Practical AI

Model Educator

Charles Chang 張佳彥

TREND MICRO™

Securing Your Journey to the Cloud

創新 和 創新的擴散

- 創新 (Research)
  - 0 to1
    - Neural Network (1943)
    - Convolutional Neural Network (1968)
    - Capsule Network (2017)

- 創新的擴散 (Engineering)
  - 1 to100
    - AlphaGo
    - 自駕車

1. Use AI as a tool, everyone could use AI.
2. AI系統像員工一樣，必須訓練！By 李宏毅

WHO AM I – Model Educator

- Join Trend Micro on 2009
  - Infra Developer
  - Threat Researcher
  - Machine Learning Researcher
- Join XGen ML project on 2015
- Now leading the Machine Learning Research/Operation team of XGen

Agenda

- Why we need model ?
- What is model ?
- How to make the model do its best ?
  - What is your problem ?
  - What is model's problem ?
  - How to educate your model ?
- How to run ML project ?
- Final Exam

Why we need model ?

- Goal:
  - A **systematic** way to create **rules/logic** by **data** to get the optimized performance
  - ➢ **ML Algorithm**

Management
& Cost

Why we need model ?

- Stage1: drop the e-mail if contains any SPAM WORDS.
  - Hi, you can buy the cheapest iPhone here.
  - Hi Charles, would you please buy a iPhone and bring back to me?

Why we need model ?

- Stage2: make more complicated rules
  - if contains any SPAM WORDS.
    - If not contains name.
      - If contains more than 3 SPAM words.
        - » If not contains signature
          - » If contains phone-number
            - » If not contains URL
              - » If….
                - » If…..

Why we need model ?

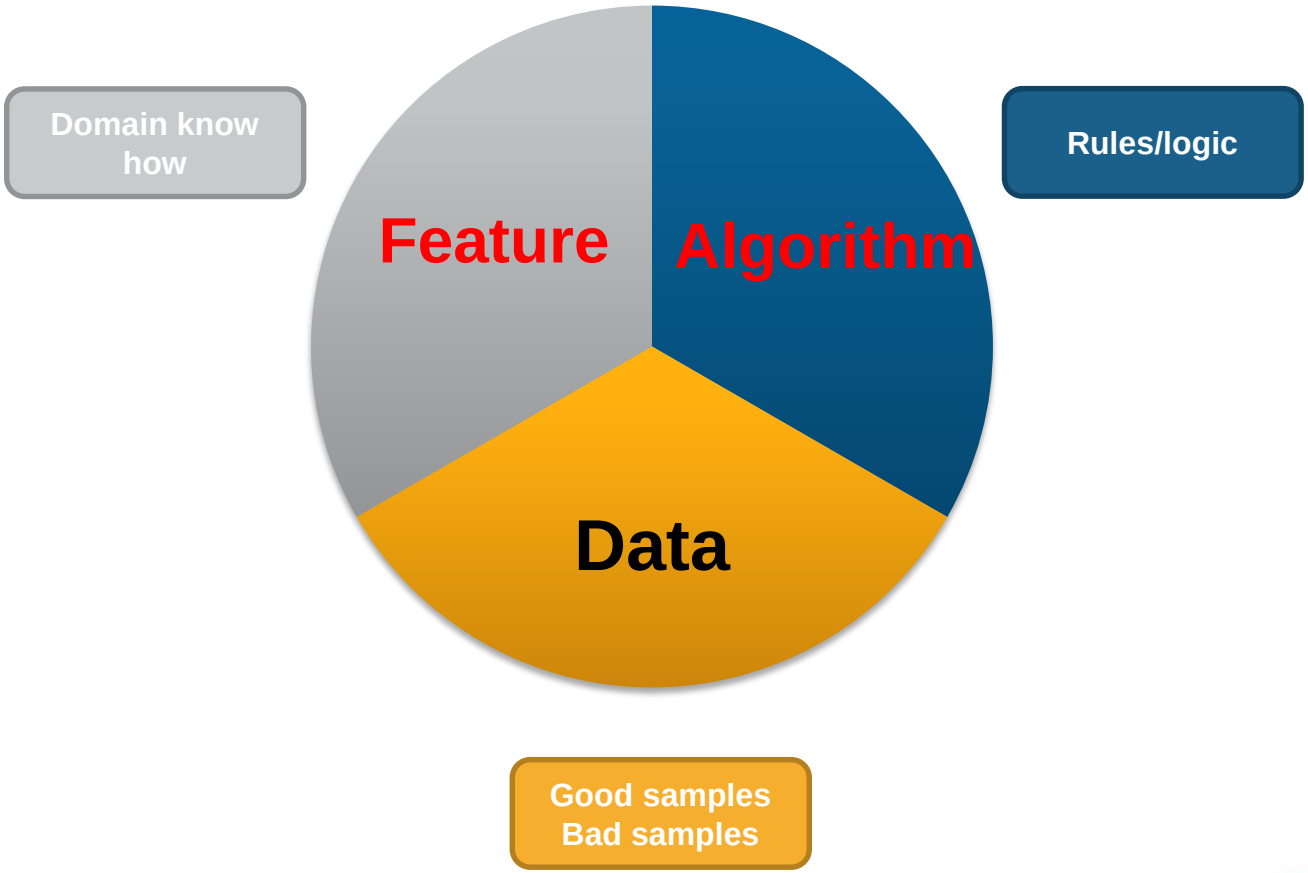
- Example:
  - Training
    - Spam:
      - Hi, you can buy the cheapest iPhone here.
      - ……
    - Benign:
      - Please help to take care of my son, thank you.
      - ……

Words	Score
buy	-0.2
cheapest	-0.6
help	0.4
thank you	0.1
…	…

  - Testing
    - Hi Charles, would you please help to buy an iPhone and bring back to me? Thank you.
      - 0.4-0.2+0.1 = **0.3** (Benign)
    - Hi, do you need any help? You can always buy the cheapest stuff here. Thank you.
      - 0.4-0.2-0.6+0.1 = **-0.3** (Spam)

Management & Cost

What is model ?



What is model not ?

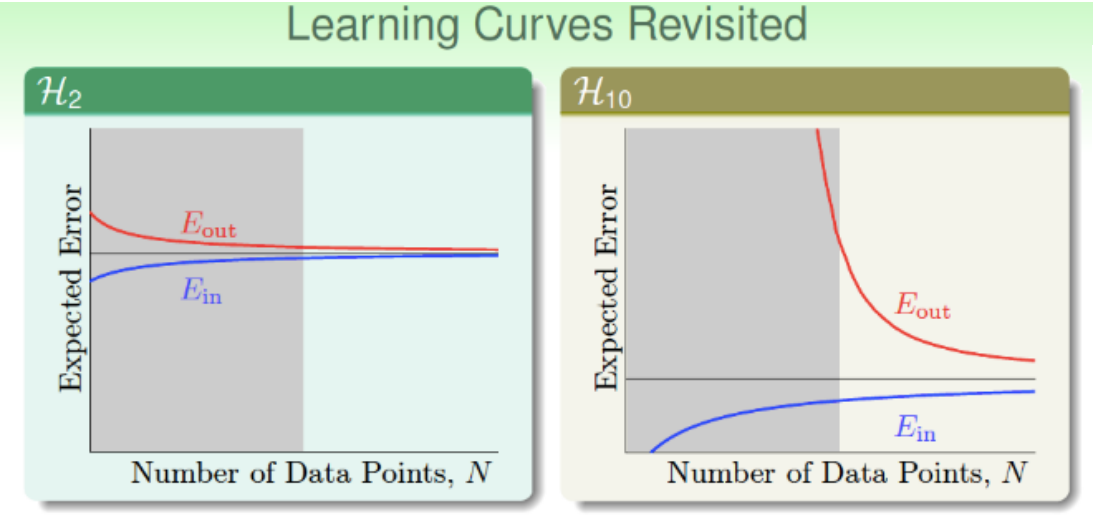
TREND MICRO

TREND MICRO
Securing Your Journey to the Cloud

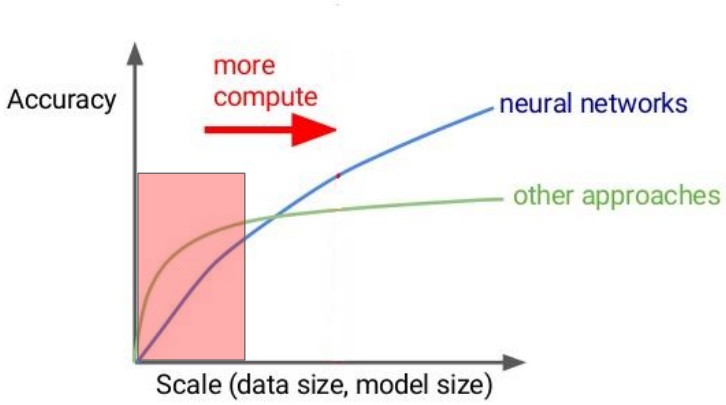
擺對位置，你才能發揮所長，放大自己的價值

What is your problem ?

- Simple V.S. Complex
- Accuracy V.S. Explanation
- **Variant V.S. Invariant**
- **Cost function**



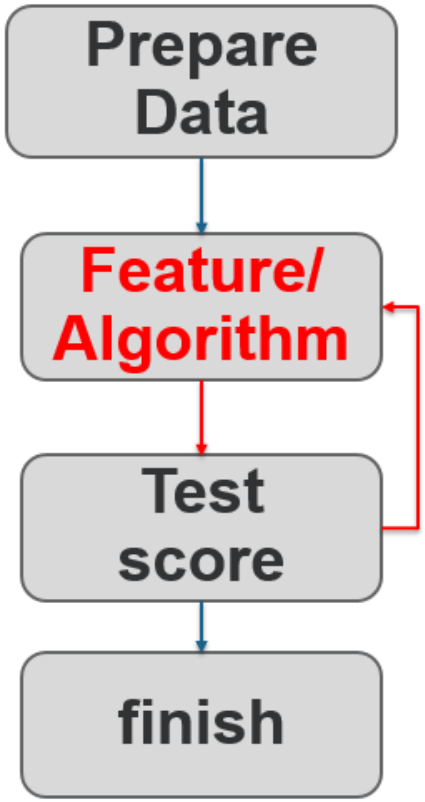
https://zh-tw.coursera.org/learn/ntumlone-mathematicalfoundations



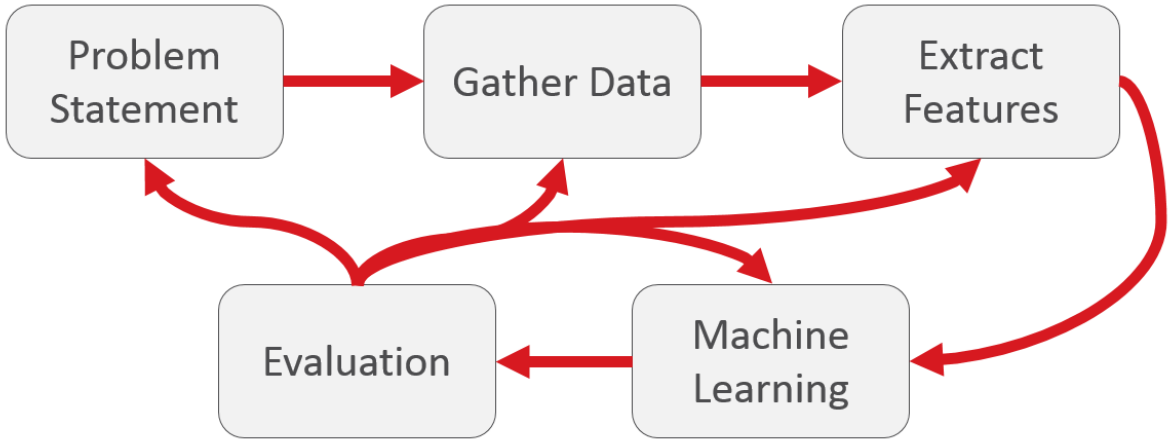
https://www.slideshare.net/AIFrontiers/jeff-dean-trends-and-developments-in-deep-learning-research

What is your problem ?

- Invariant V.S. Variant



1-time cost



Operation cost

What is your problem ?

- Invariant V.S. Variant
  - Invariant
    - 手寫辨識
    - 語音辨識
    - 圍棋

> 750 Million NTD

  - Variant
    - 金融預測
      - Daily update
    - 病毒預測
      - Monthly update

What is model's problem ?

- How do you know if a student is leaning well or not ?
  - Testing
    - Testing with many different kinds of data and meta data
      - 歷史
        - » 中國史，西洋史，近代史，藝術史
      - 銷售預測
        - » 周間，周末，上班，下班
      - 病毒預測
        - » 勒索病毒，木馬，蠕蟲，廣告

Cost function
Cost function
Cost function

What is model's problem ?

- How do you know if a student is leaning well or not ?
  - Asking why
    - Linear model
    - Review training data
      - Leave 1 data out model
      - Influence Functions
    - Class Activation Mapping
    - Attention

https://github.com/jacobgil/pytorch-grad-cam

Example

- Business Email Compromise

CEO
Immediate Wire Transfer
收件人: **Chief Financial Office**

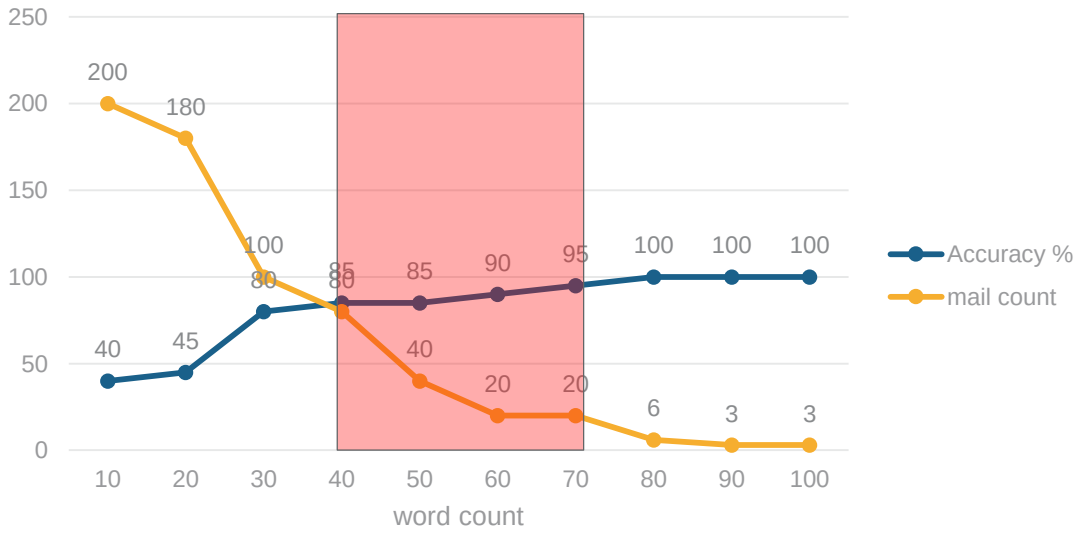
2015年2月3日 上午8:09

C

Please process a wire transfer payment in the amount of $250,000 and code to “admin expenses” by COB today.

Wiring instructions below…

書寫風格異常!!



Accuracy: 60.1%

Accuracy: 86.9%

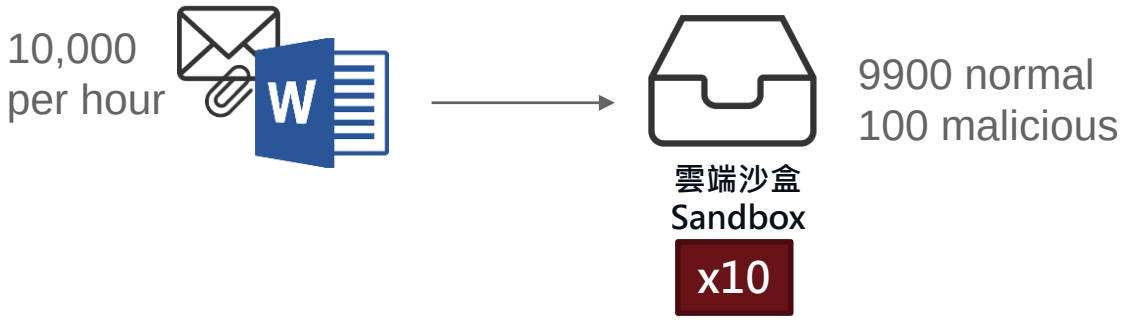
How to educate your model ?

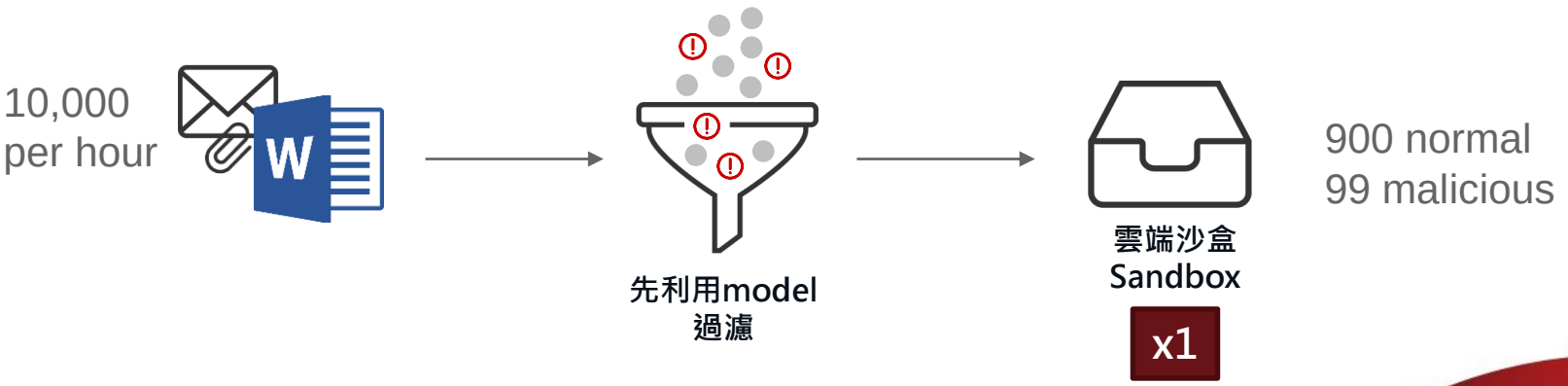
- How to make a student good ?
  - Data is the King
  - Teacher need to **know** and **prepare** the data.
    - This is the major work of the model educator

- How to make a good student ?
  - Feature and Algorithm
  - You teach 1,000 different students, and choose a best.

How to educate your model ?

- ~~Data is the King~~
- Data is the Queen, Label is the King.

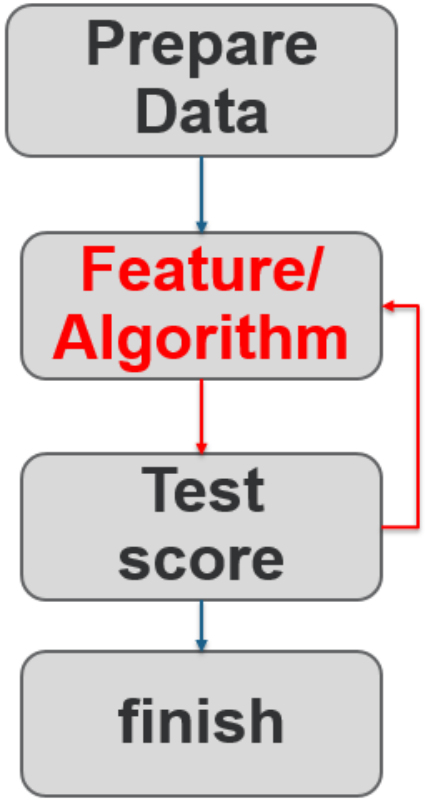




How to run ML project ?

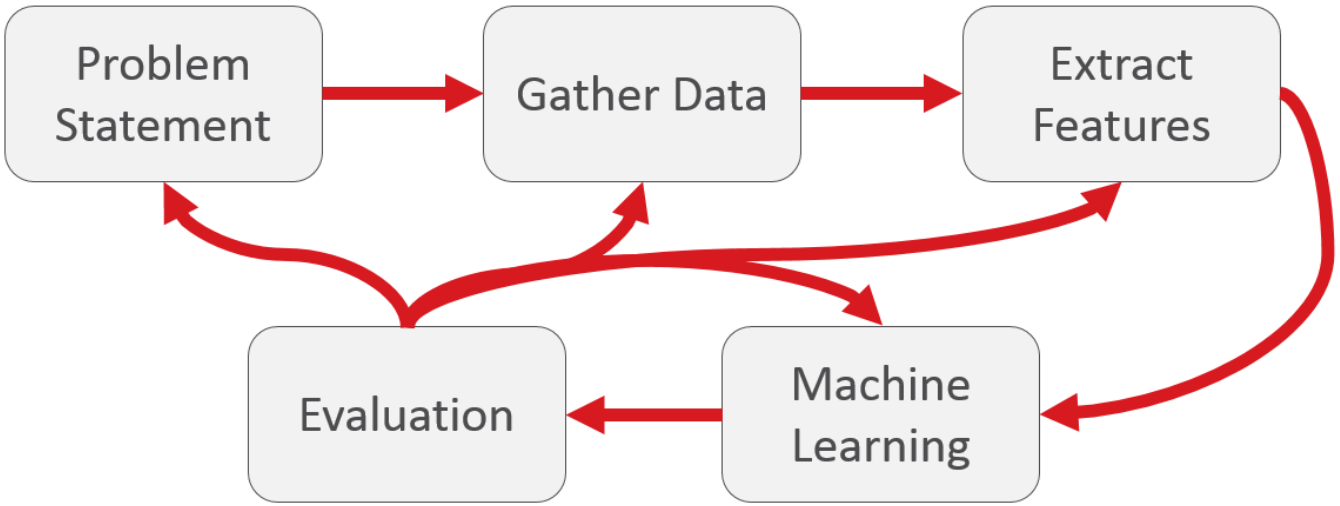
How to run ML project ?

- Competition
  - Too narrow



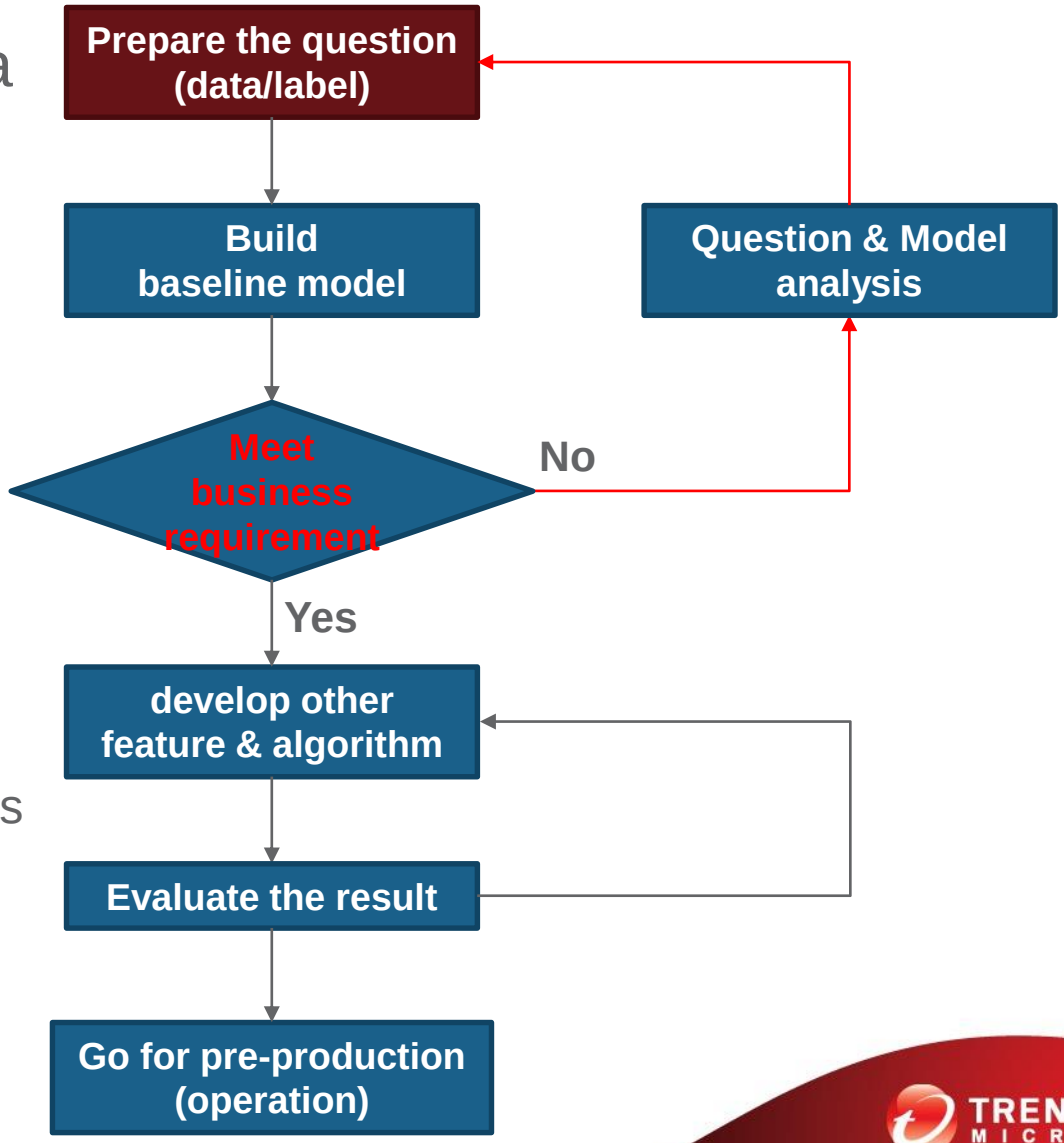
How to run ML project ?

- Real world
  - Too broad



How to run ML project ?

- P0: Prepare the data
- P1: Fast
  - Tuning your problem
  - Go for pre-production
- P2: Widely
  - Evaluate feature and algorithm
    - Use simple features
    - Try existed algorithms



Example

- Business Email Compromise

CEO
Immediate Wire Transfer
收件人: **Chief Financial Office**

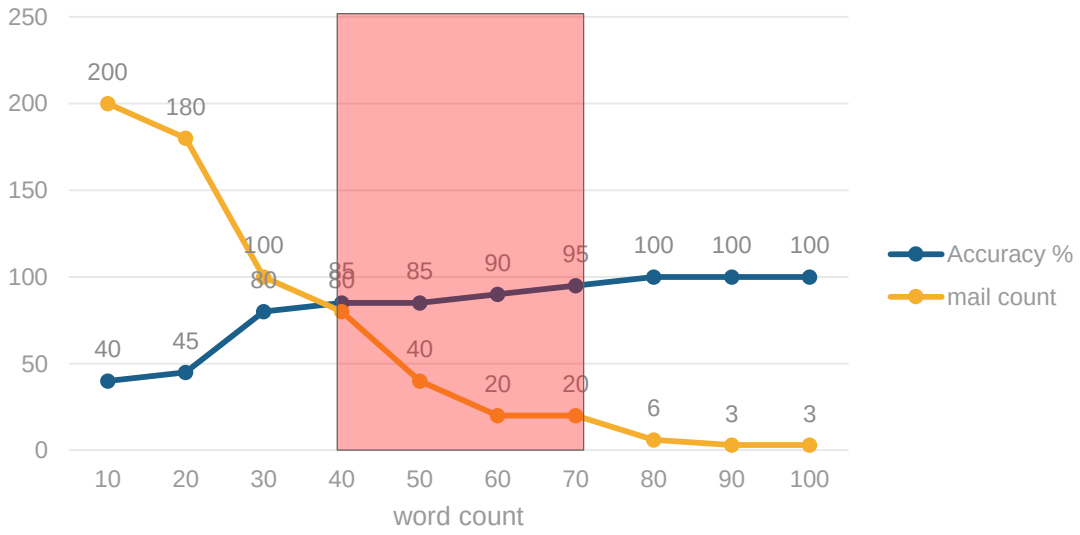
2015年2月3日 上午8:09

C

Please process a wire transfer payment in the amount of $250,000 and code to “admin expenses” by COB today.

Wiring instructions below…

書寫風格異常!!



Accuracy: 60.1%

Accuracy: 86.9%

How to run ML project ?

- Concept
  - Model is not perfect, it is just a tool.
  - It can help if asking correct question.
    - Low hanging fruit
  - Go for pre-production ASAP
  - Always think about cost.
    - Research outcome may not be predictable like product feature.
    - Operation cost may be the major cost

How to run ML project ?

- Process
  - Initial stage
    - Identify the problem(data) first
    - Evaluate all of the known algorithm fast
    - Consider operation process and cost
  - Operation stage
    - Data driven decision
    - Keep enhancing
    - Change the problem if necessary
    - Dedicated research resource with real world data

Accuracy doesn't matter. Only business value does.