ML Process

1. Prepare & Transform Data
2. Modeling
3. Evaluating
4. Backtesting
5. Stress Test
6. Deploy Strategy / Real Time Event Handling

ML 101

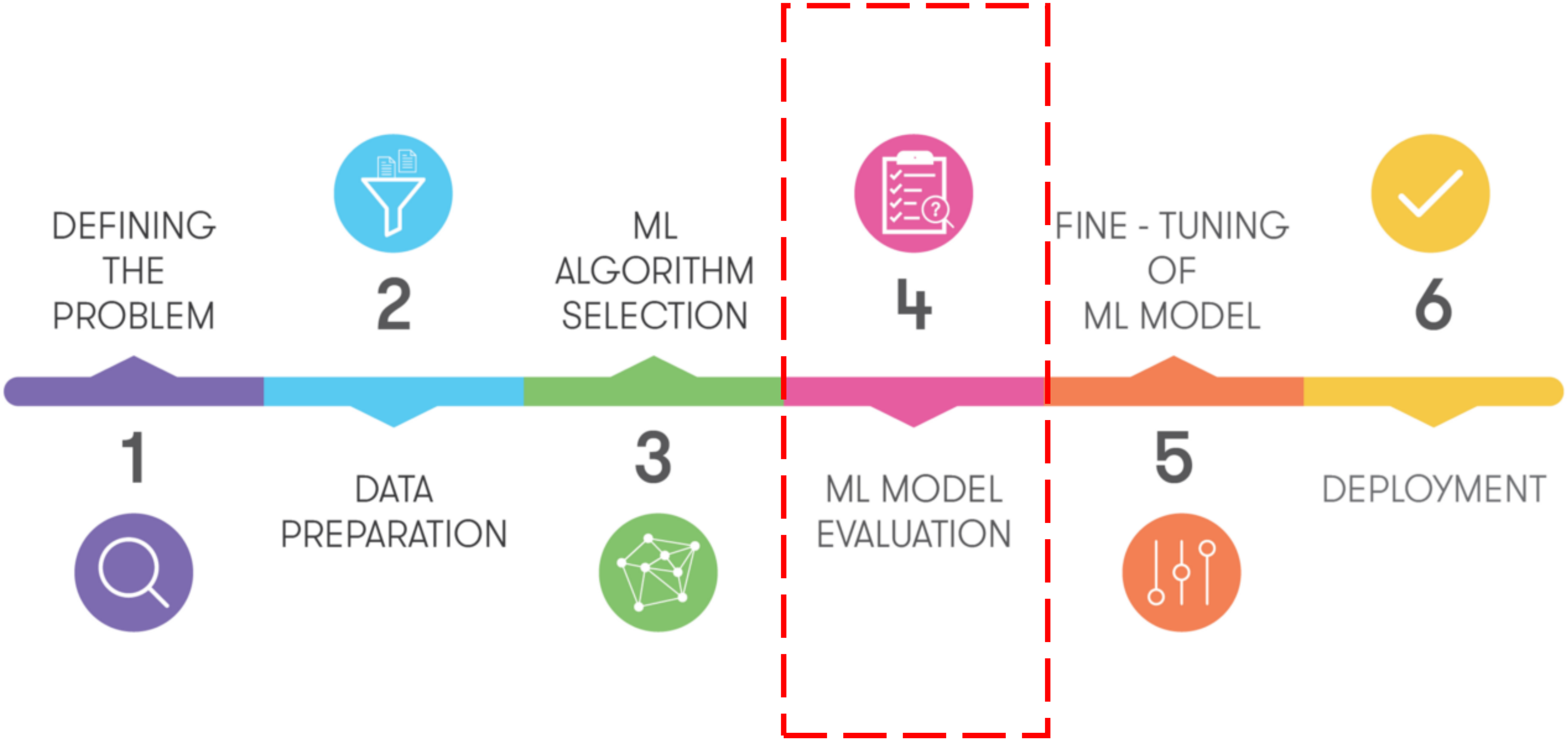
ถ้าใช้ข้อมูลที่ไม่ได้ RC

Part of 4. (เหตุการณ์crisis)

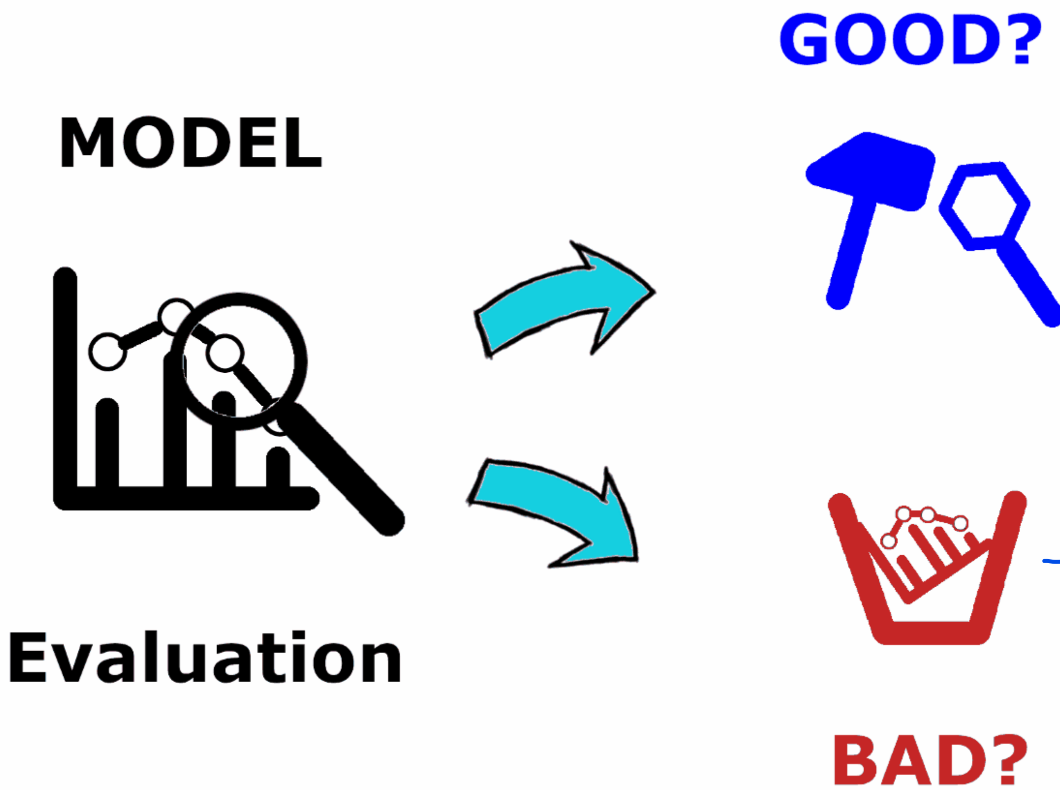
σ, Σ, scale

, normalize

Typical Modeling



How do you know



→ เอาไปใช้ ด้านไหนก็รอด

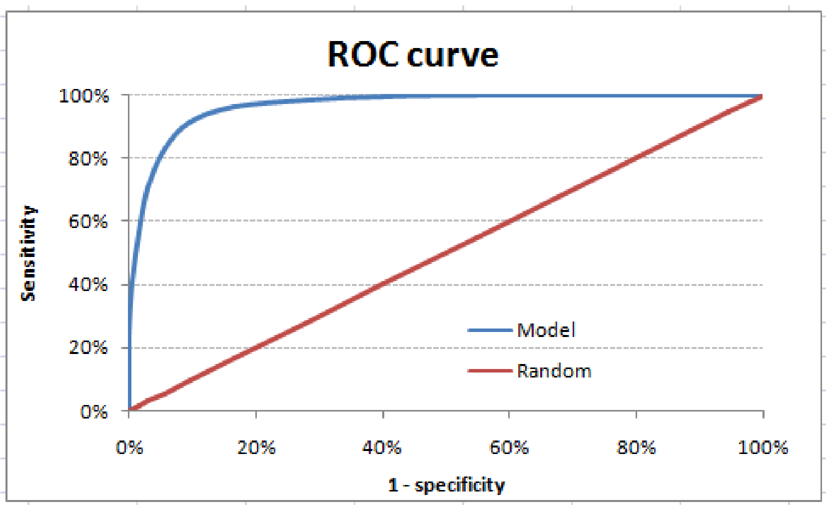
→ ผิดยังไง ช่วงไหน
- overfit

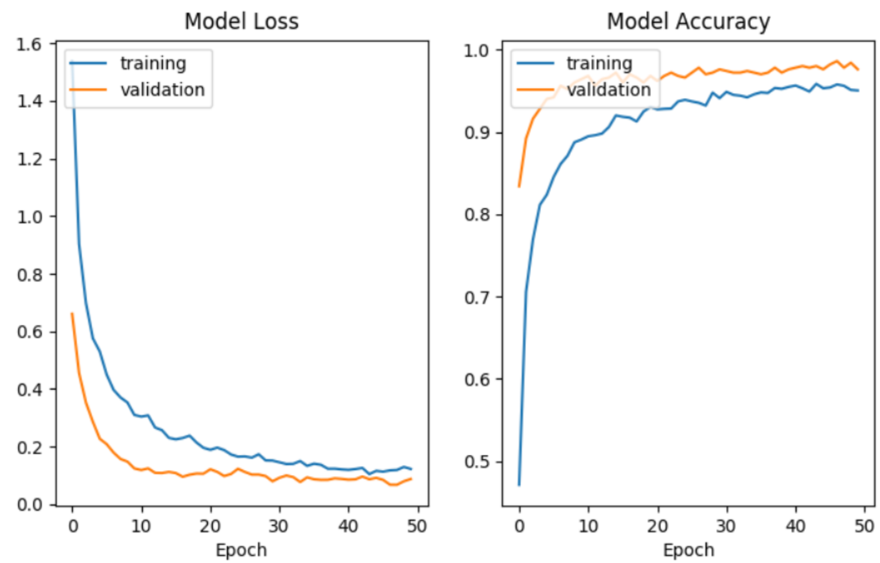
R.Brilenkov

INVESTIC

Evaluation

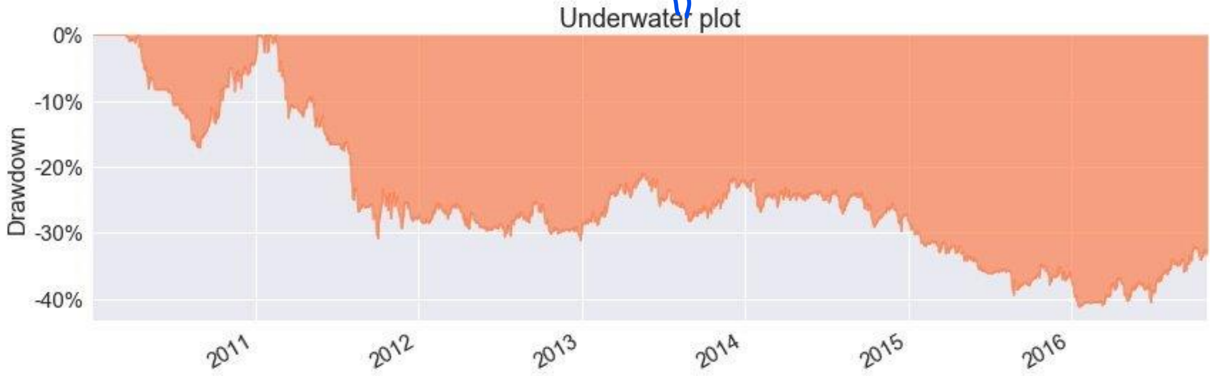
model ทั่วไป Data sci

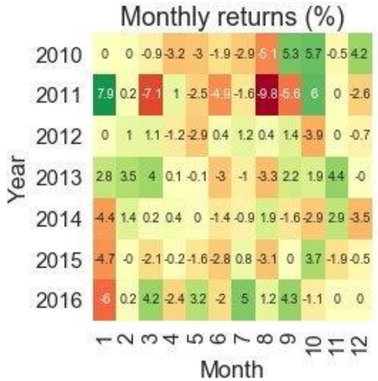


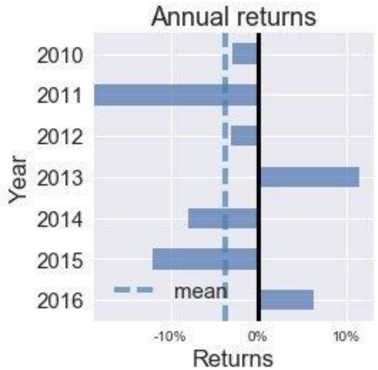


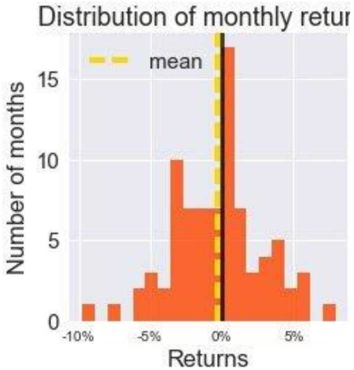
Evaluation Metric for Portfolio Management

ไมเอา รหวางทุน









INVESTIC

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

ใช้ดูว่าผิดตรงไหน / ถูกต้อง
— ผิดพลาด

Confusion Matrix

	Actually Positive (1) +	Actually Negative (0) -
Predicted Positive (1) +	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0) -	False Negatives (FNs)	True Negatives (TNs)

INVESTIC

Regression

→ ได้เป็นตัวเลข | classification ได้มาเป็น 1 2 3 label แยกกลุ่ม ค่าไรแต่ละกลุ่ม

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

ใช้ 2 อย่าง ก็พอแล้ว วัดว่า R-squared ก็ได้

Root Mean Squared Error (RMSE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Root Mean Squared Log Error (RMSLE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(p_i + 1) - \log(a_i + 1))^2}$$

prediction

actual

$$R^2 = 1 - \frac{\text{MSE(model)}}{\text{MSE(baseline)}}$$

$$\frac{\text{MSE(model)}}{\text{MSE(baseline)}}$$

$$\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(\overline{y}_i - \hat{y}_i)^2}$$

Regression

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

Root เพื่อขยาย Bias

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}}$$

$$R^2 = 1 - \frac{\text{MSE(model)}}{\text{MSE(baseline)}}$$

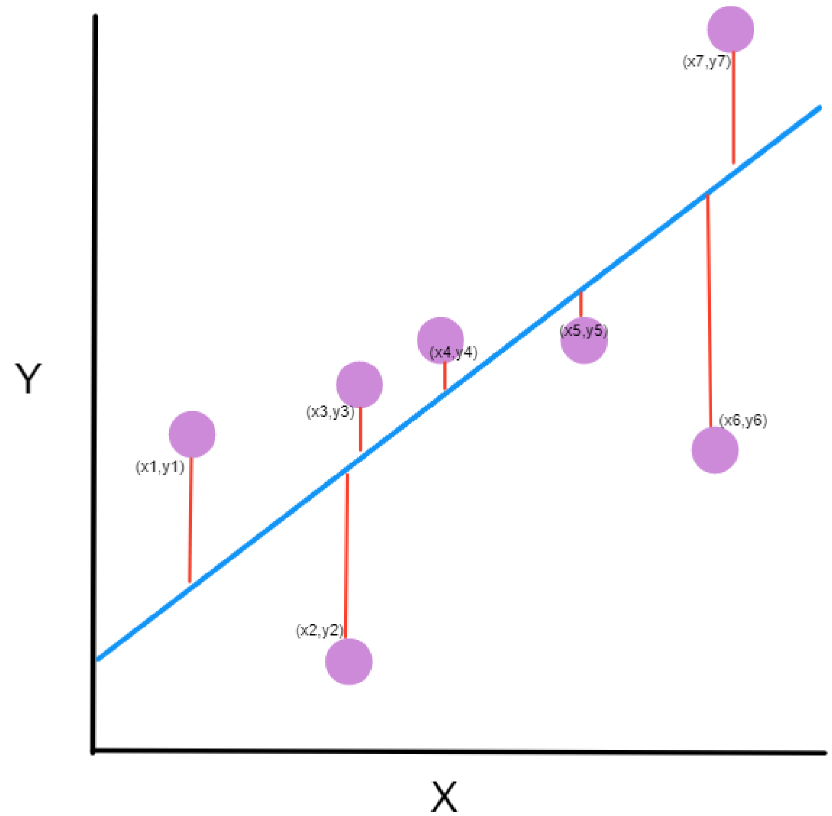
$$\frac{\text{MSE(model)}}{\text{MSE(baseline)}}$$

$$\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(\overline{y}_i - \hat{y}_i)^2}$$

Regression

ตัด Outlier ว่าถ้าดูมาก็ไม่เอานั้

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$



measure of how close a fitted line is to data points

Sum Square → Cancelling Positive/Negative errors

Regression

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

ขยาย error ให้ชัดขึ้น

‘square root’ → show large number deviations
punishes large errors.

Sum Square → Cancelling Positive/Negative errors

RMSE is highly affected by outlier values

INVESTIC

Other Type??

Root Mean Squared Error (RMSE)

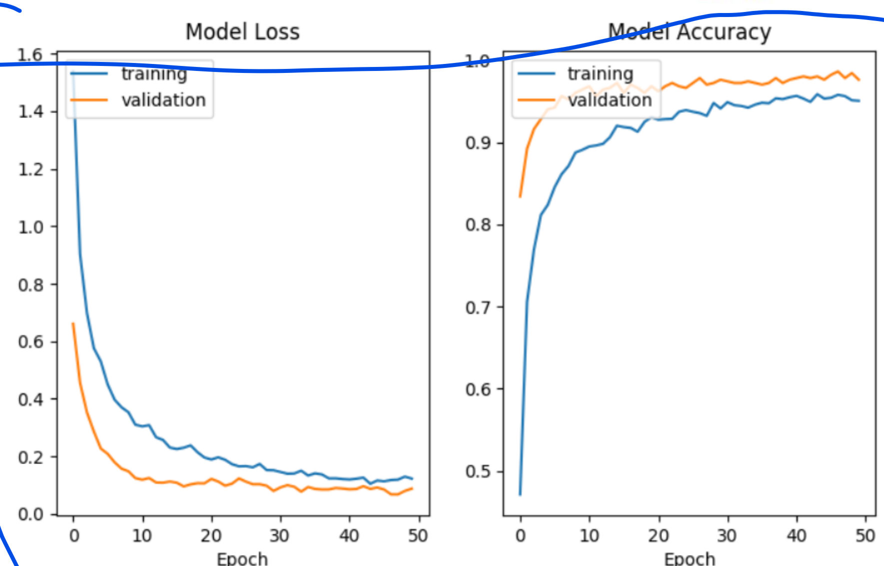
$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\boxed{y_i-\hat{y}_i}^2}$$

Root Mean Squared Log Error (RMSLE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\boxed{\log(p_i+1)-\log(a_i+1)}^2}$$

prediction

actual



Other Type??

Root Mean Squared Error (RMSE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\boxed{(y_i - \hat{y}_i)}^2}$$

Root Mean Squared Log Error (RMSLE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\boxed{(\log(p_i + 1) - \log(a_i + 1))}^2}$$

prediction

actual

1.If both predicted and actual values are small: RMSE and RMSLE are same.
2.If either predicted or the actual value is big: RMSE > RMSLE
3.If both predicted and actual values are big: RMSE > RMSLE (RMSLE becomes almost negligible)

INVESTIC

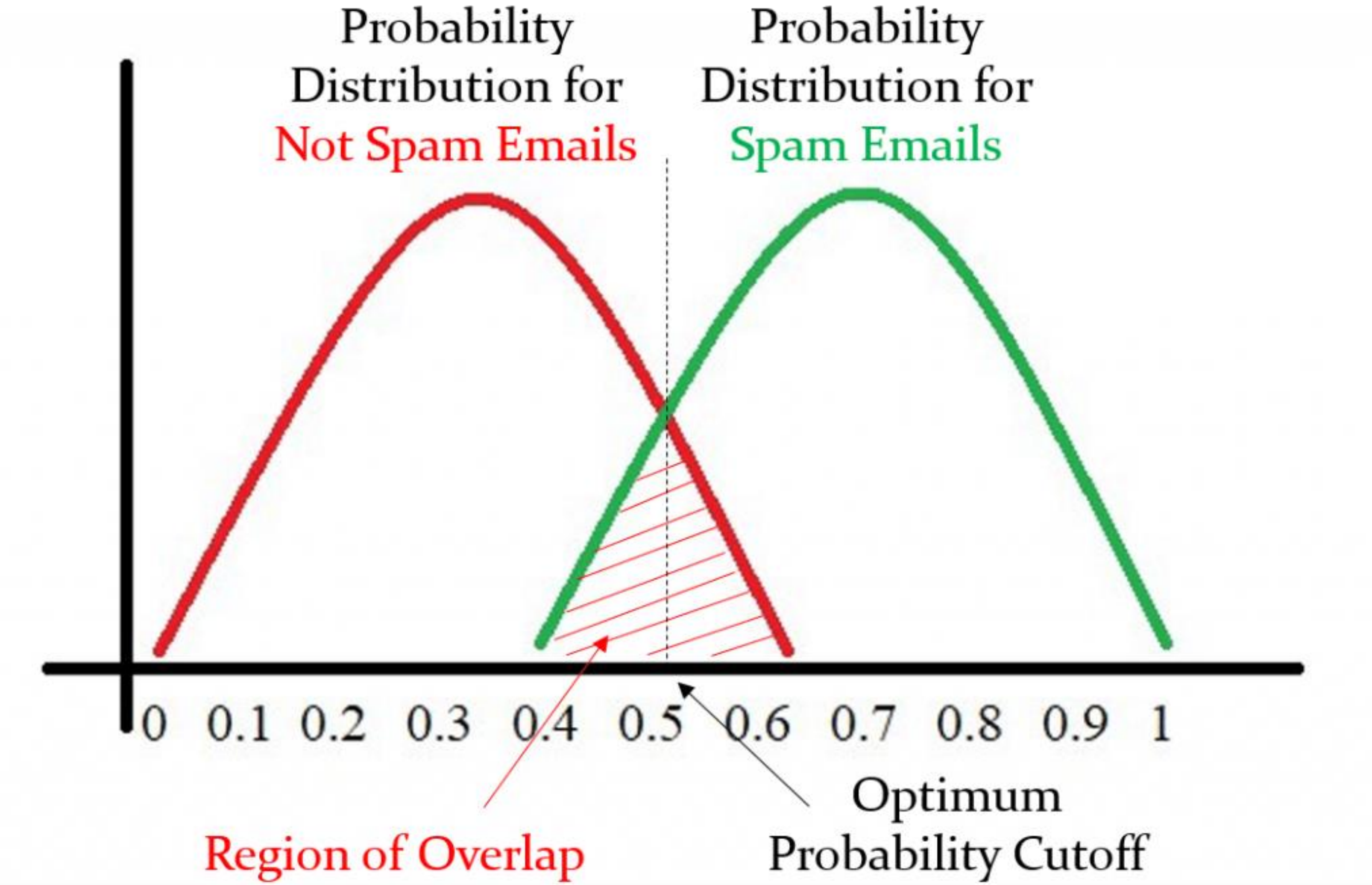
Classification

ใช้มากสุด

Confusion Matrix		**Target**			
		Positive	Negative		
Model	Positive	a	b	*Positive Predictive Value*	a/(a+b)
	Negative	c	d	*Negative Predictive Value*	d/(c+d)
		Sensitivity	*Specificity*	**Accuracy** = (a+d)/(a+b+c+d)	
		a/(a+c)	d/(b+d)		

$$F_1 = \left(\frac{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}}{2}\right)^{-1} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}.$$

Evaluating



INVESTIC

Evaluating

Type I and Type II Error

Null hypothesis is...	True	False
Rejected	**Type I error** False positive Probability = α	**Correct decision** True positive Probability = 1−β
Not rejected	**Correct decision** True negative Probability = 1−α	**Type II error** False negative Probability = β

Scribbr

Confusion Matrix

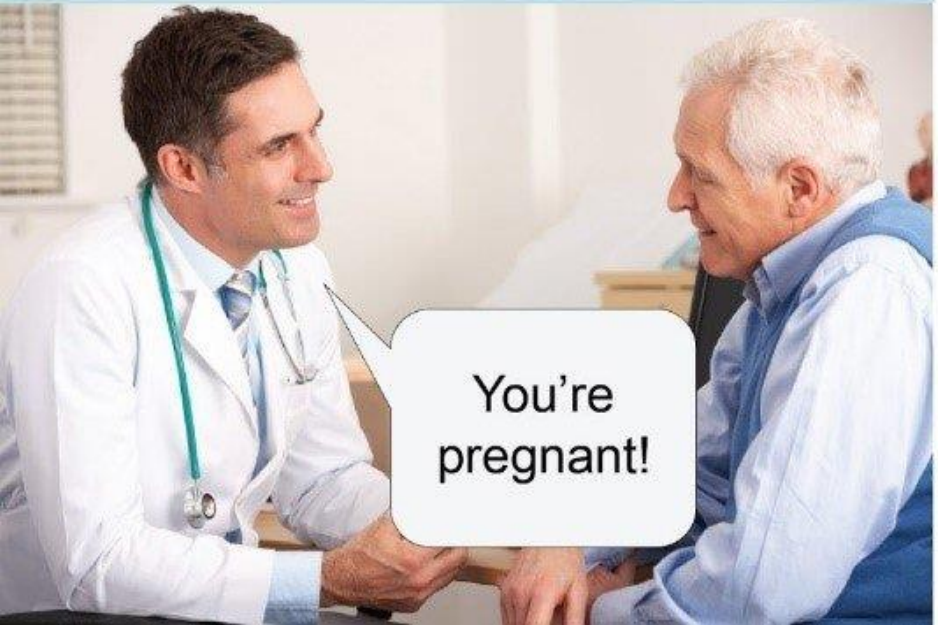
	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Evaluating

บอก ผู้ชายว่าท้อง

ผิดบวก ปกติ

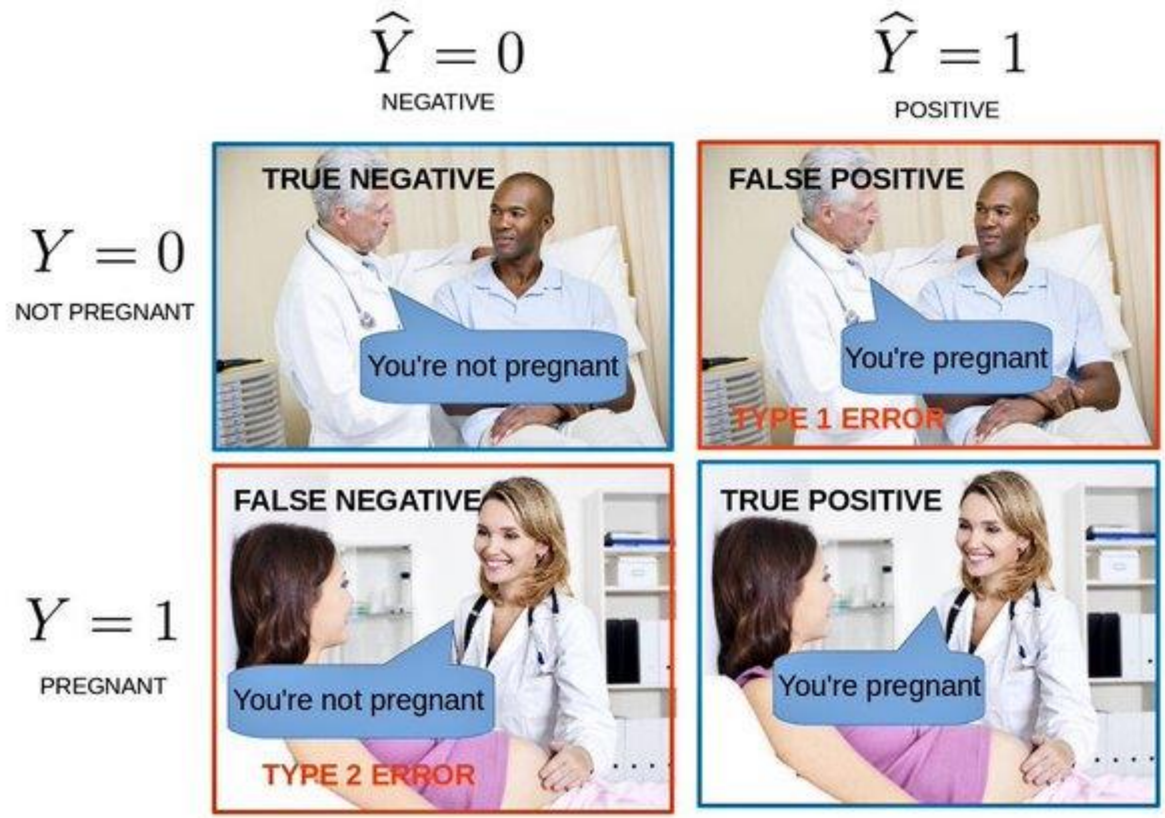
Type I Error



Type II Error

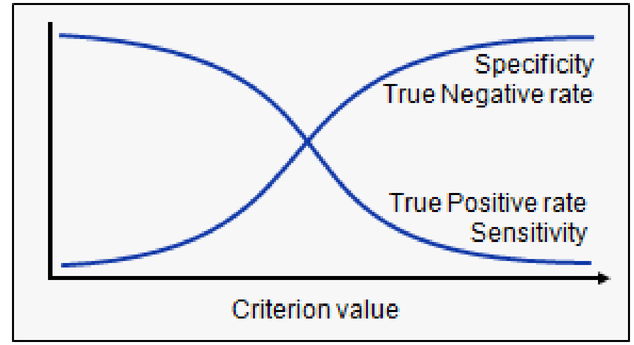
Evaluating

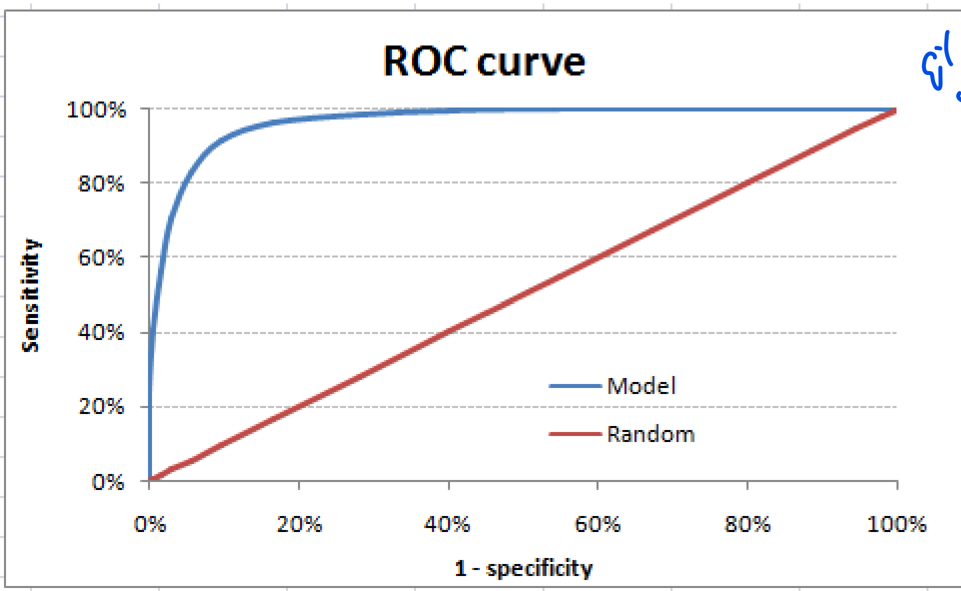
ต้นทุน ไปดู tradeoffs เท่าไร



Classification - ROC

Confusion Matrix		Target	
		Positive	Negative
Model	Positive	a	b
	Negative	c	d
		Sensitivity	*Specificity*
		a/(a+c)	d/(b+d)



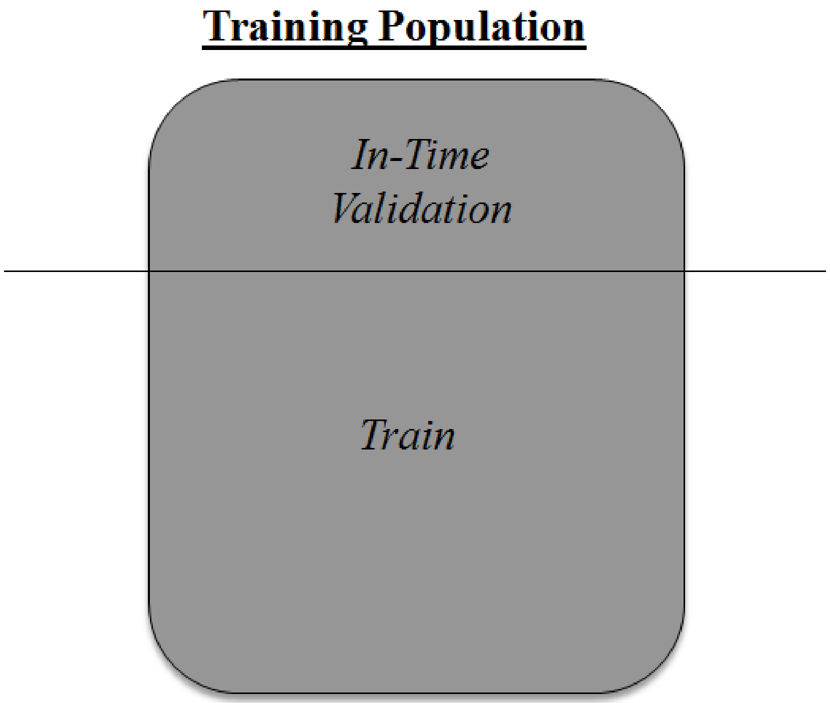


ยิ่งใหญ่ / เต็ม ยิ่งดี

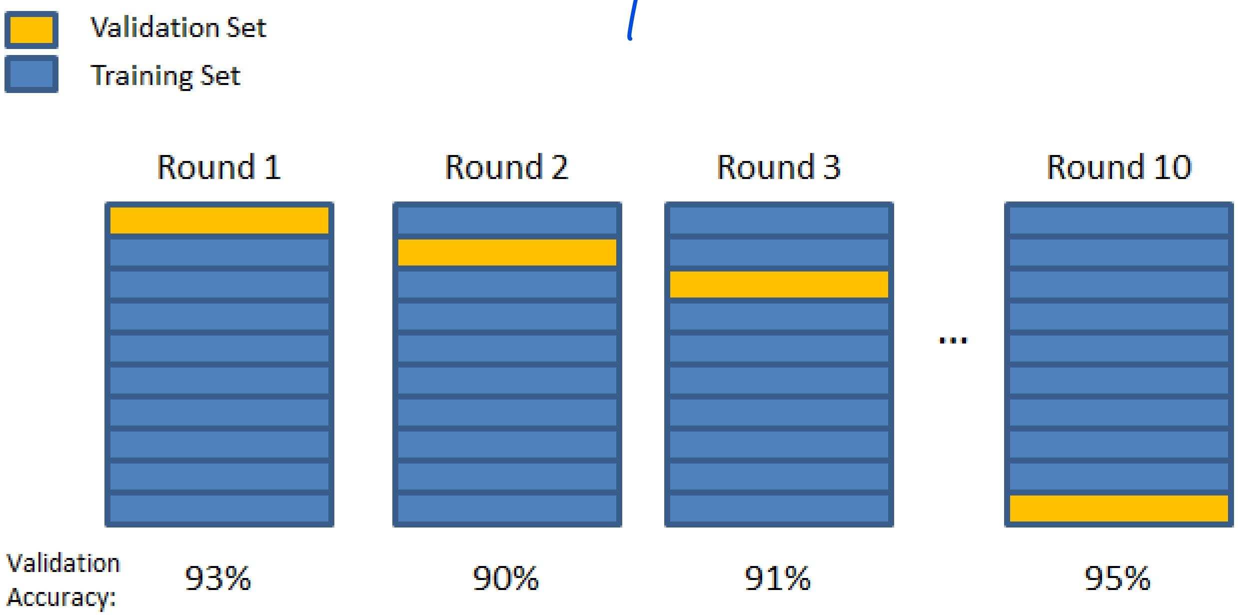
- .90-1 = excellent (A)
- .80-.90 = good (B)
- .70-.80 = fair (C)
- .60-.70 = poor (D)
- .50-.60 = fail (F)

INVESTIC

Cross Validation

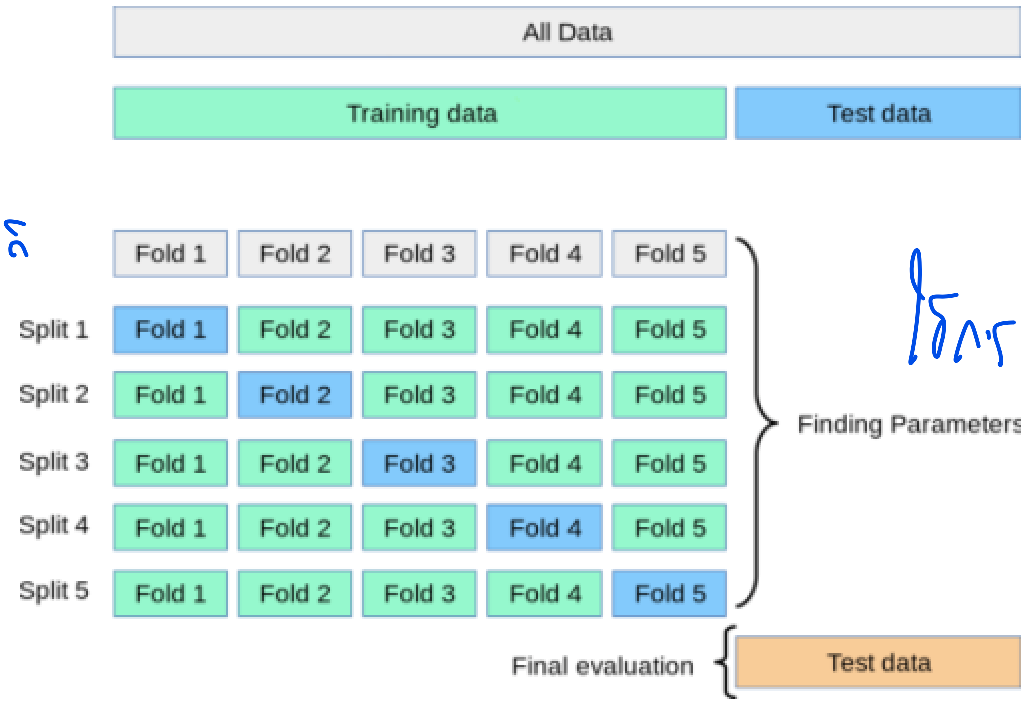


K Fold Cross Validation



Final Accuracy = Average(Round 1, Round 2, ...)

Good Practice??



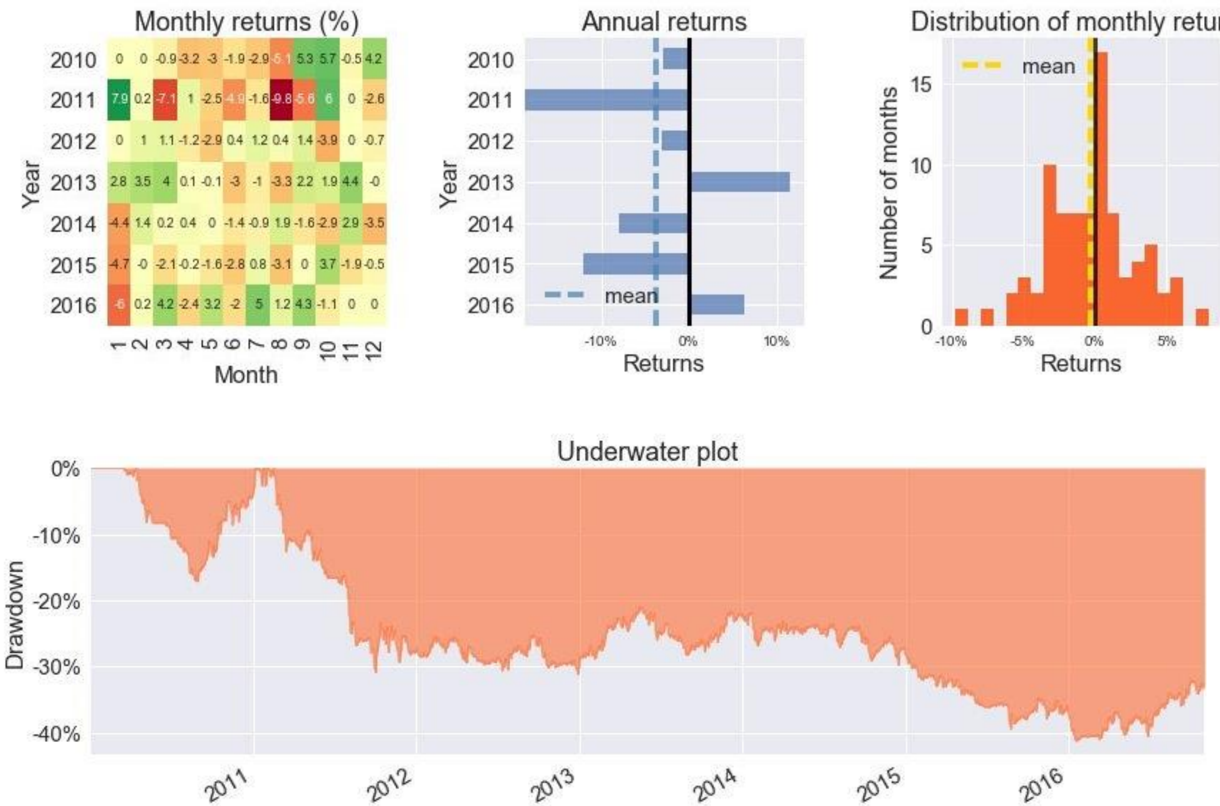
Why ML failed for Investing

INVESTIC

Evaluation Metric for *Trading Strategy*

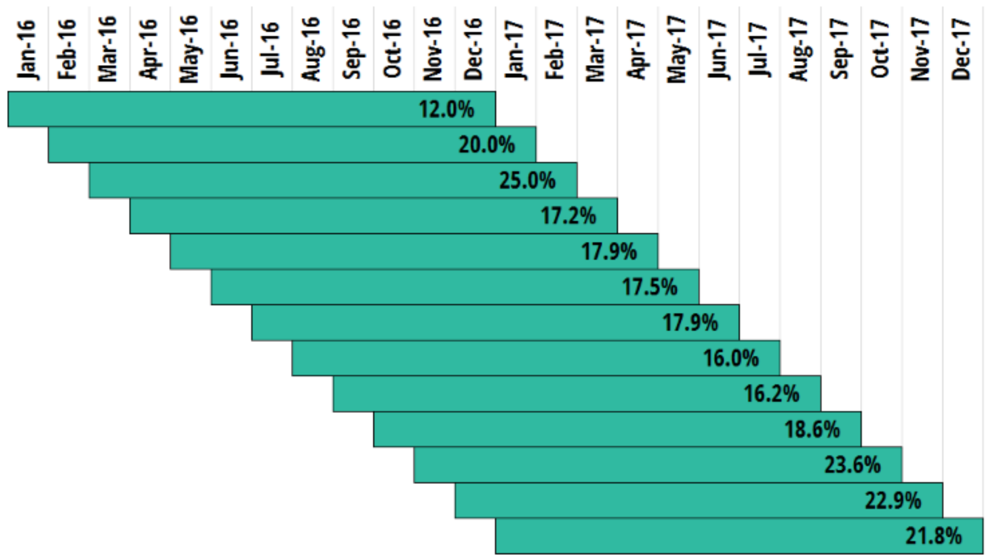
Trading Strategy Statistics	
52-Week High Breakout System - 5% trailing stop loss	
Total Number of Trades	**1927**
Testing Period Duration	**19 years**
Average Trades per Year	**96.4**
Average Trade Hold Time	**31.33 days**
Cumulative R	**168.71**
Percent of Winning Trades	**38.40%**
Average Return of Winners	**6.04%**
Average Return of Losers	**-3.06%**
Average Win-to-Loss Ratio (R-Multiple)	**1.98**
Largest Winning Streak	**8 consecutive trades**
Largest Losing Streak	**13 consecutive trades**
3 Largest Winners	**54.49%, 38.42%, 35.82%**
3 Largest Losers	**-13.2%, -12.9%, -12.52%**

Evaluation Metric for *Portfolio Management*



INVESTIC

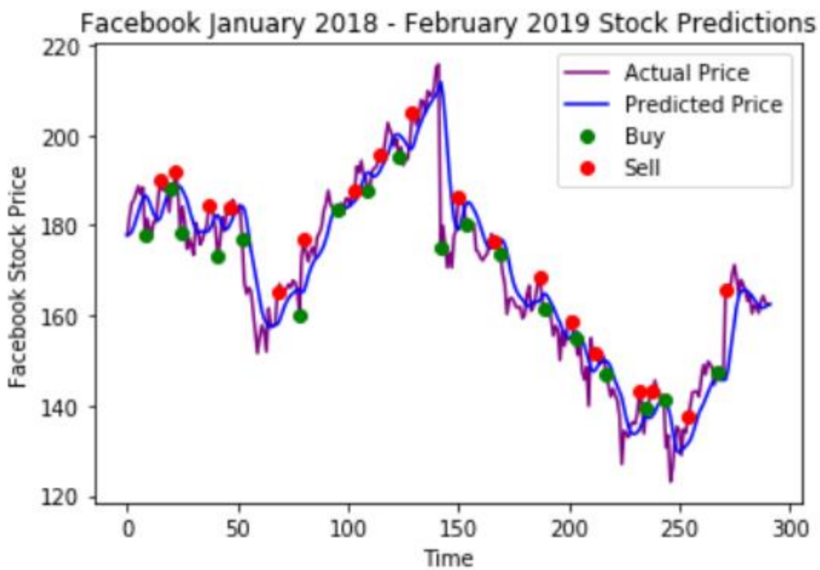
Evaluation Metric for *Portfolio Management*

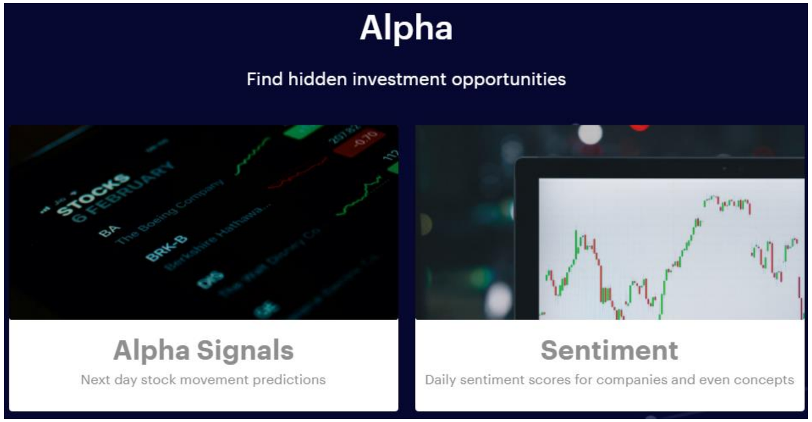


Any ideas for ML

Example of Applications from ML

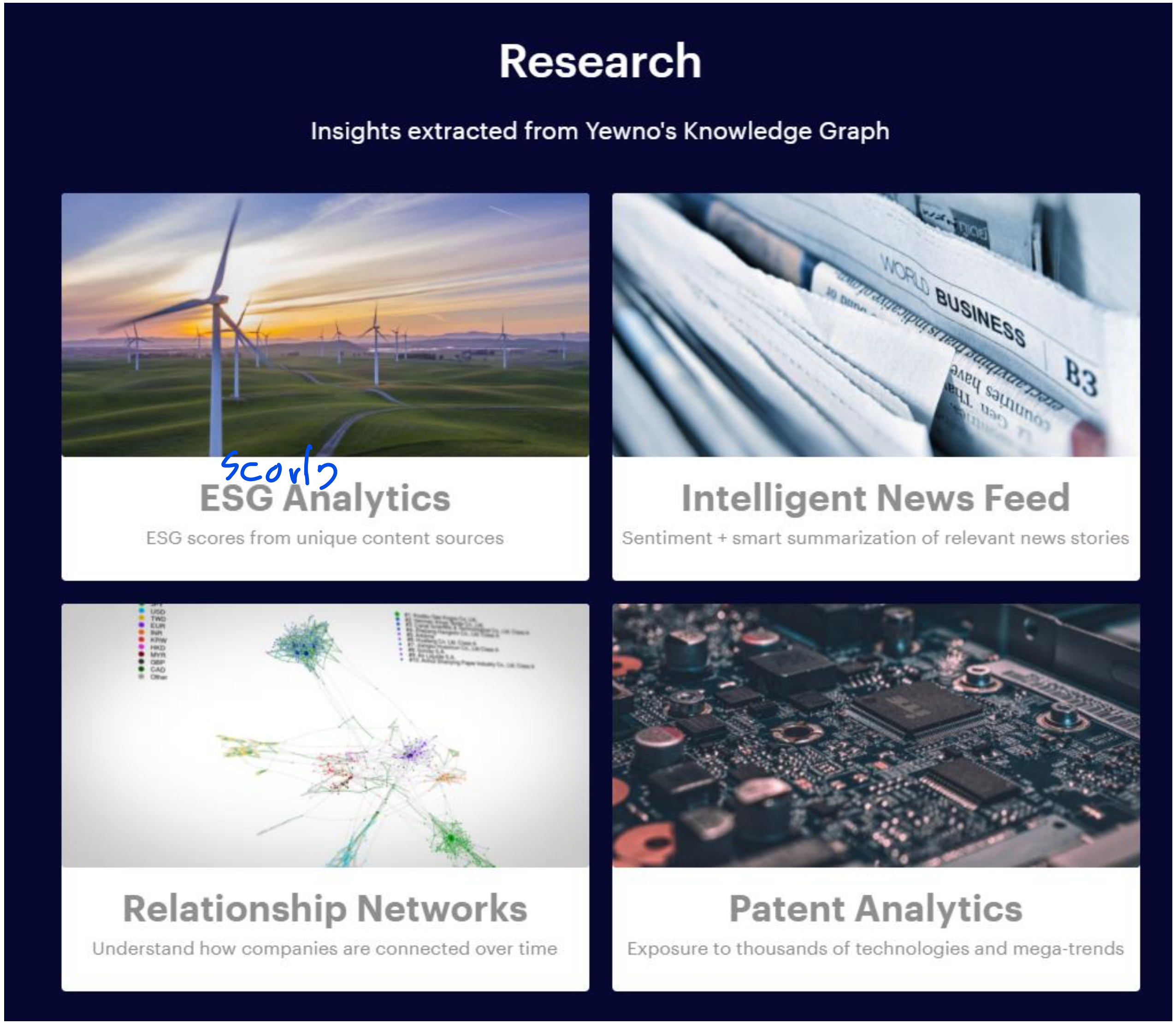
Is Signal the Input





Example of Applications from ML

เอาไปผนวกกับได้



Example of Applications from ML