



ML Process

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

ML 101

ML 101

ML 101

ML 101

ML 101

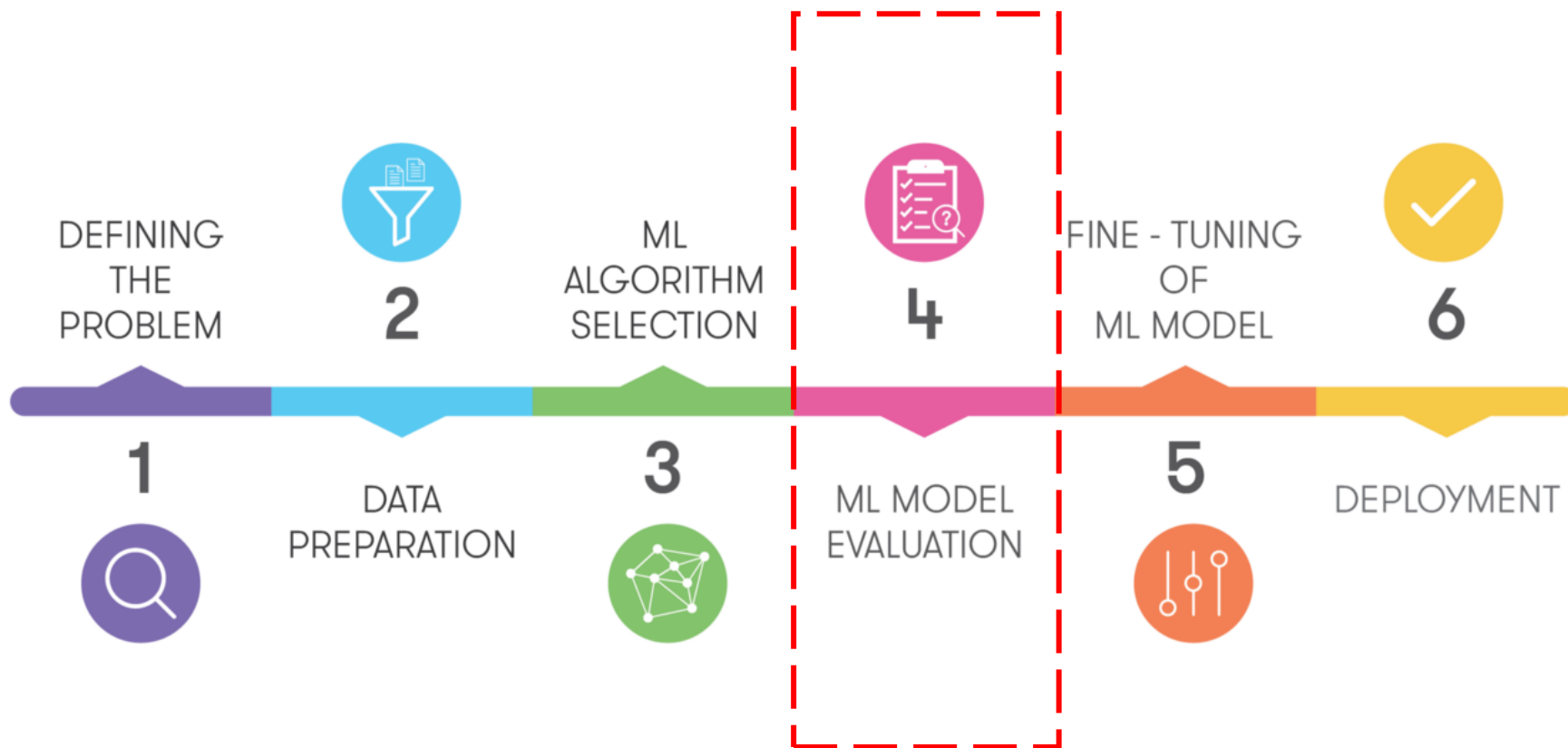
ML 101

Part of the ML process





Typical Modeling

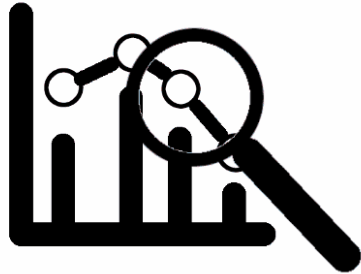




How do you know

–

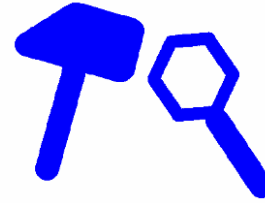
MODEL



Evaluation



GOOD?



→ 100%
အောင်မြင်စွာ



→ မအောင်မြင်စွာ
- overfit

BAD?

R.Brilenkov

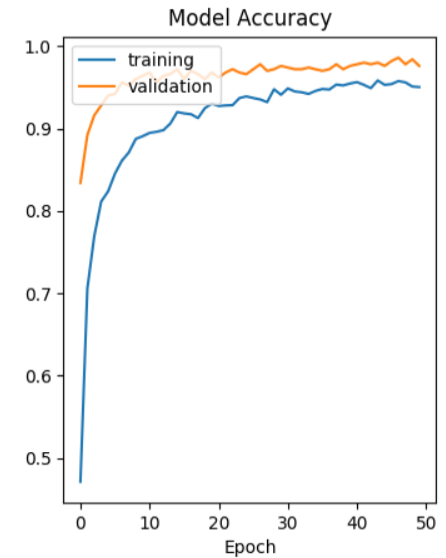
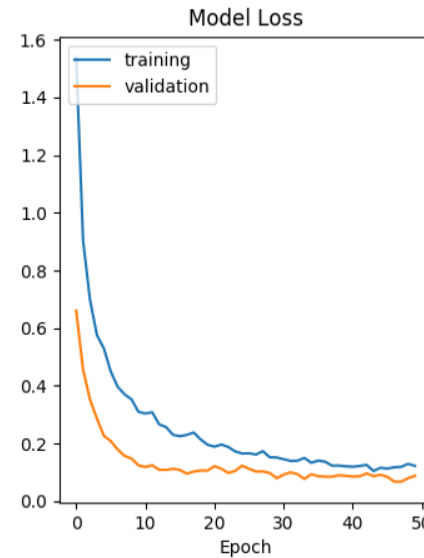
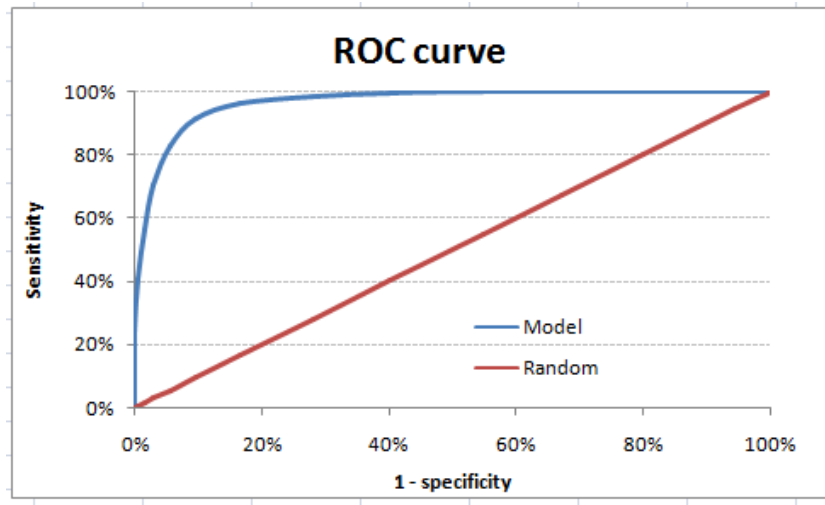


INVESTIC



Evaluation

model တွဲလျ Data sci

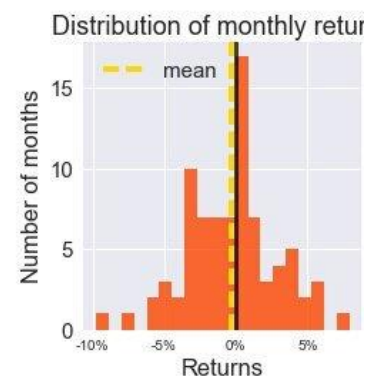
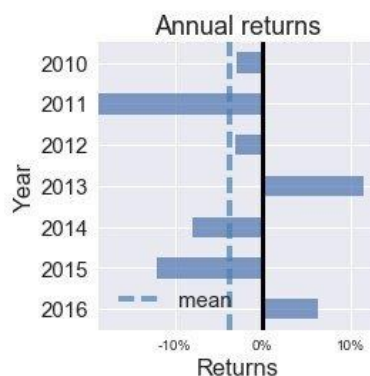


INVESTIC



Evaluation Metric for Portfolio Management

92107 5425974





Which is which

—

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

15 gini index / 0.00476
— 0.00476

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)





1. classif. cation ^{card} Tarihler 1 2 3

To 200, 10
 To 1000, 10
 To 10000, 10

Root Mean Squared Error (RMSE)

Root Mean Squared Log Error (RMSLE)

prediction

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

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

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



Regression

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

Root Mean Square Error

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (\text{Predicted}_i - \text{Actual}_i)^2}{N}}$$

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

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

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





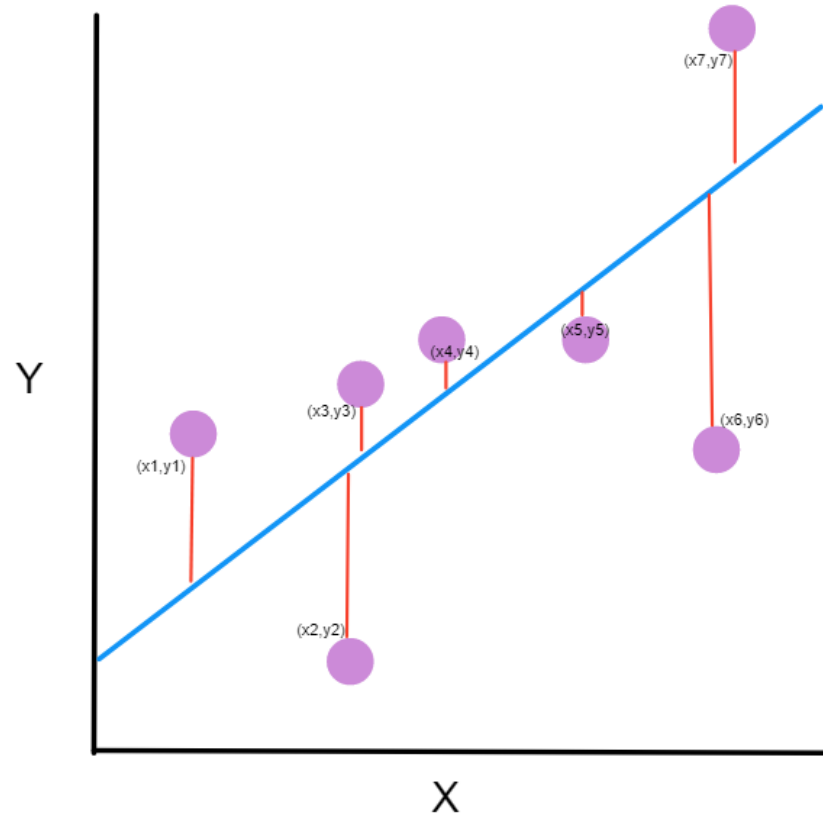
Regression

मा ऑल्लेव डेटा पॉइंट्स वर फिट लाइन

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

measure of how close a fitted line is to data points

Sum Square → Cancelling
Positive/Negative errors





Regression

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

very good for BO

'square root' → show large number deviations
punishes large errors.

Sum Square → Cancelling Positive/Negative errors

RMSE is highly affected by outlier values





Other Type??

Root Mean Squared Error (RMSE)

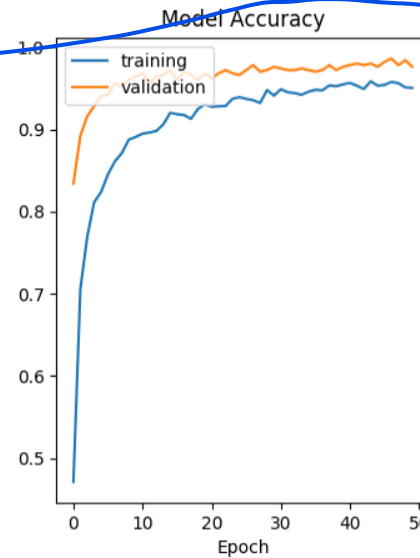
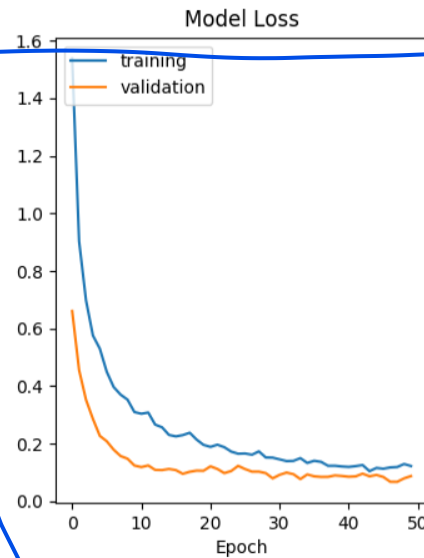
$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

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



✓ 18.5%





Other Type??

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

- 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)





Classification

9/21/21

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 = $(a+d)/(a+b+c+d)$	
		$a/(a+c)$	$d/(b+d)$		

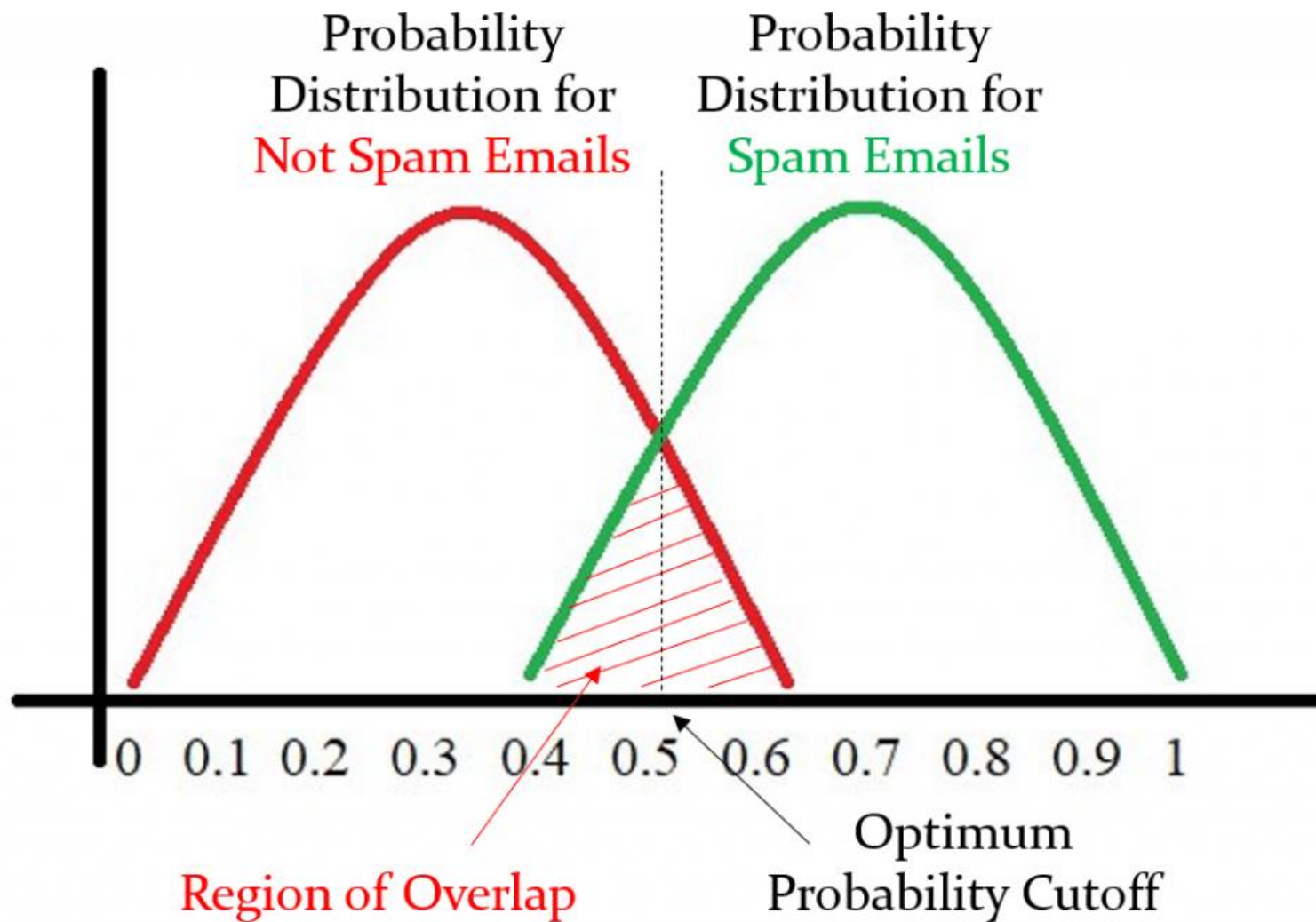
$$F_1 = \left(\frac{\text{recall}^{-1} + \text{precision}^{-1}}{2} \right)^{-1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$





Evaluating

–





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 = $1 - \beta$
Not rejected	Correct decision True negative Probability = $1 - \alpha$	Type II error False negative Probability = β



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)

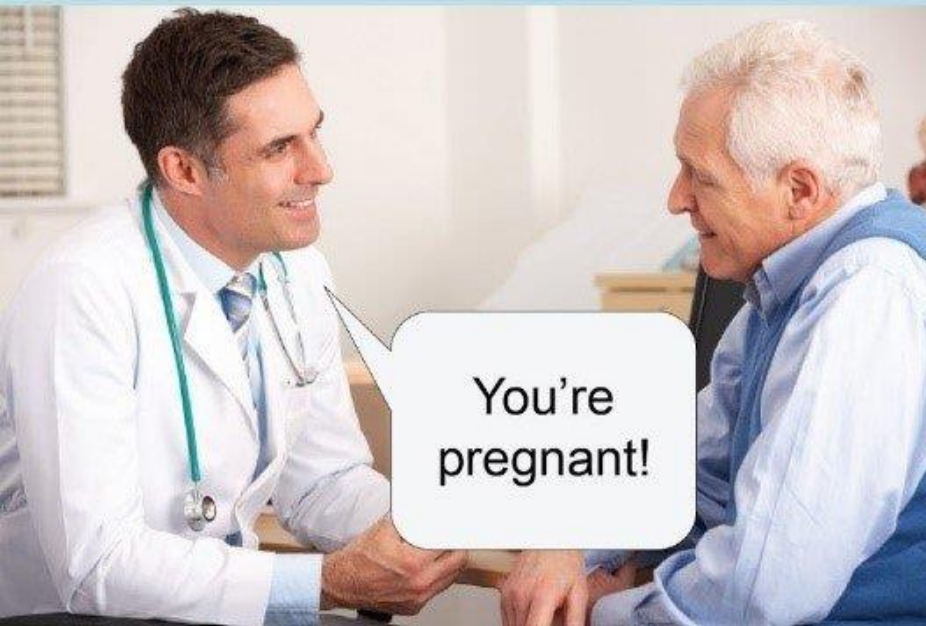




Evaluating

မျက်မှန်စစ်ဆေးခြင်း

Type I Error



အမှန်ကို မမြင်နိုင်ခြင်း

Type II Error





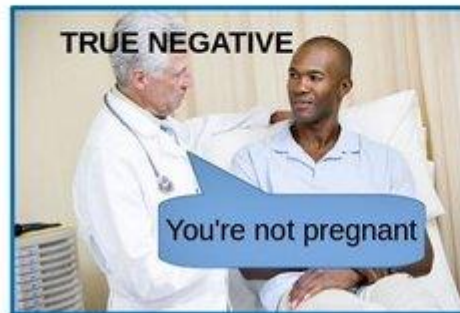
Evaluating

သိပ္ပံနာယက တွေ့ရှိချက် 1.1/α

$\hat{Y} = 0$
NEGATIVE

$\hat{Y} = 1$
POSITIVE

$Y = 0$
NOT PREGNANT



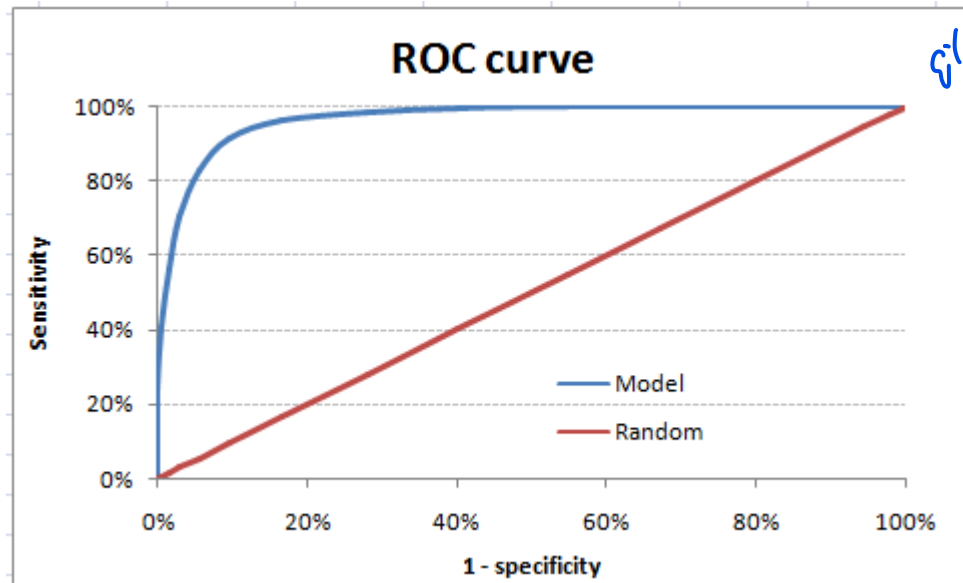
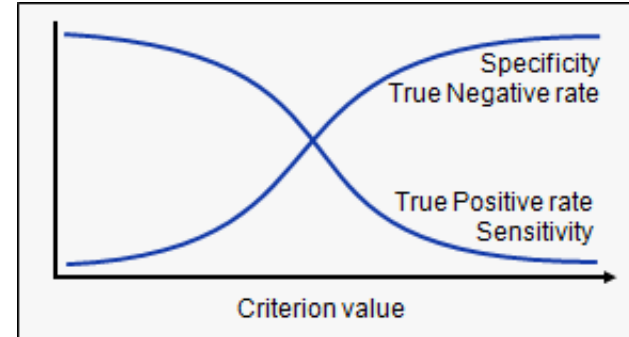
$Y = 1$
PREGNANT





Classification - ROC

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



ရရှိမှု / ကိန်း

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

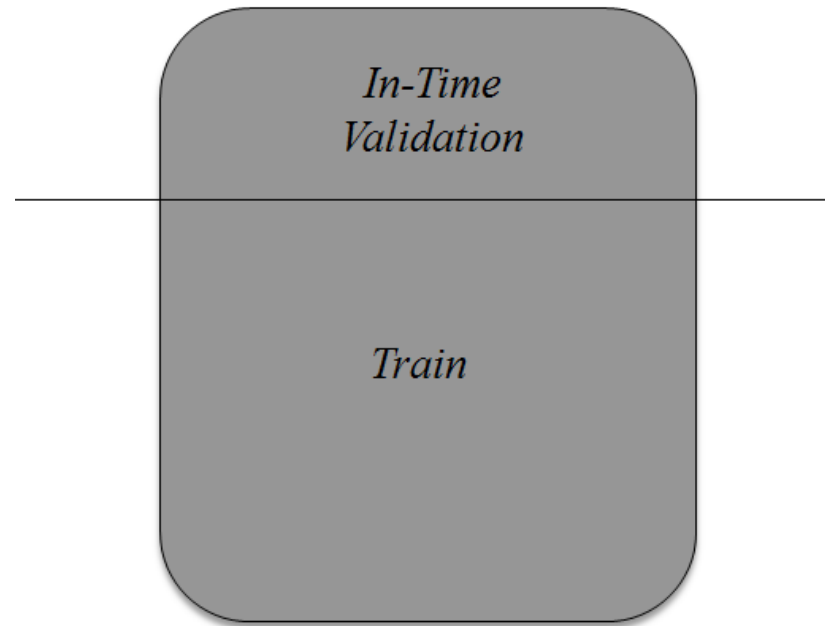




Cross Validation

–

Training Population



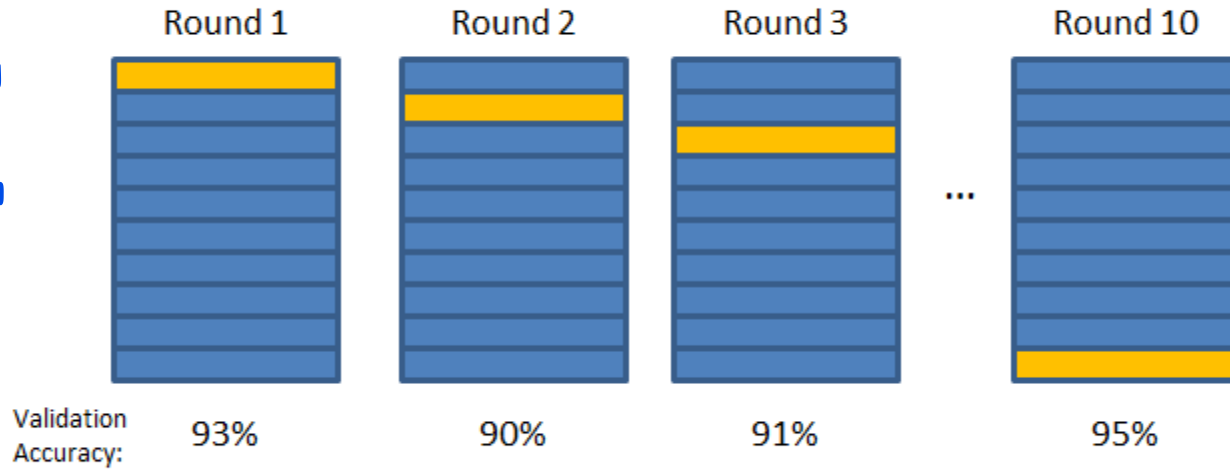


K Fold Cross Validation

အဲဒါကို 10 ခု၊ ၅ ခု ခုနစ် ၁၀ ကပ်စ် ၁၀

Validation Set
Training Set

ဒီ window
အဲဒါကို ၁၀ ခု

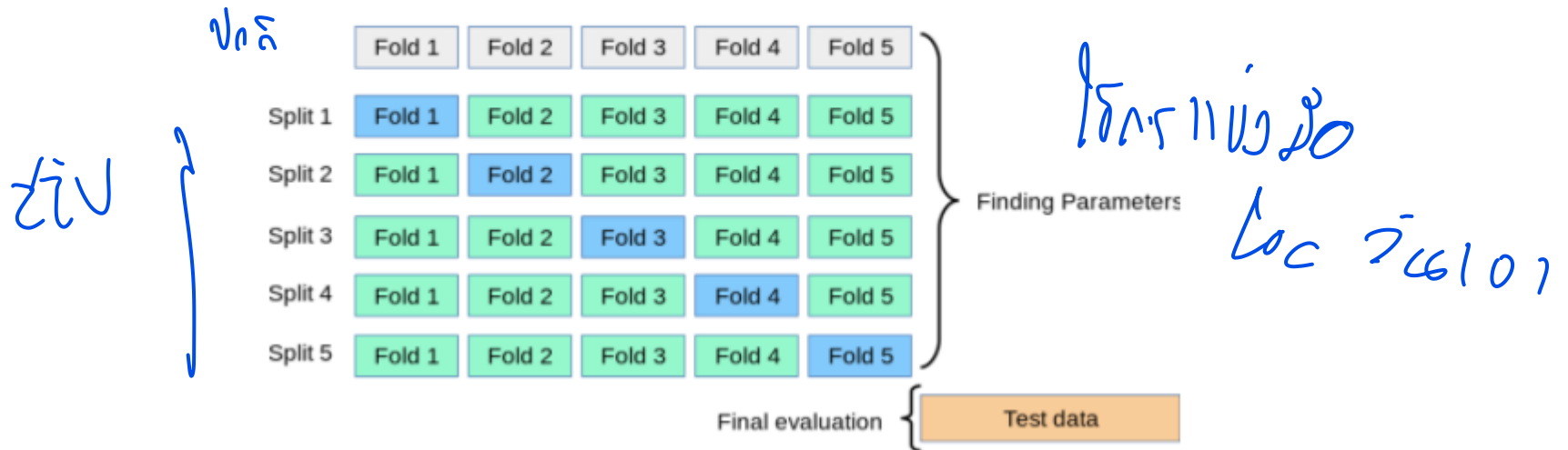
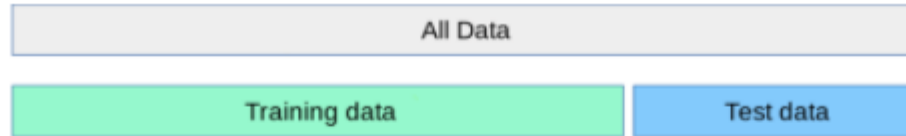


Final Accuracy = Average(Round 1, Round 2, ...)





Good Practice??



Why ML failed for Investing



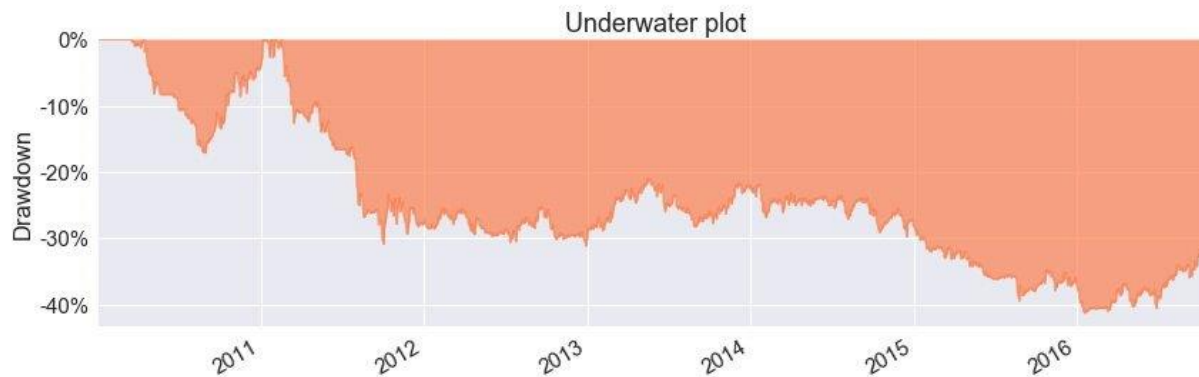
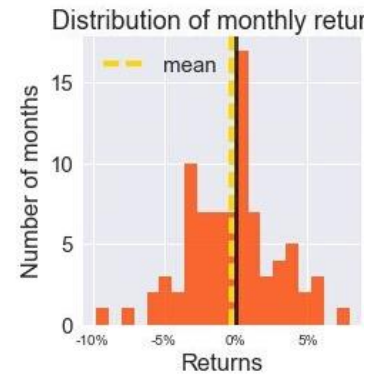
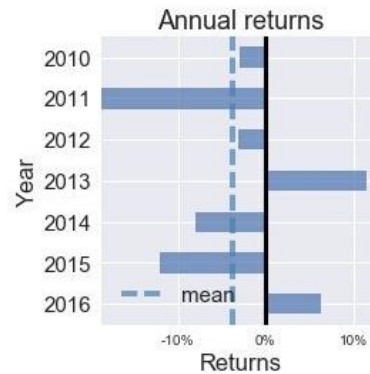
Evaluation Metric for *Trading Strategy*

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%





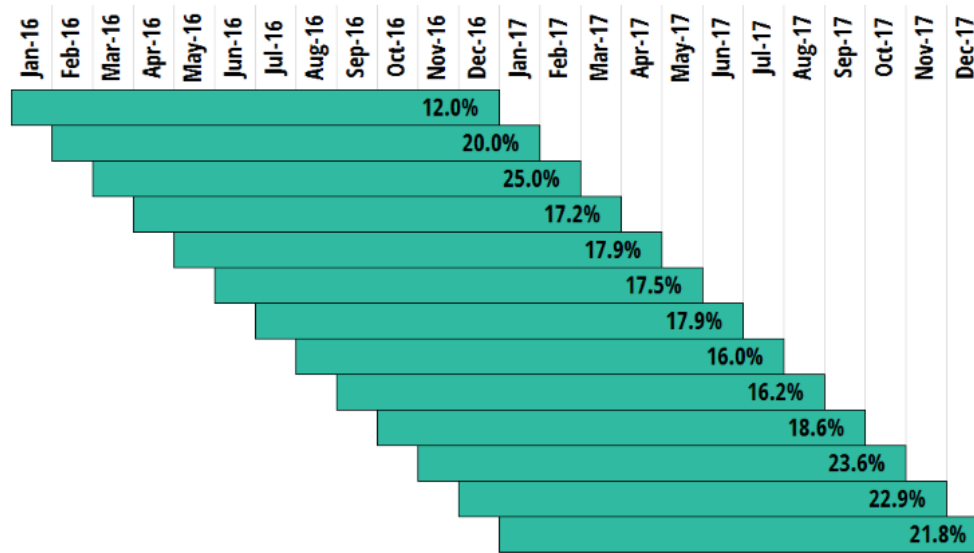
Evaluation Metric for *Portfolio Management*





Evaluation Metric for *Portfolio Management*

Rolling Return
or
Fama
WPS

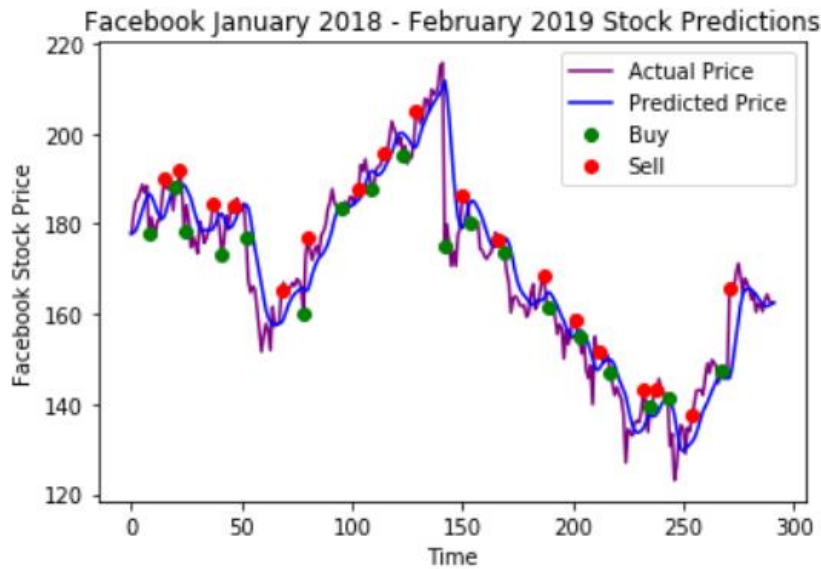


Any ideas for ML



Example of Applications from ML

Is Signal the Profit



Alpha

Find hidden investment opportunities

Alpha Signals

Next day stock movement predictions

Sentiment

Daily sentiment scores for companies and even concepts



INVESTIC



Example of Applications from ML

Handwritten blue text: "Yewno" with an arrow pointing upwards.

Research

Insights extracted from Yewno's Knowledge Graph



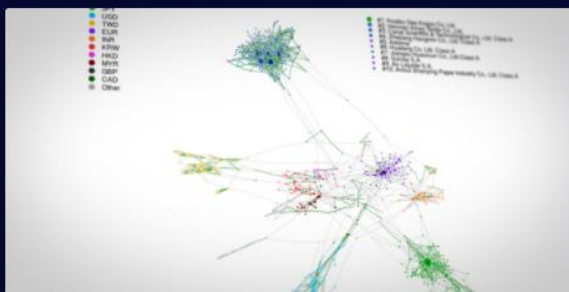
ESG Analytics

ESG scores from unique content sources



Intelligent News Feed

Sentiment + smart summarization of relevant news stories



Relationship Networks

Understand how companies are connected over time



Patent Analytics

Exposure to thousands of technologies and mega-trends



INVESTIC

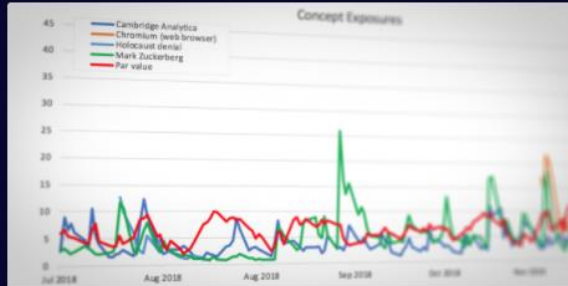


Example of Applications from ML

bt sec ✓
usdt 5425/✓

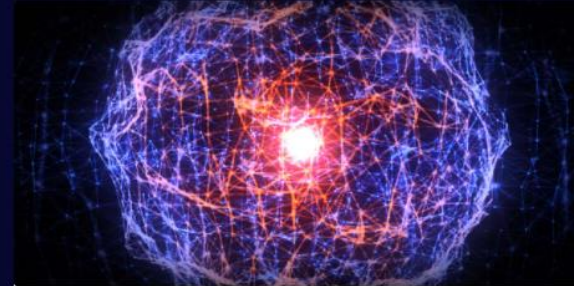
Risk Analysis

Exposures derived from Yewno's Knowledge Graph



Concept Exposures

Dynamic exposures to non-traditional risk factors



Systemic Risk

Daily network-level company systemic risk scores



Judicial Exposure

Judicial performance analytics and exposures for
global companies



Country Exposure

Company level network exposure to countries across
time



INVESTIC