Agenda

- A new Business Case
- Imbalanced Data
- Problem with Accuracy
- More Robust Metrics
  - Confusion Matrix
  - Precision Score
  - Recall Score
  - F1-Score

Business Case

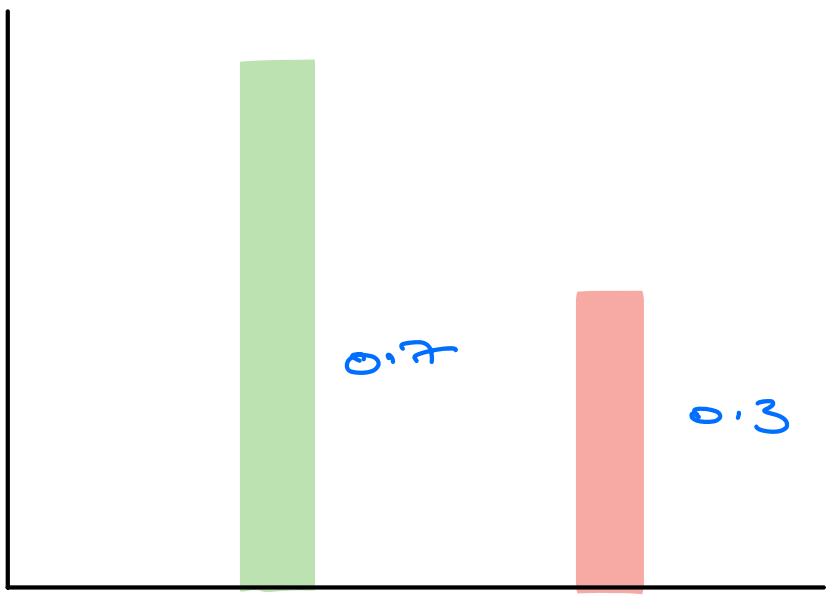
Text	Spam
Hi -------	1
Jim scales ----	0

0 : Non Spam

1 : Spam

* Objective : Build a ML Model which can classify emails Spam or Not Spam.

Data Distribution



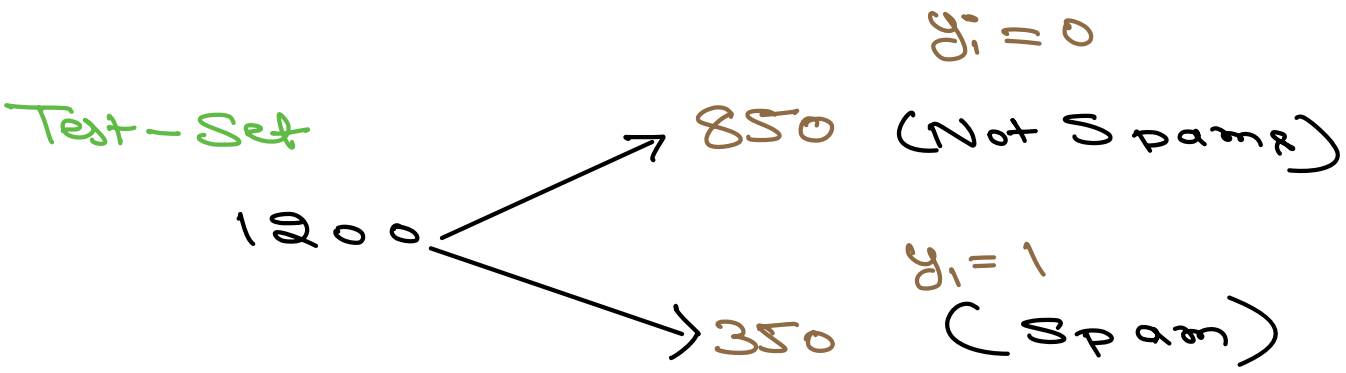
Acc @ 0.93

93%

20 : 100 D1 : Imbalance

80 : 100 D2 : Slightly imbalanced

Problem with Accuracy



Dumb Model: Predicts Every Email as Not-Spam

$\hat{y} = 0$

Accuracy = $\frac{850}{1200} = 70.8$

Case 2:

Credit Card Transaction → Fraudulent / Non F

NF 99.5% | F 0.5%

Extreme imbalanced

Always Says → NF

Acc. 99.95%

How to Handle Imbalance:

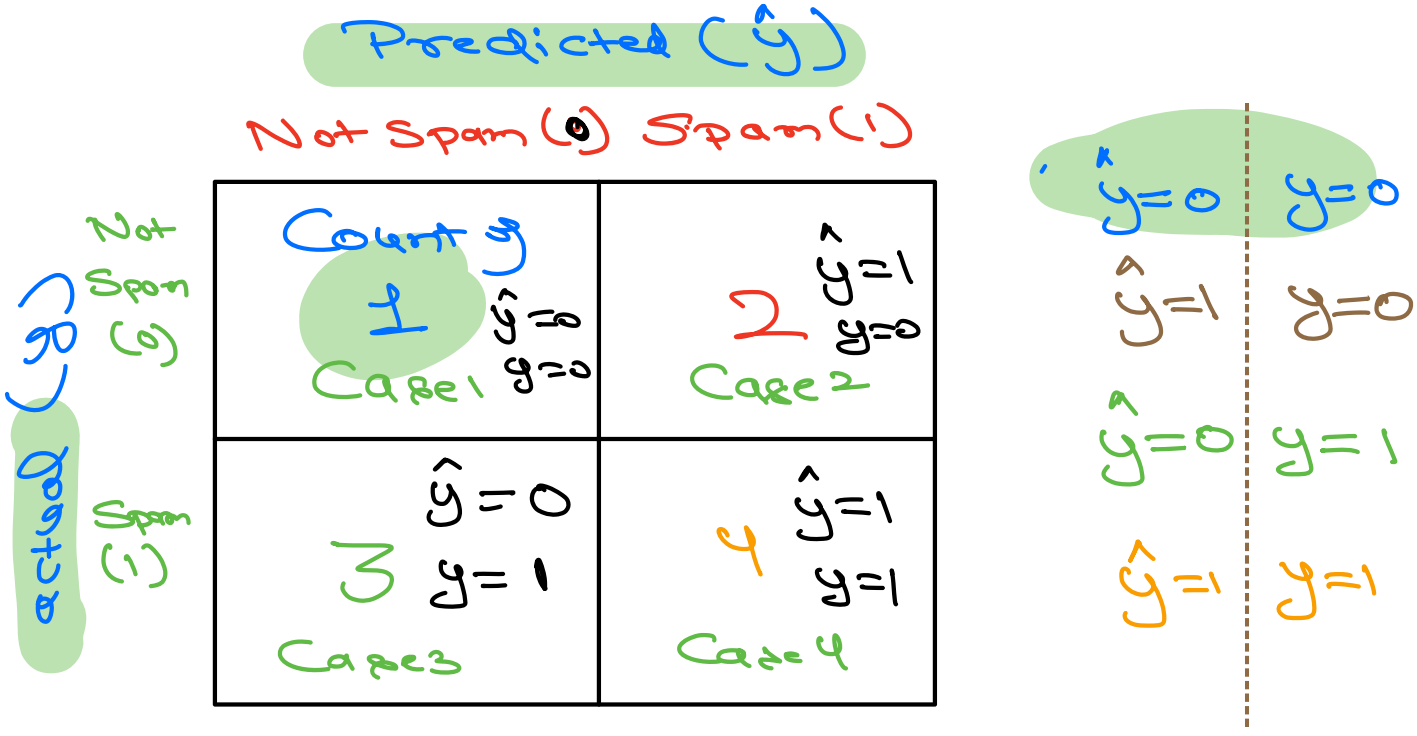
1. Data Augmentation

2. Update Loss Function

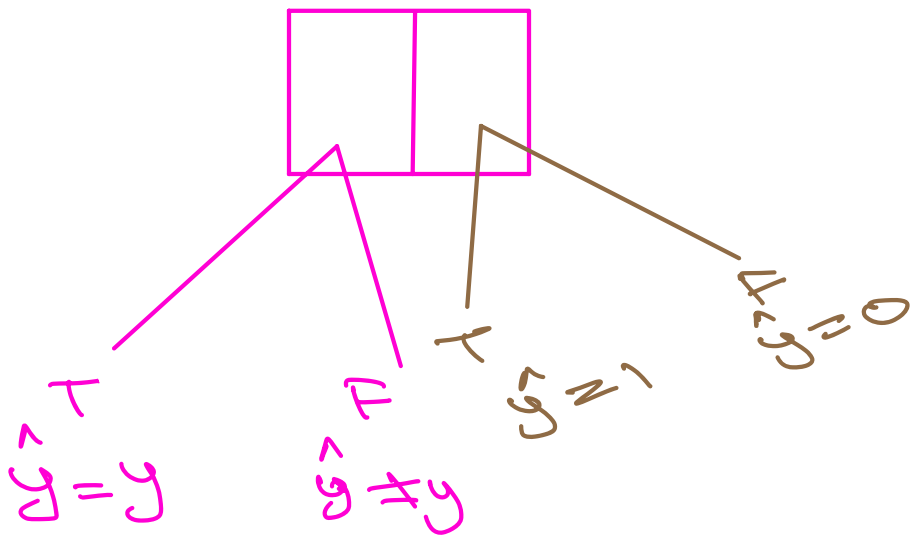
3. Change the Metric ✓

Learn Alternative of Accuracy that give us fair Estimate of Performance irrespect of Distributi

Confusion Matrix

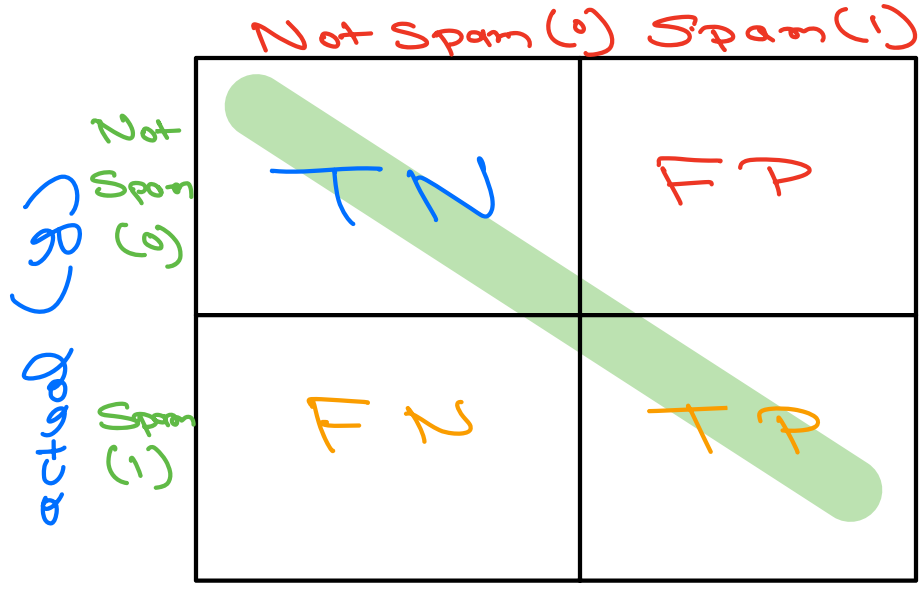


①	$\hat{y}=0$	$y=0$	True Negative (TN)
②	$\hat{y}=1$	$y=0$	False Positive (FP)
③	$\hat{y}=0$	$y=1$	False Negative
④	$\hat{y}=1$	$y=1$	True Positive



* First term tells us 'y model pred is Correct/incorrect

* Second Term tell us pred

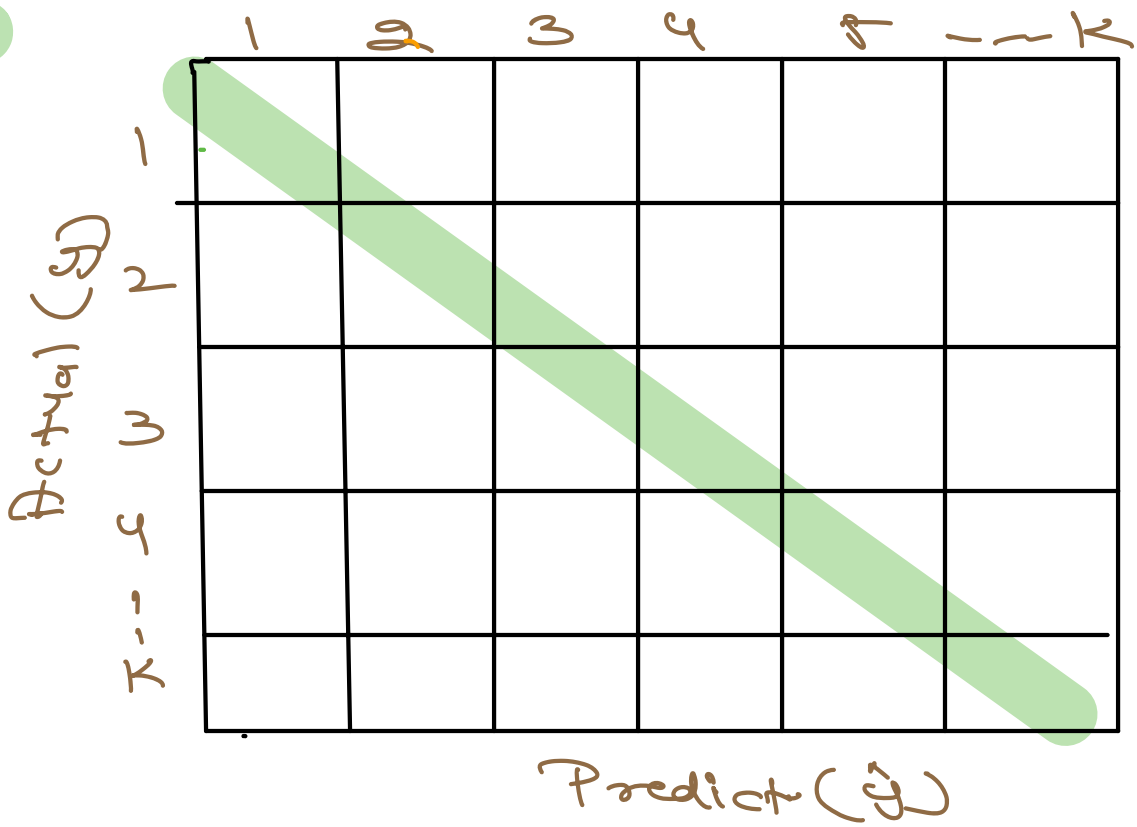


Questions

$$\text{Acc} \Rightarrow \frac{TP + TN}{TN + FN + TP + FP}$$

How about multi-class Classification

K x K



Understanding Confusion Matrix

Test data
400 DP

- 360 Not Spam
- 40 Spam

Pred

Actu

Dumb Model

⊙ $\hat{y}$ = majority

Actual \ Pred	Not Spam (0)	Spam (1)
Not Spam (0)	360 TP	0 Type-1
Spam (1)	40 Type-2	0 TN

Acc: $\frac{360}{400}$

⊙ 90%

Ideal Model

Actual \ Pred	Not Spam (0)	Spam (1)
Not Spam (0)	360	0
Spam (1)	0	40

$\frac{400}{400} = 100\%$ Acc

For Ideal Model, which of the following is true?

4 options
Active Duration (Most preferred: 30 seconds)

Appears for 60 Secs

- A FP and FN ⇓, while TP and TN ⇑
- B TP and TN ⇓, while FP and FN ⇑
- C TP and FN ⇓, while FP and TN ⇑
- D FP and TN ⇓, while TP and FN ⇑

Type of Errors:

⇒ FP ⇒

⇒ FN ⇒

Precision

Scenario $y = 1$ $y = 0$ ⇒ FN

1) Recieving a Spam in Normal Mailbox

2) Offer letter sent to Spam

Actual ⇒ 0

Pred ⇒ 1

Spam-folder

⇒ FP

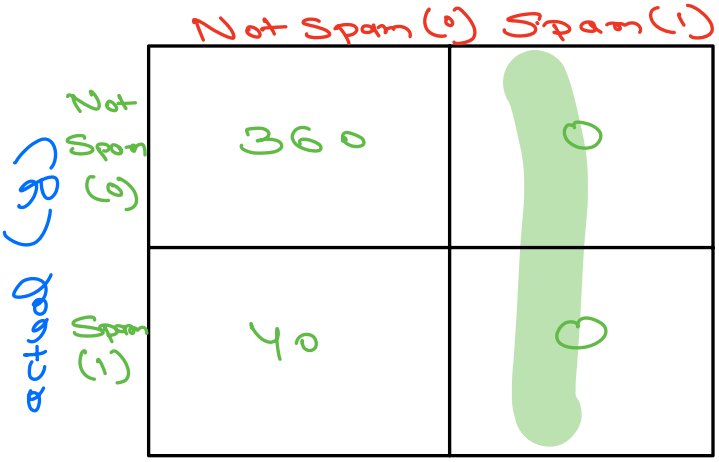
FP is Dangerous

Goal should be to minimize FP

Precision @ $\frac{TP}{TP+FP}$ $\hat{y}=1$

Among all $\hat{y}_{pred}=1$, How much was actual=1

Dumb Model



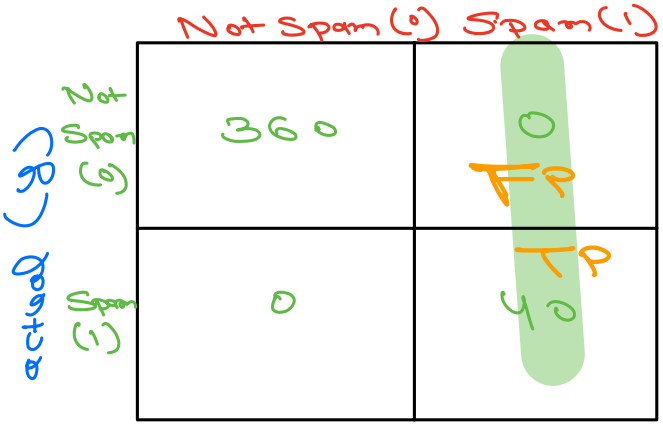
F

$\frac{TP}{TP+FP}$

$\frac{0}{0+0+40}$

⇒ Precision ⇒ 50

Ideal Model



$\frac{TP}{TP+FP}$

⇒ $\frac{40}{40+0}$

Range of Precision @ $[0, 1]$

Recall

Cancer Diagnostic

Class 0 : No Cancer — Negative

Class 1 : Cancer

① Healthy patient diagnosed as Cancer

Actual 0 0 / 0 1 → F → P

② Cancer patient Diagnosed as Healthy

Y 0 / Y 0 → F Negative

FN is dangerous

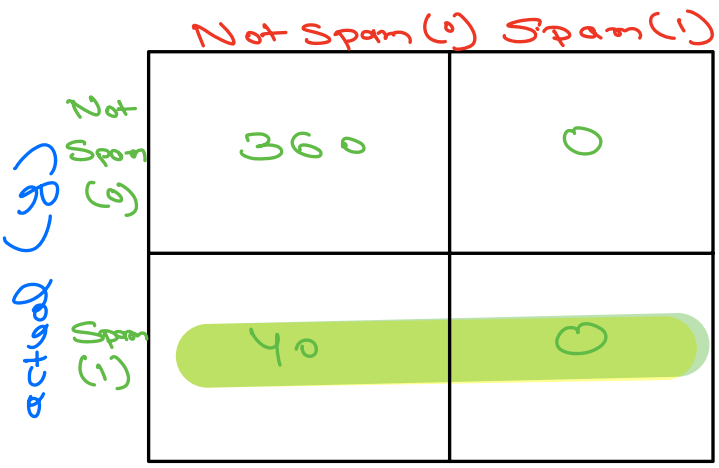
$$\text{Recall} = \frac{TP}{TP+FN}$$

Ideal Model

Actual \ Predicted	Not Spam (0)	Spam (1)
Not Spam (0)	360	0
Spam (1)	0	40

$$= \frac{40}{40+0} = 1$$

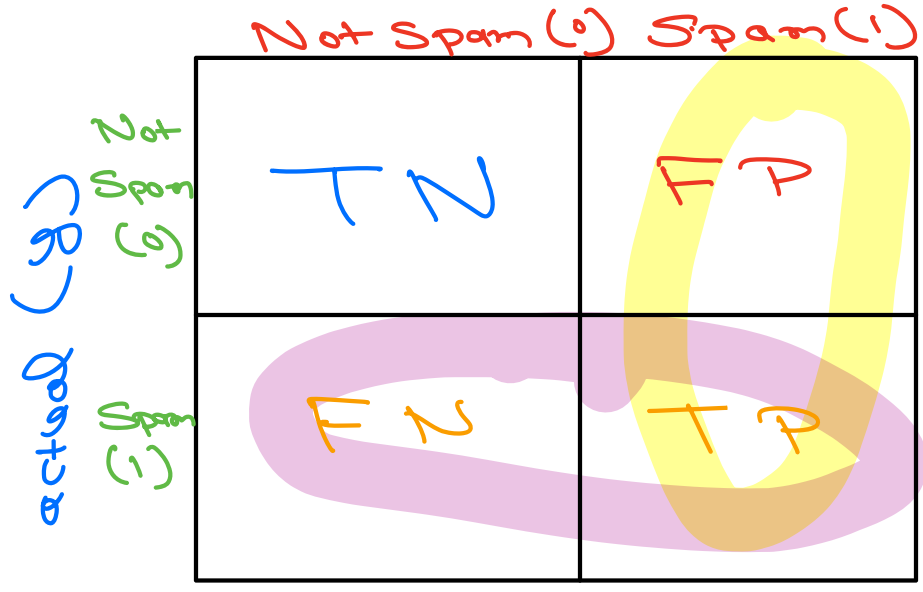
Dumb Model



Recall
0 0
0 + 40
0 0

Range of Recall) (0, 1)

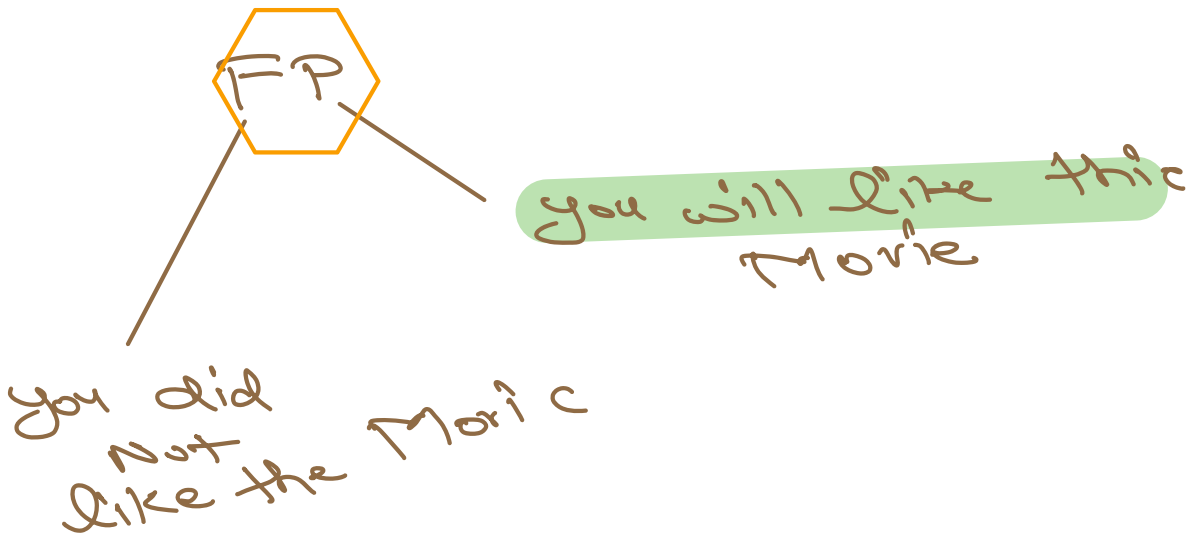
Confusion Matrix



Numerator is alway TP

Precision ⇒ Denom is last Col

Recall ⇒ Denom is last Row



min FP ⇒ Precision Max

* min FP ⇒ Precision Max

---

if Both FP and FN are your Mistakes

F1 - Score : $\frac{1}{\frac{1}{P}+\frac{1}{R}} \Rightarrow \frac{2PR}{P+R}$

$\frac{P+R}{2}$

	P	R	Avg $(P+R)/2$
M1	0.3	0.8	0.55
M2	0.9	0.3	0.55
M3	0.7	0.40	0.55

	P	R	F-Score $\frac{2PR}{P+R}$
M1	0.3	0.88	0.43
M2	0.9	0.2	0.33
M3	0.7	0.40	0.51

Best model

F1 - Score adds penalty if either of P and R is Low

Why does the F-1 score use Harmonic Mean (HM) instead of Arithmetic Mean (AM) ?

4 options

Active Duration (Most preferred: 30 seconds)

Appears for	60 Secs

A	AM penalizes models the most when even Precision and Recall are low.
B	HM penalizes models the most when even Precision and Recall are low.
C	HM penalizes models the most when even Precision and Recall are high.

P is Low → F1
R is High

R is Low → F1
P is High

CC → NP
P