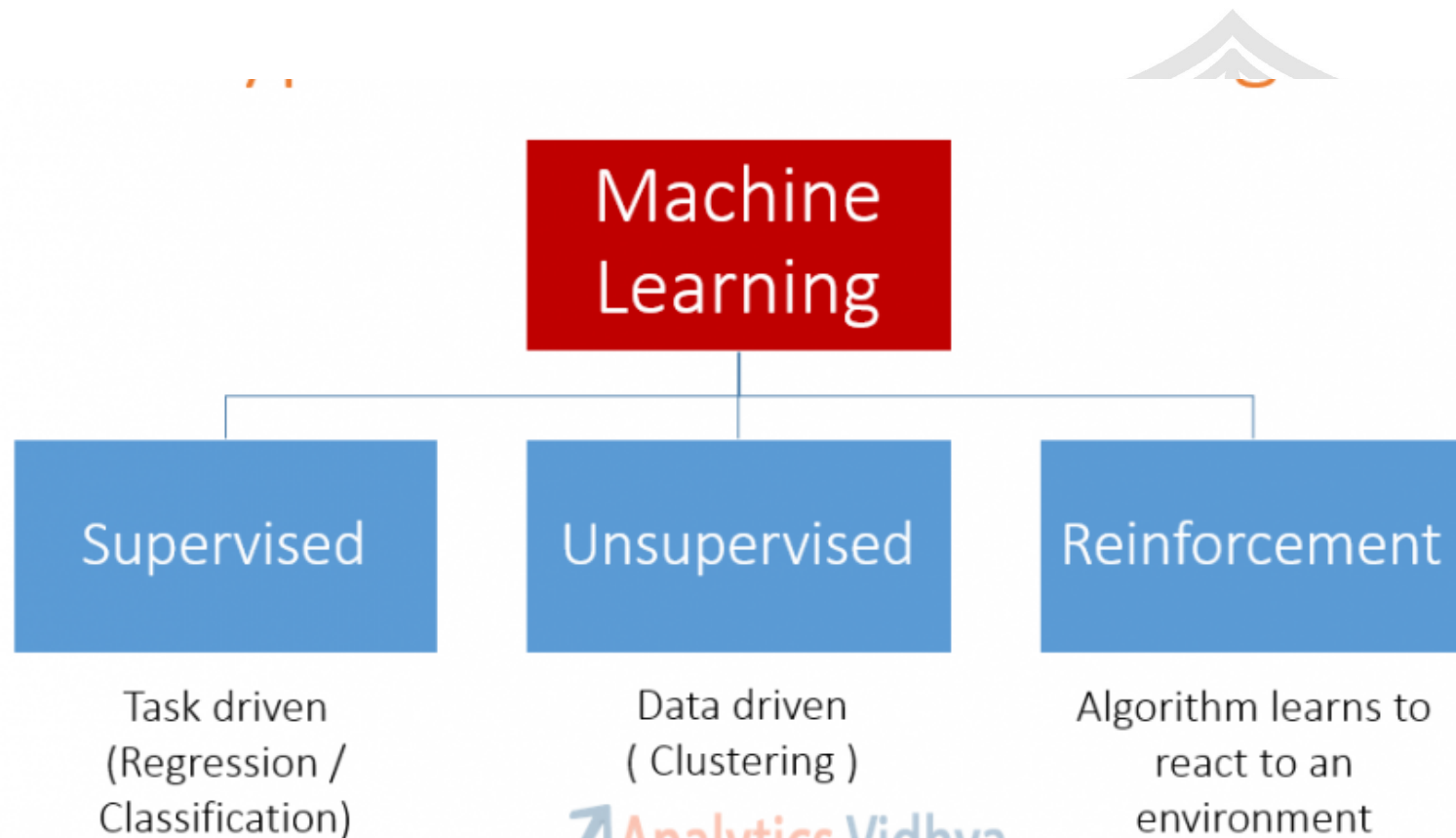


Machine learning algorithms

- 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

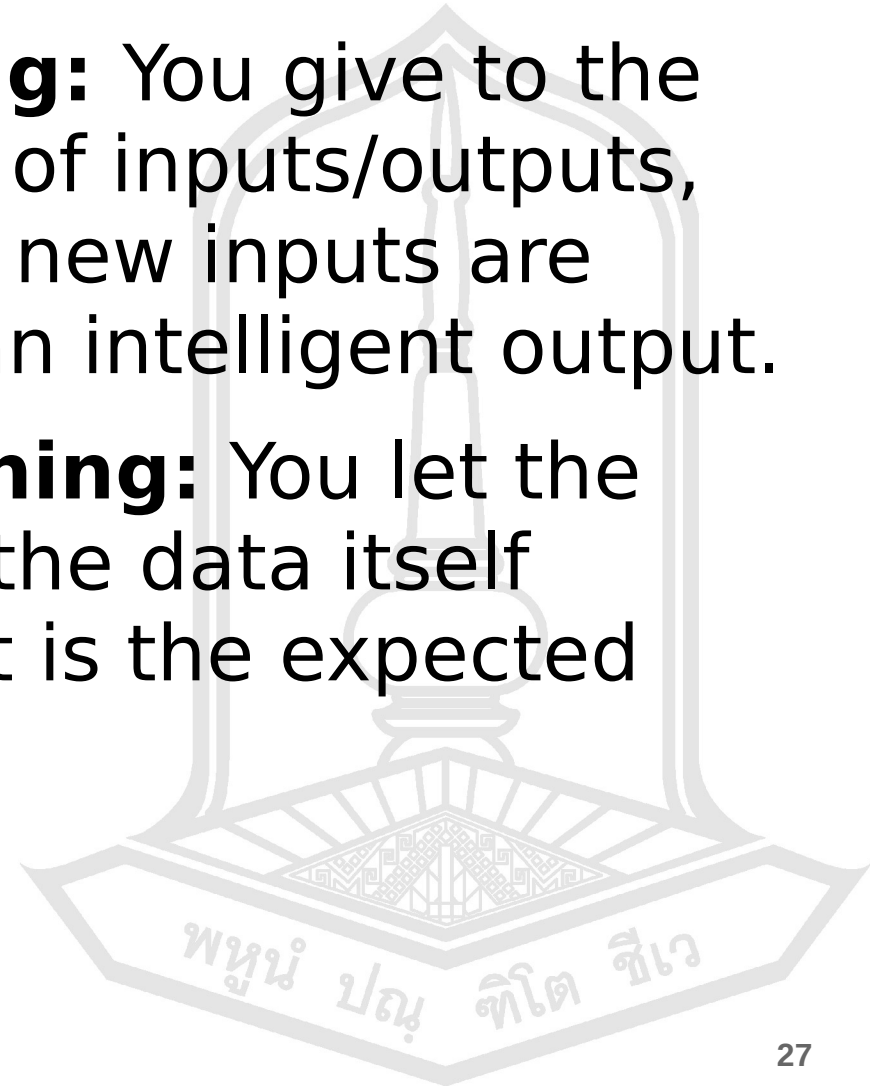


Types of machine learning



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.



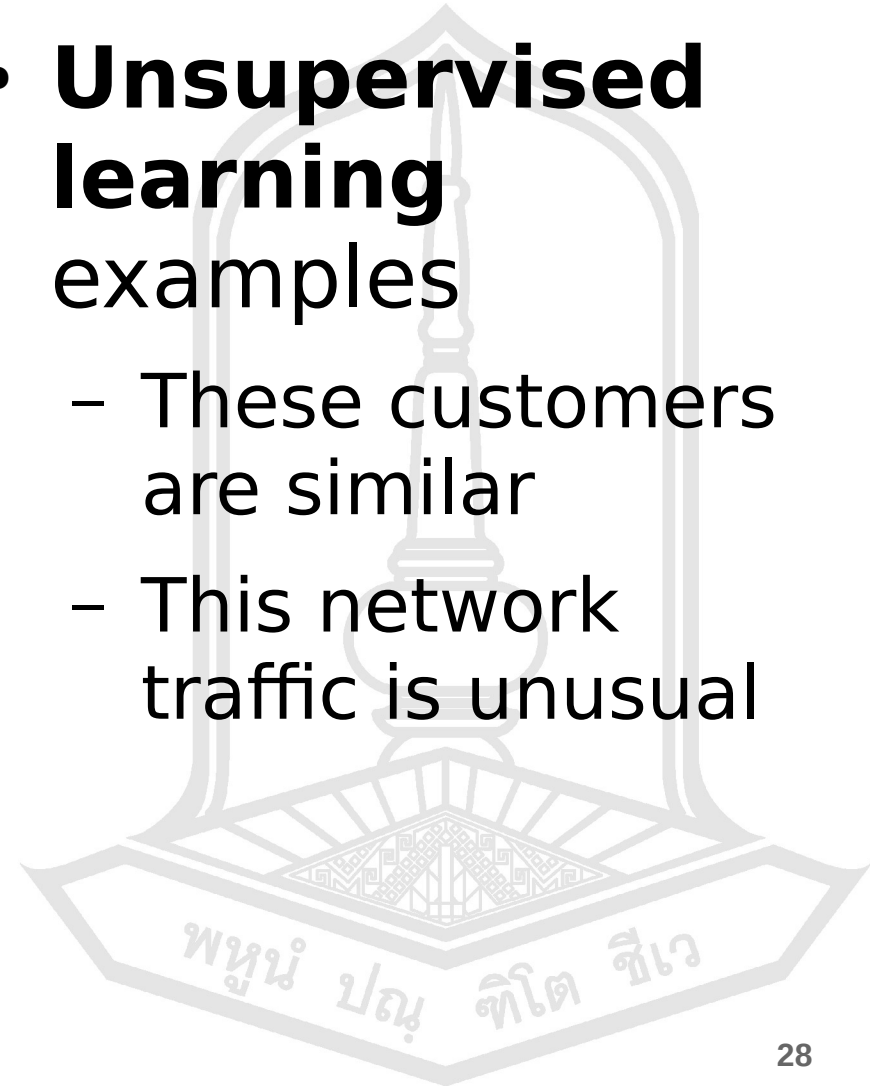
Supervised vs Unsupervised learning

- **Supervised learning**
examples

- This customer will like coffee
- This network traffic indicates a denial of service attack

- **Unsupervised learning**
examples

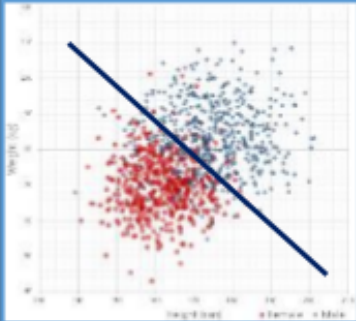
- These customers are similar
- This network traffic is unusual



Common classes of algorithms

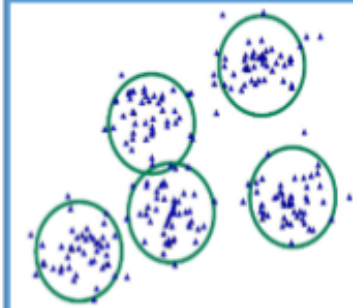


Classification



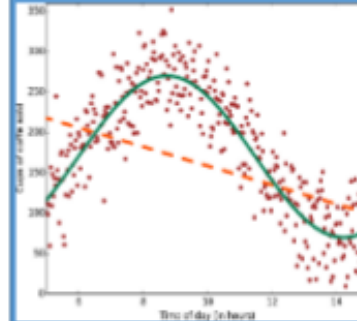
Supervised

Clustering



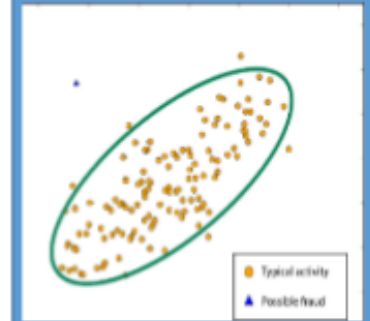
UnSupervised

Regression



Supervised

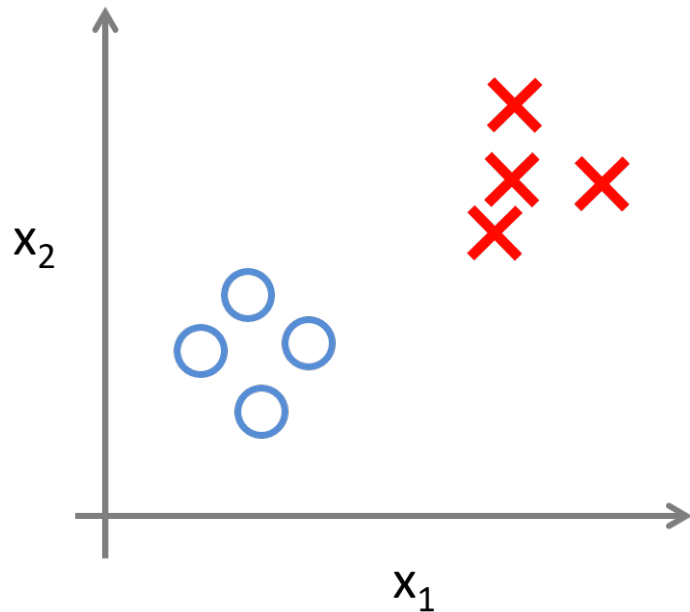
Anomaly Detection



Supervised

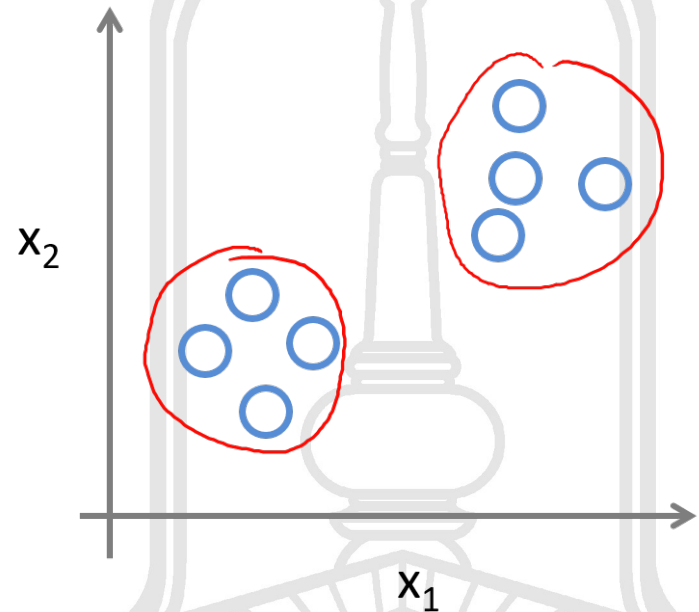
Supervised vs Unsupervised learning

Supervised Learning

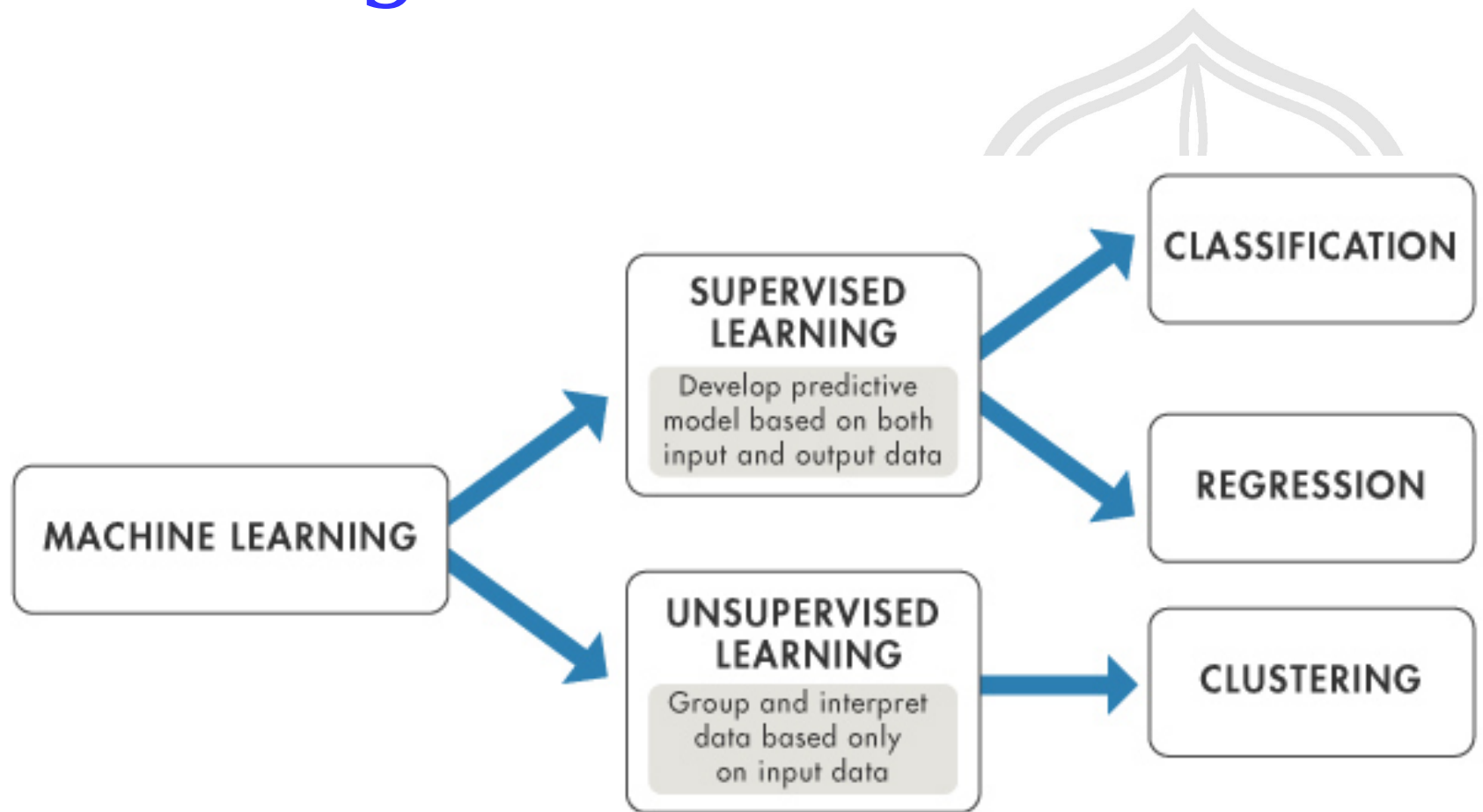


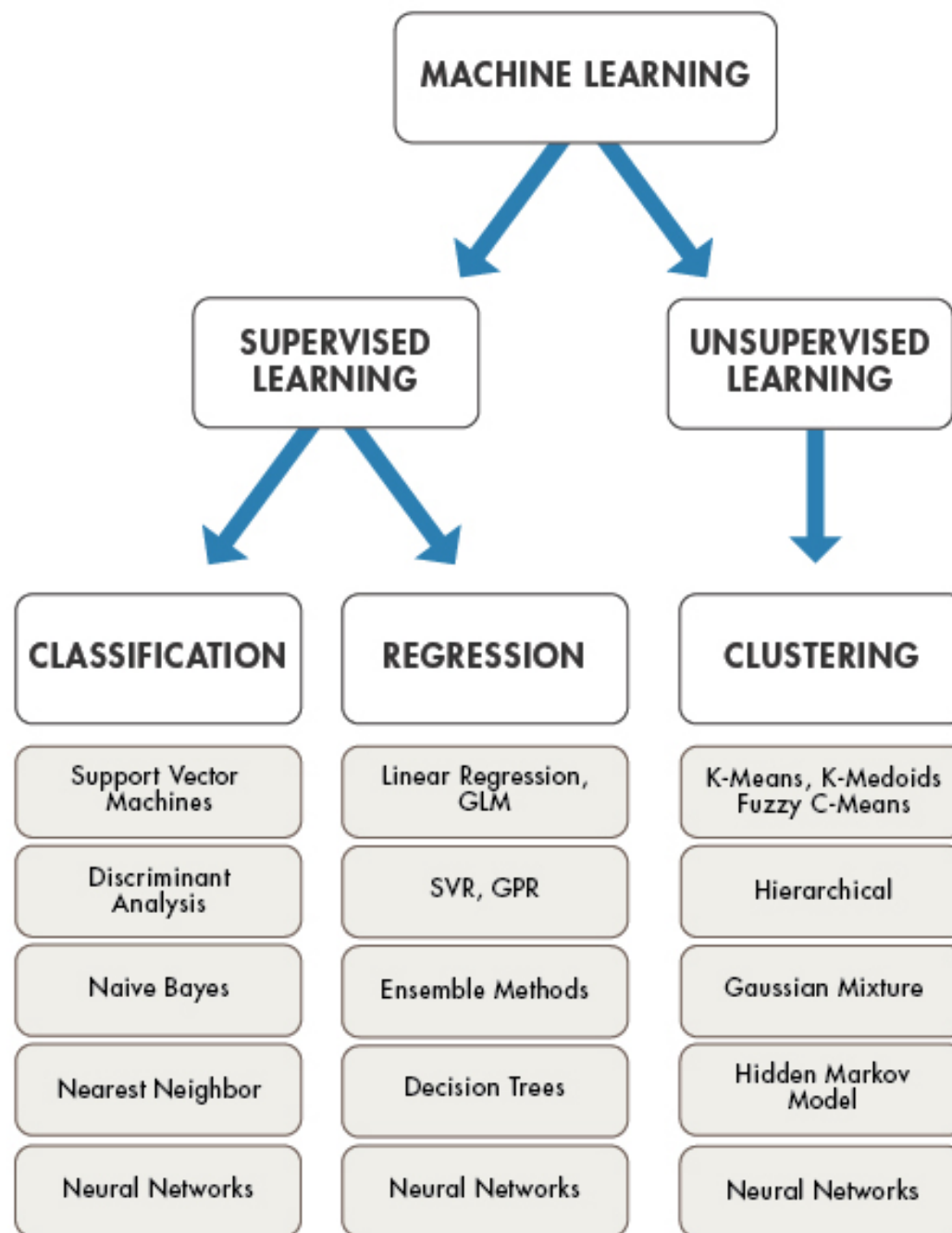
MAHASARAKHAM
UNIVERSITY

Unsupervised Learning



Supervised vs Unsupervised learning

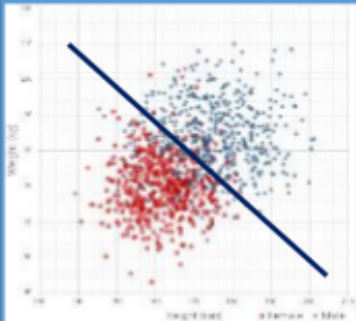




Common classes of algorithms

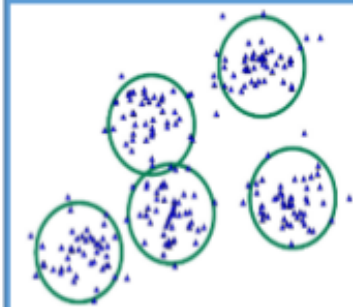


Classification



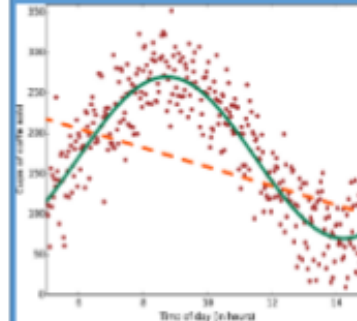
Supervised

Clustering



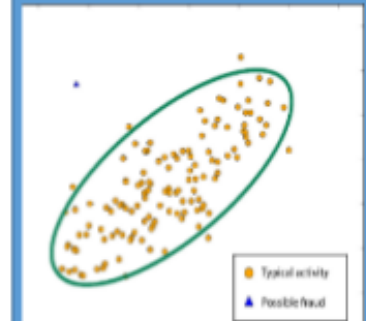
UnSupervised

Regression



Supervised

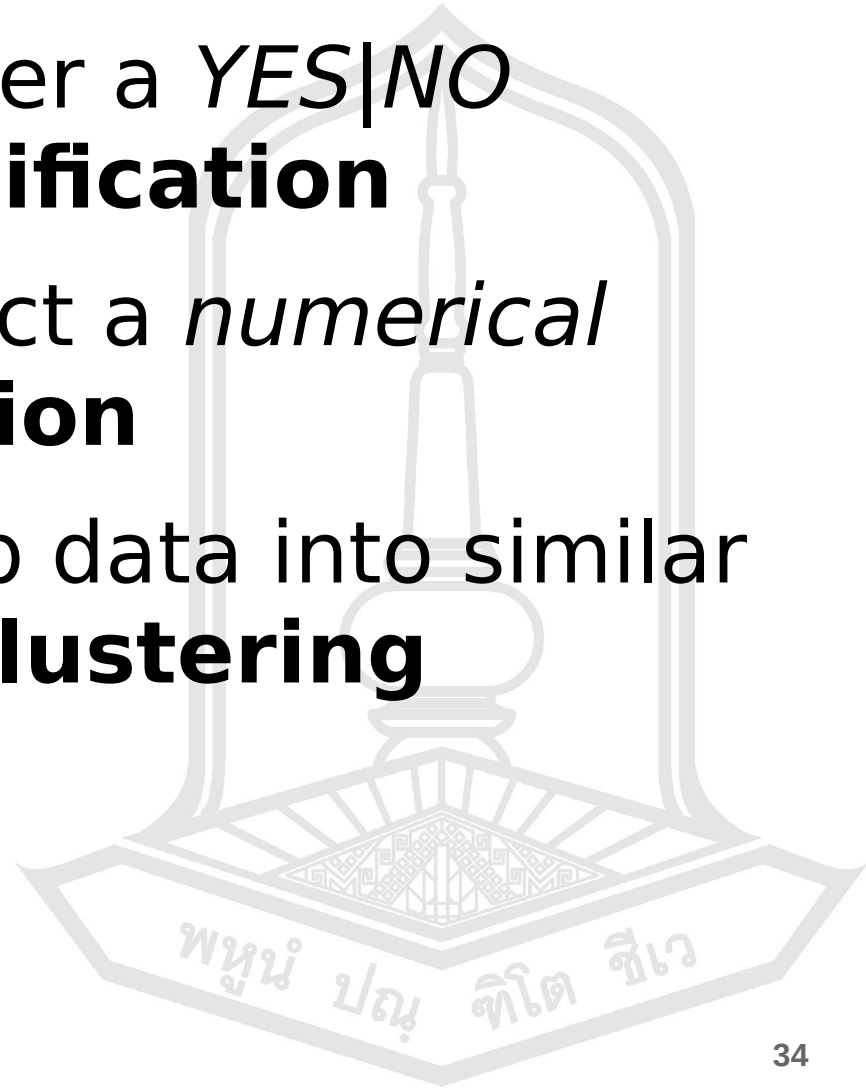
Anomaly Detection



Supervised

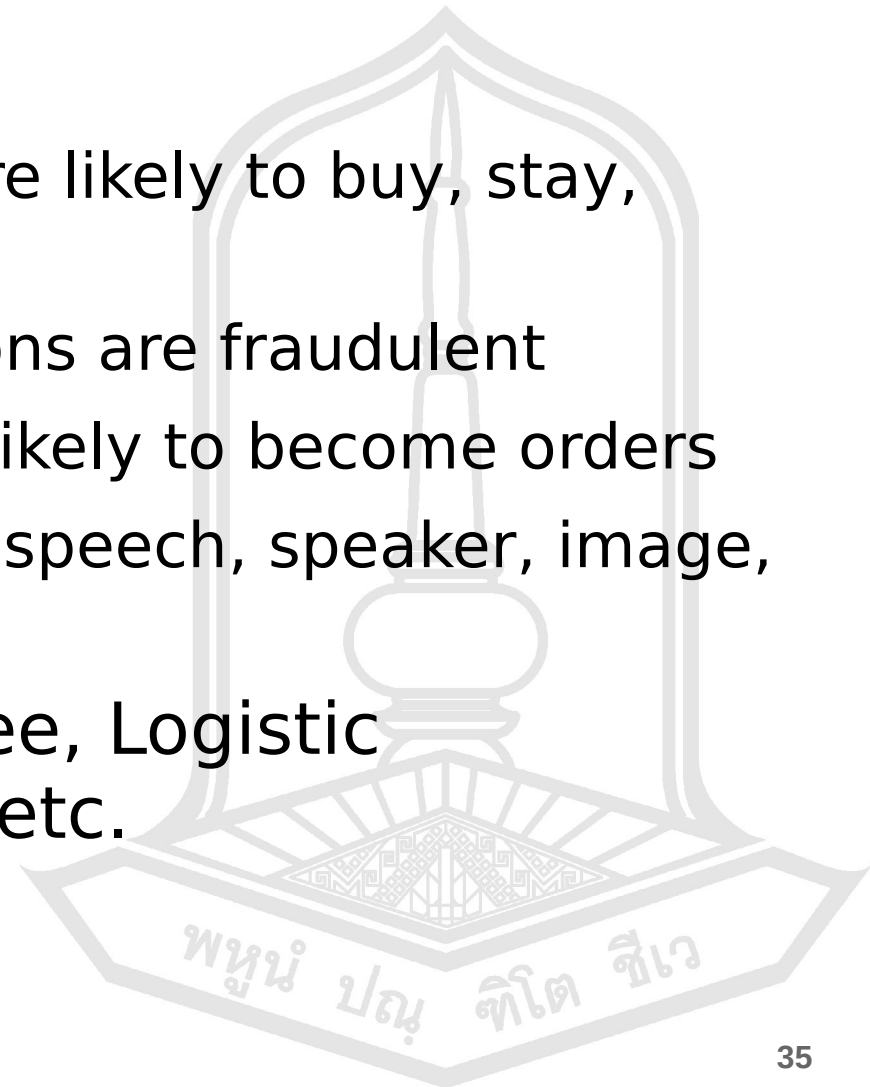
Why you need to know these algorithms?

- 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**



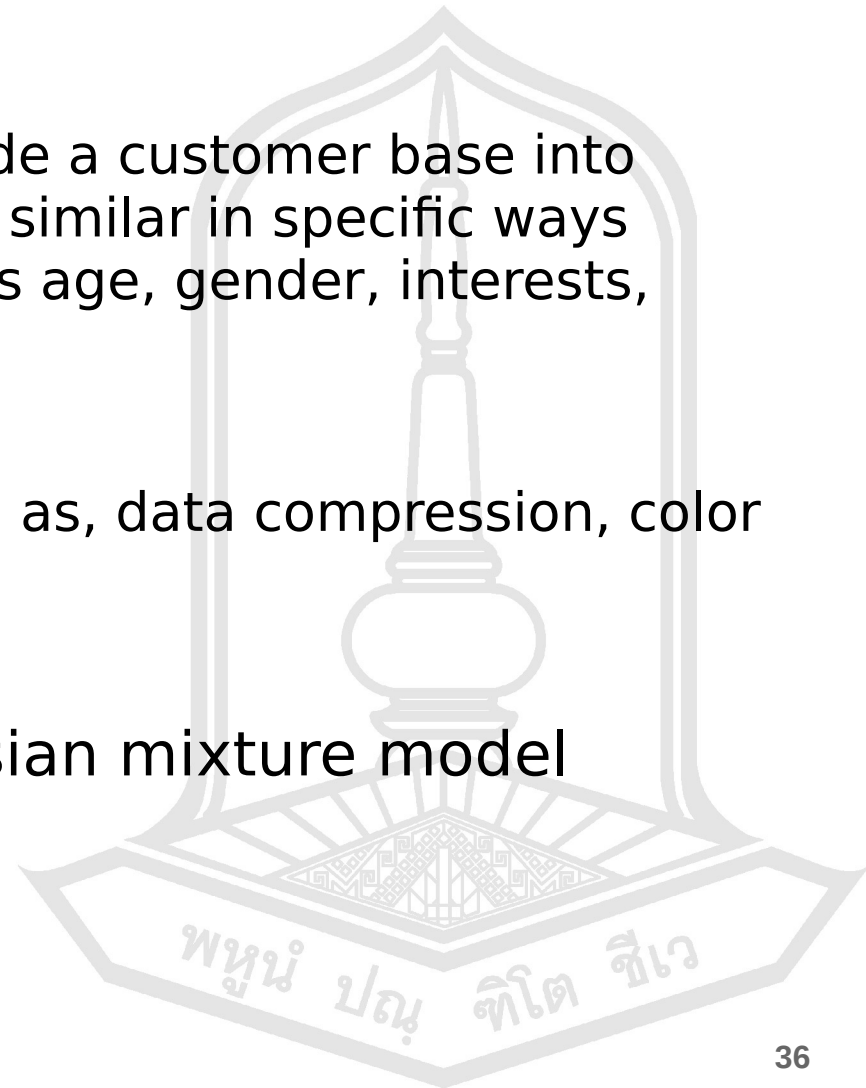
Classification

- 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.



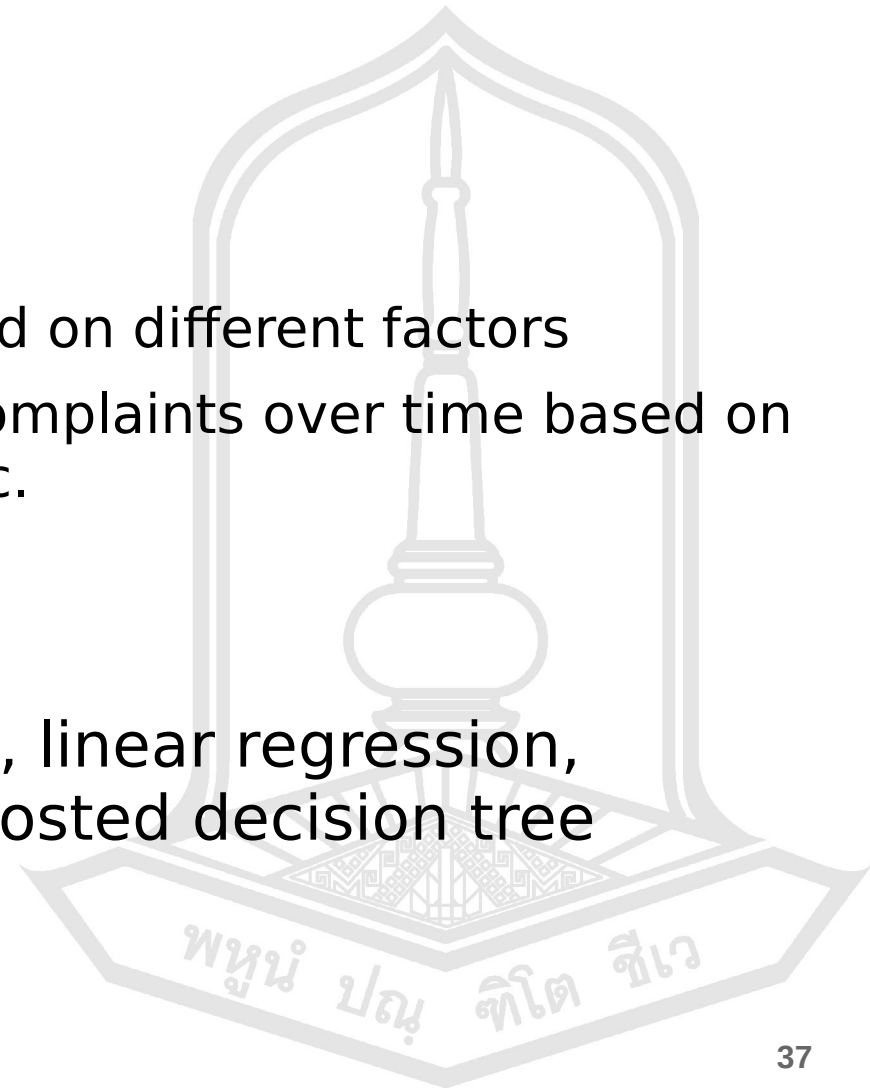
Clustering

- 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



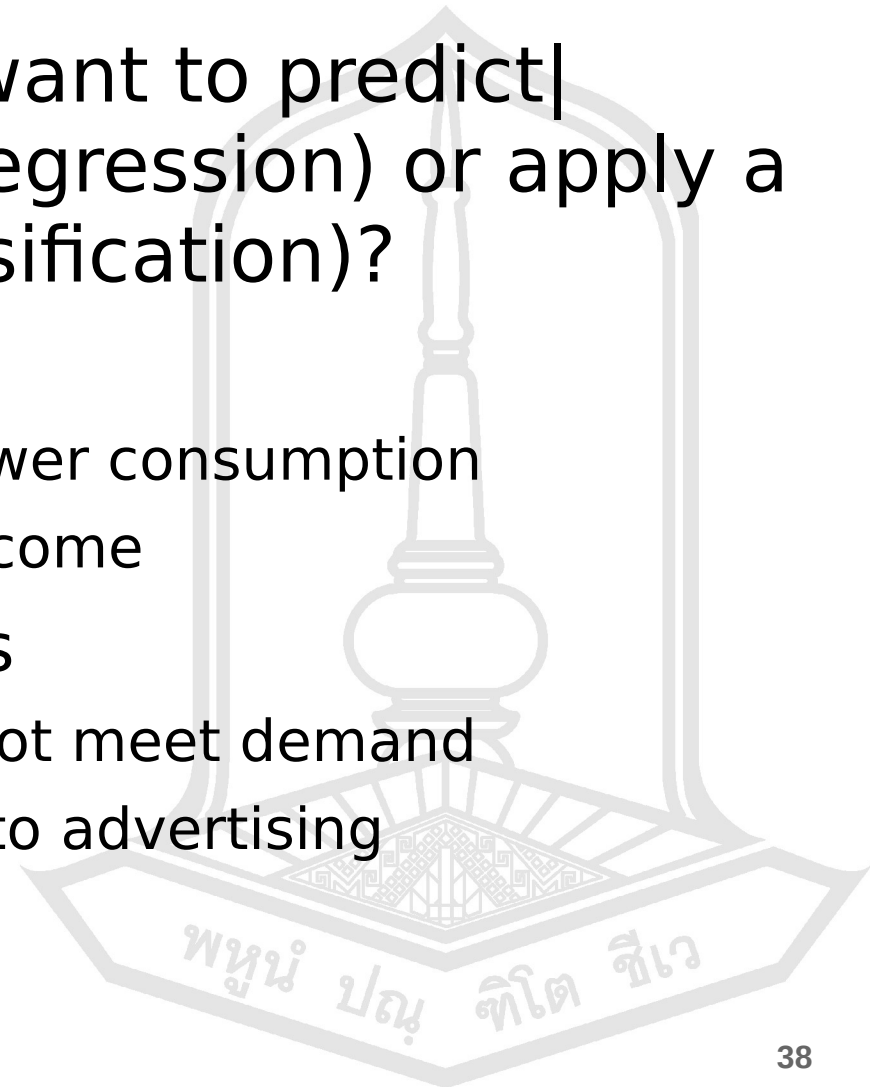
Regression

- 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



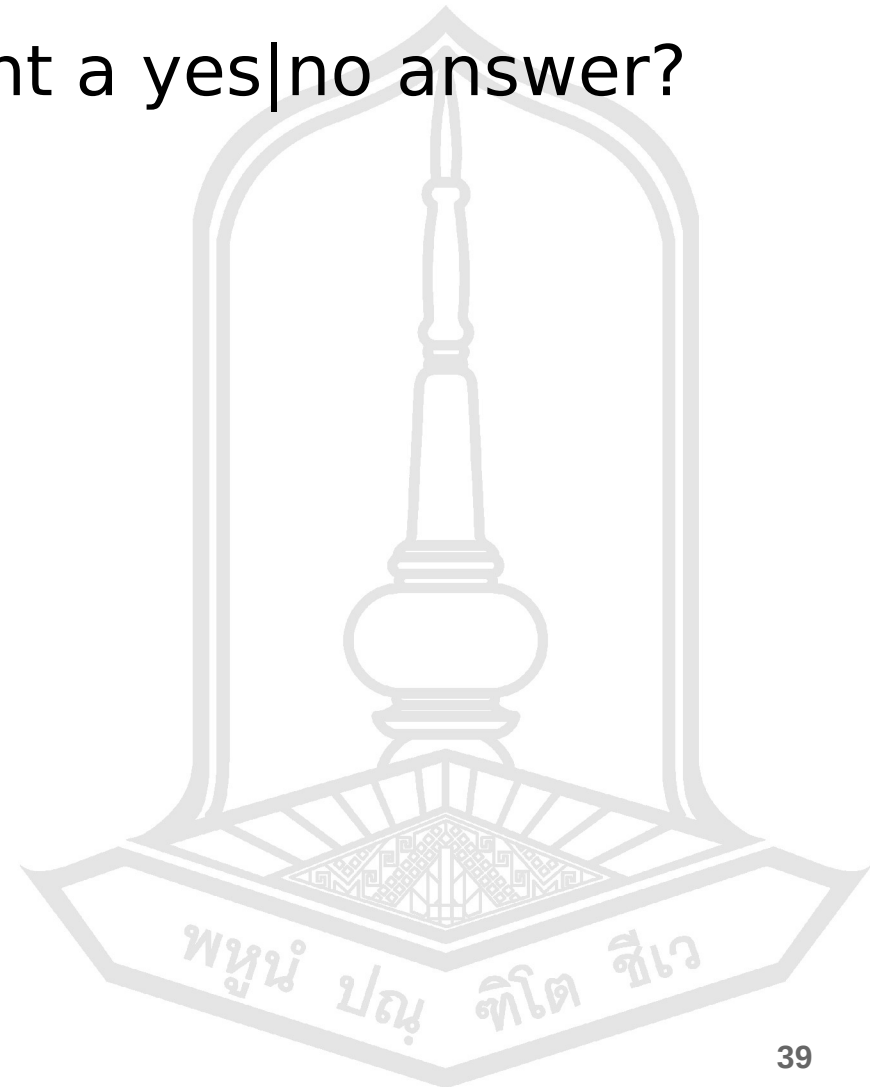
Regression vs Classification

- 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



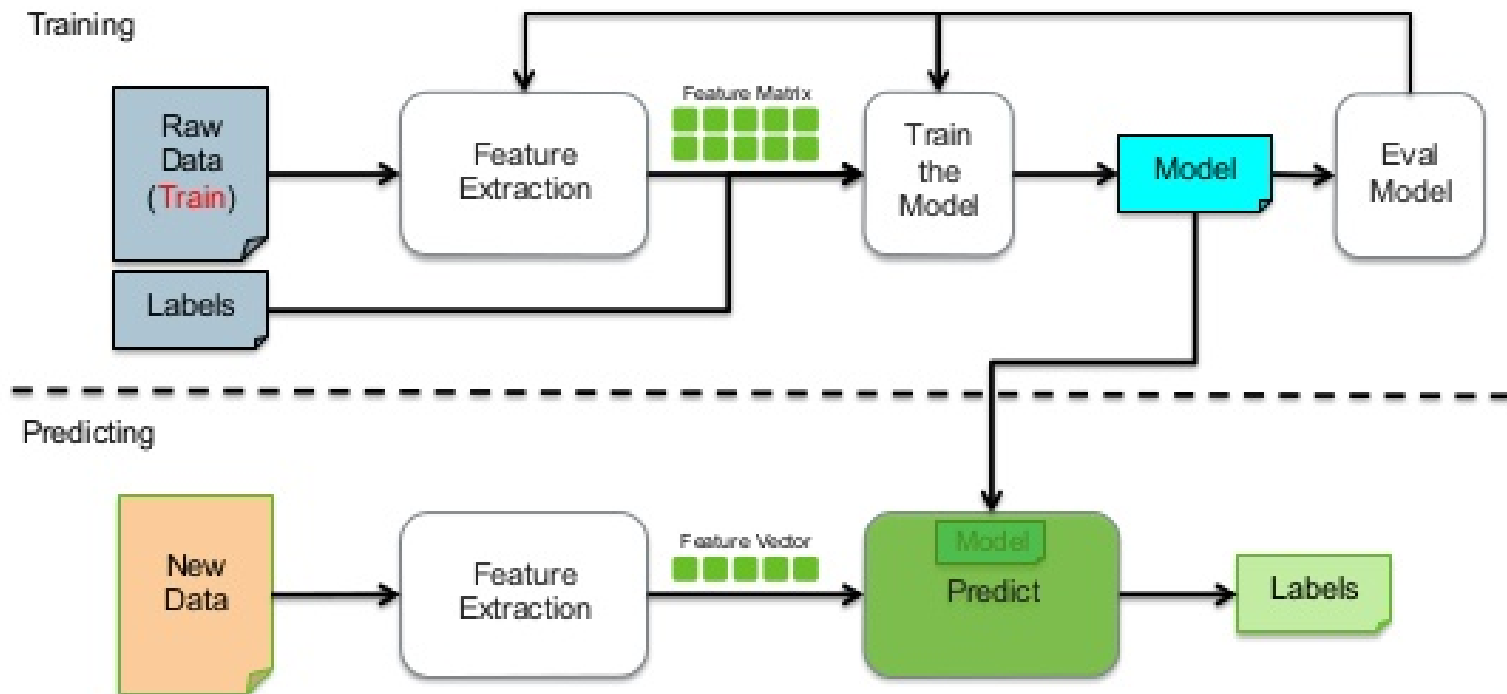
Binary vs Multiclass classification

- 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

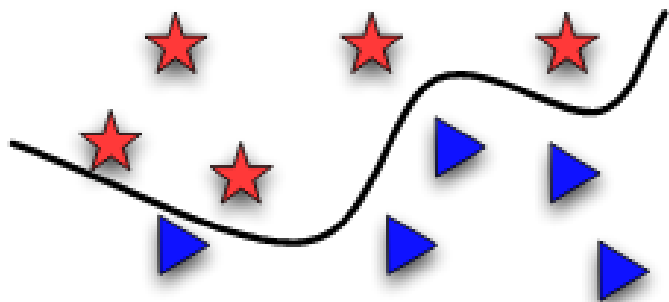


Supervised learning

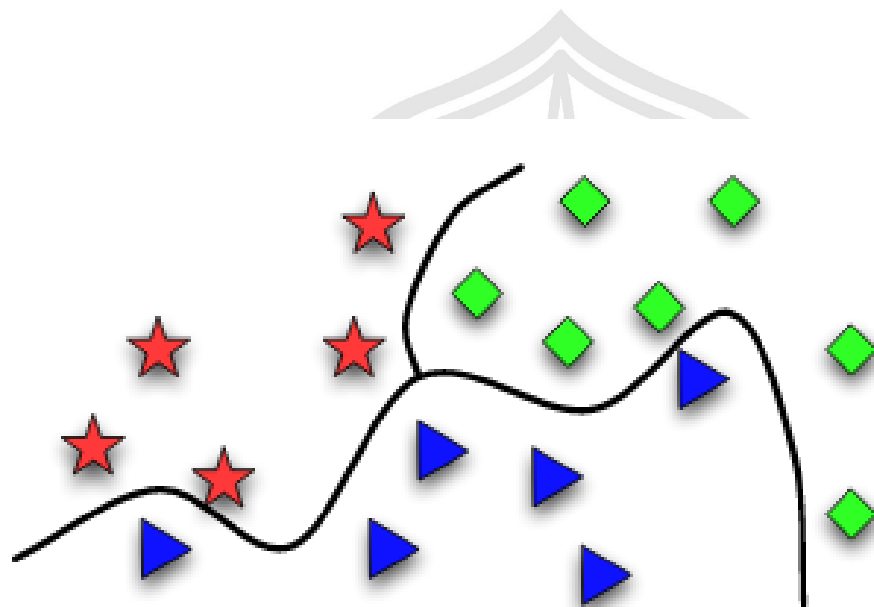
Supervised Learning Workflow



Classification problems

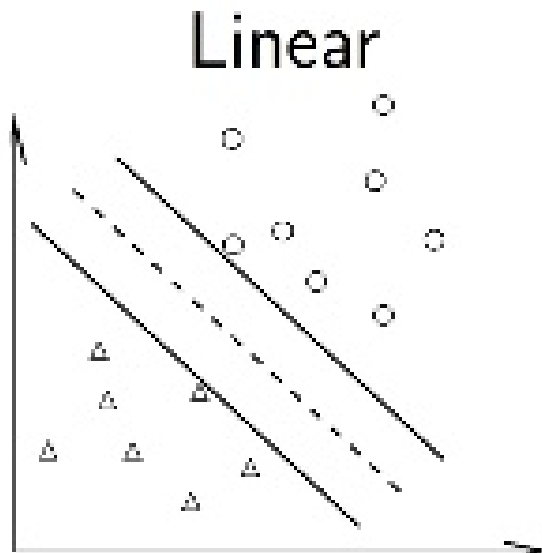


Binary classification

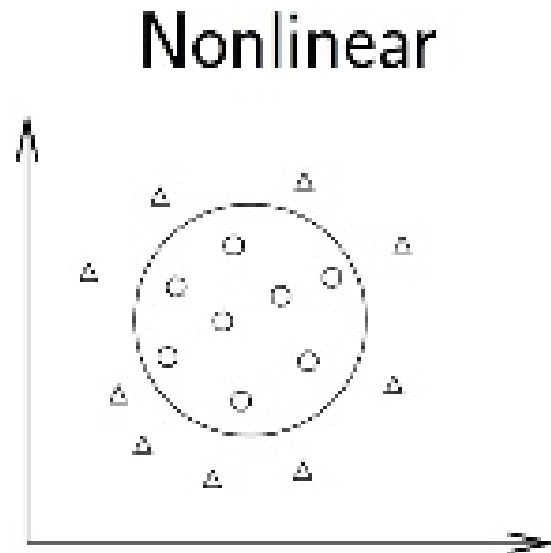


3-class classification

Linear / Nonlinear Classification

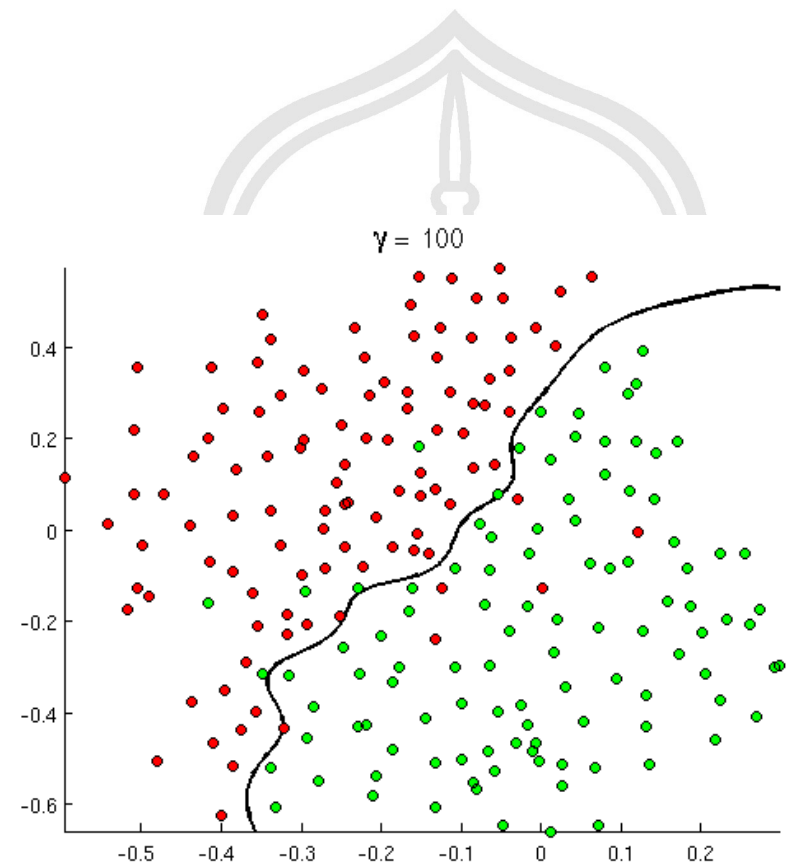
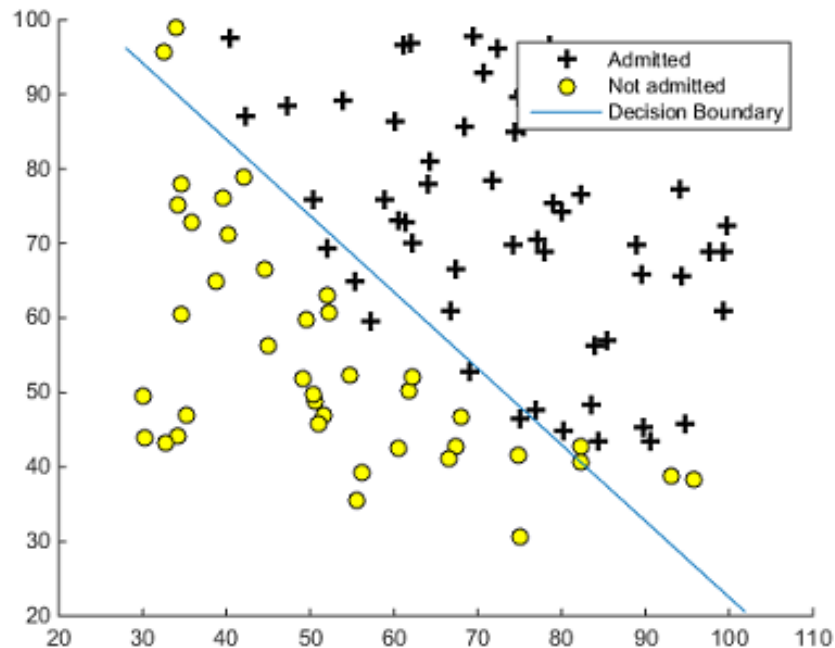


Linear – In the data's original input space, labels can be classified by a linear decision boundary.



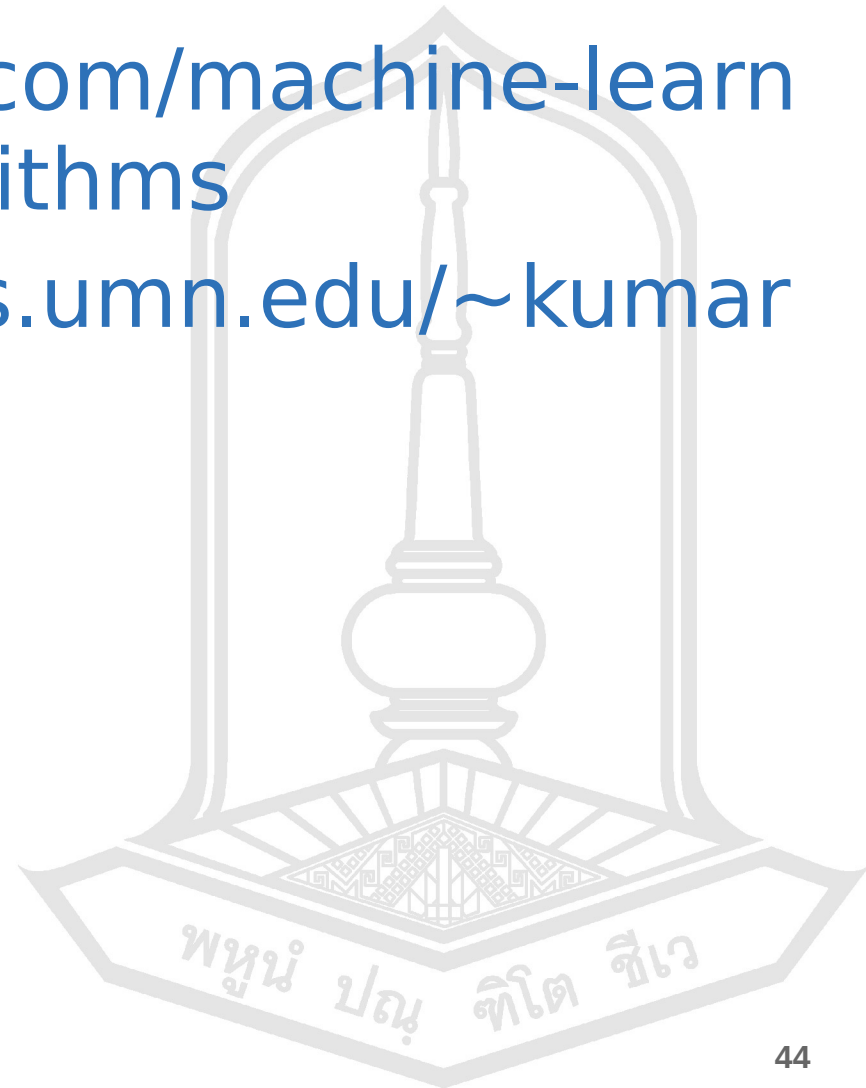
Nonlinear – The classifiers have nonlinear, and possibly discontinuous decision boundaries.

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>

