Introduction : Multi Class Classification

Classification.

1. Binary → email
   - Spam ①
   - Non spam ②

   2 class

2. Multi class Classification.

   >2 class

   Email
   - Promotional ✓
   - Social ✓
   - Primary ✓

How do machine learn: ?

— looks for the eg:

Promotional → Offers ✓
→ discount ✓
→ 25% discount ✓

} Independent Variable | Feature

The result is called Dependent Variable | Target
→ Promotional ✓
→ Social ✓
→ Primary. ✓

2-Step

1. Training. ✓
2. Prediction

We are given information
→ Various biz ✓
→ location details ✓
→ Txn mode ✓
→ Time line txn ✓

} Independent Variable.

What would be the status of their licence

→ Approved → ①

→ Cancelled } ②

→ Revoked ③ etc

1. Data Imbalance ✓

2. Missing data. ✓

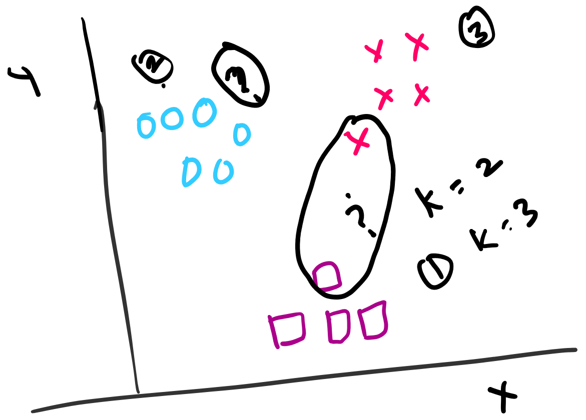
3. Data leakage

4. Outlier Treatment.

5. Encoding.

KNN - K Nearest Neighbors

KNOW A PERSON by looking at people surronding him.



1) Euclidean → continous
2) Hamming → discrete

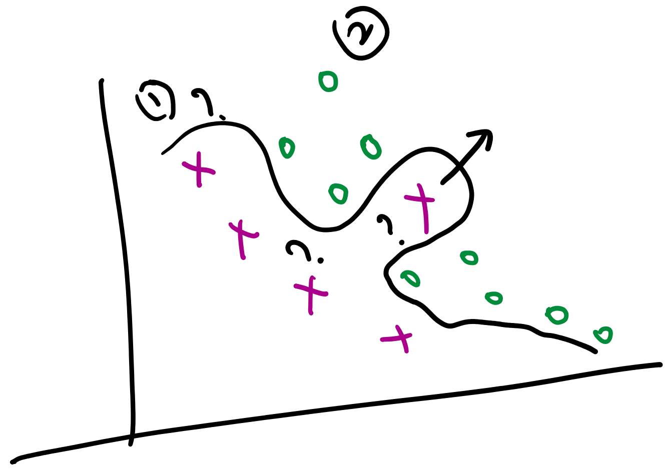
How to choose the value of K → no. of neighbours

→ odd value K = 5, 7, 9

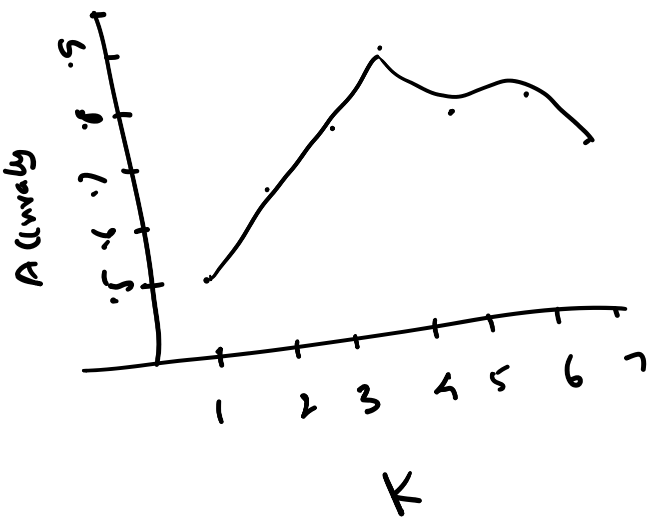
→ Not too low (or) too large.

K = 1
→ high variance

K = 10
→ high bias
→ Biased class of.

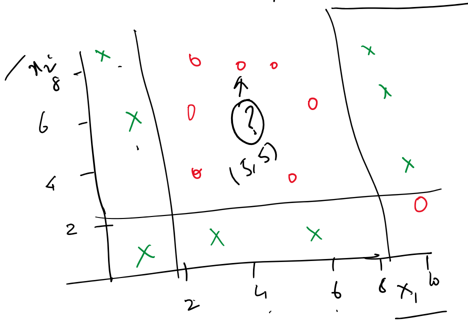


Plot of K (VS) Accuracy.

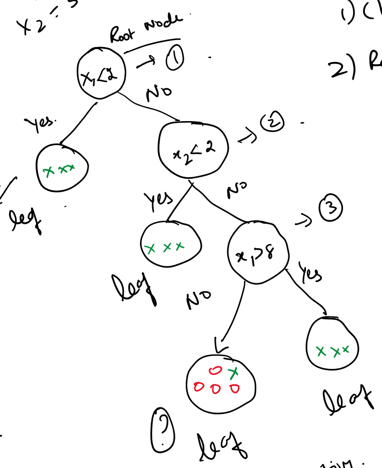

Accuracy
.5
.6
.7
.8
.9
1
2
3
4
5
6
7
K

Decision Tree

Good classifier for Non-linear data points.



$x_1 = 5$
$x_2 = 5$



Pure Index

1) Classification.
2) Regression.

linear (vs) Non linear.
↓ Single
↓ Classifier

How do we arrive at the criteria.

① We go for conditional which give pure Node.

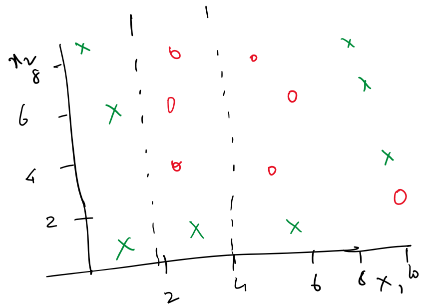
Majority voting.
5 data.
4 - Red → Red.
1 - green

How do we know which condition gives pure Node.

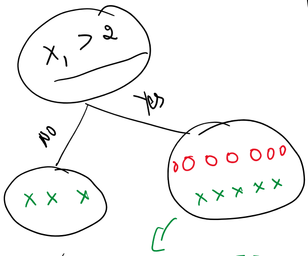
(i) Gini Index.

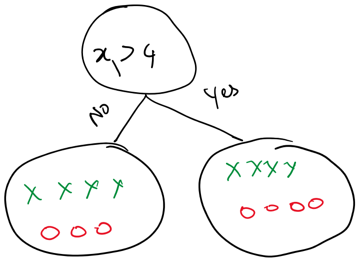
$$= 1 - \sum_{i=1}^{n} P_{(i)}^2$$

$n \to$ no. of class
$P_i \to$ probability of class



Let us understand it better with actual example we saw above.





$P(x) = 3/3 = 1$

$P(o) = 0$

$P(x) = 5/12$

$P(o) = 7/12$

$P(x) = 4/7$

$P(o) = 3/7$

$= 0.48$

$P(x) = 4/8$

$P(o) = 4/8$

$= 0.5$

$\text{Gini}_{\text{left}} = 1 - (3/3)^2 - 0^2$

$= 0$

$1 - (5/12)^2 - (7/12)^2$

$\text{Gini}_{\text{Right}} = 0.48$

$\text{Total Gini} = 7/15 \times 0.48 + 8/15 \times 0.5$

$= 0.49.$

G impurity is high.

$\text{Total Gini} = 3/15 \times 0 + 12/15 \times 0.48$

$= 0.38$

↓ lower is better

impurity is less

Advantage:

data points.

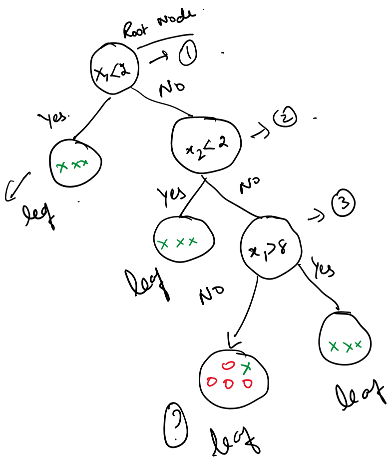
Advantage :

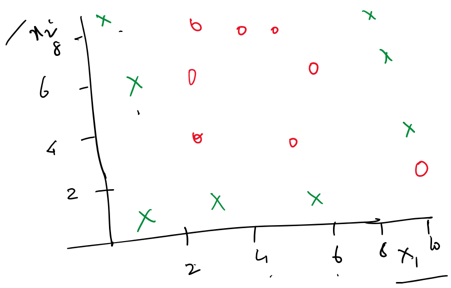
1) Non-linear data points. ↪ imp.

Greedy Algo.

DisAdvantage.

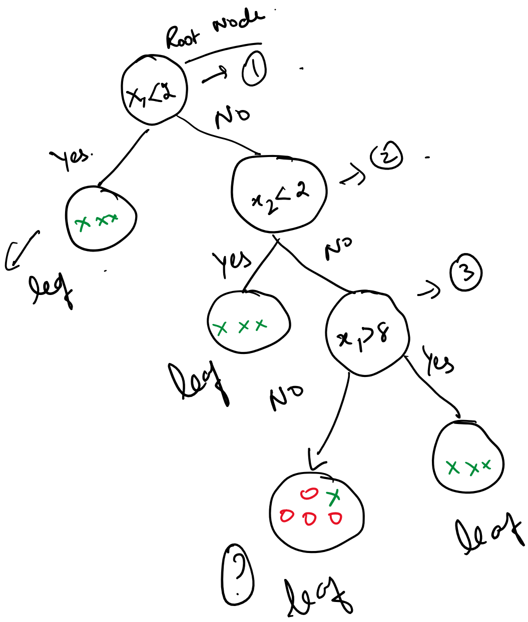
1) High Variance / Over fitting.

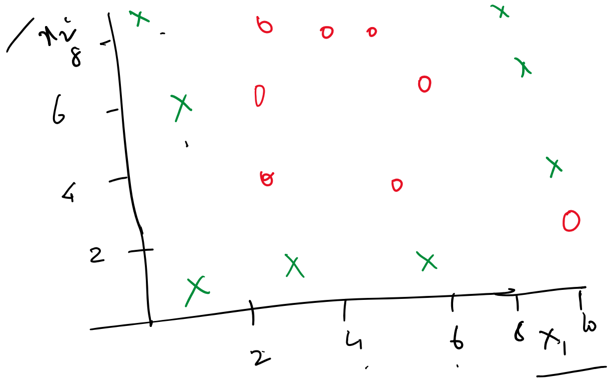




Model Comparison

	Accuracy	Precision	Recall	F1 score
KNN	0.78	0.62	0.53	0.56
Naive Bias	0.73	0.29	0.27	0.27
Logistic	0.68	0.58	0.79	0.59
Decision Tree	0.86	0.89	0.92	0.90

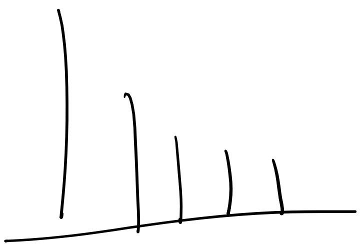




EDA

Univariate Analysis.

Catagorical - Bar chart.



Continous. → Histogram
(distribution.

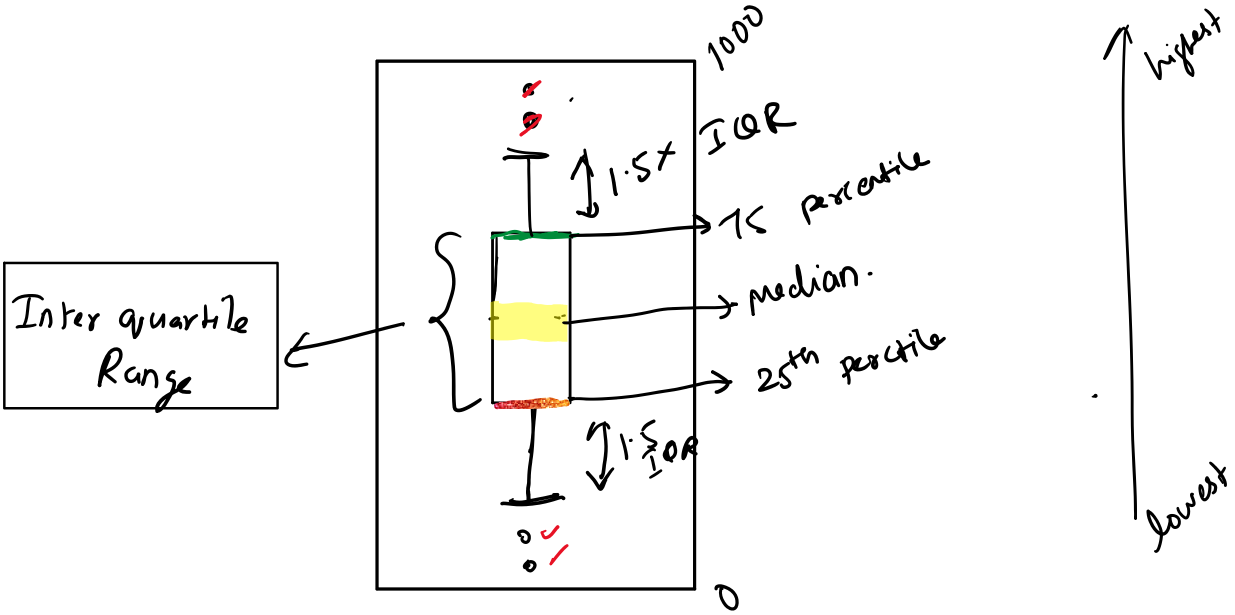
Bivariate Analysis. → 2 Variable
at a time

1) Stacked bar.

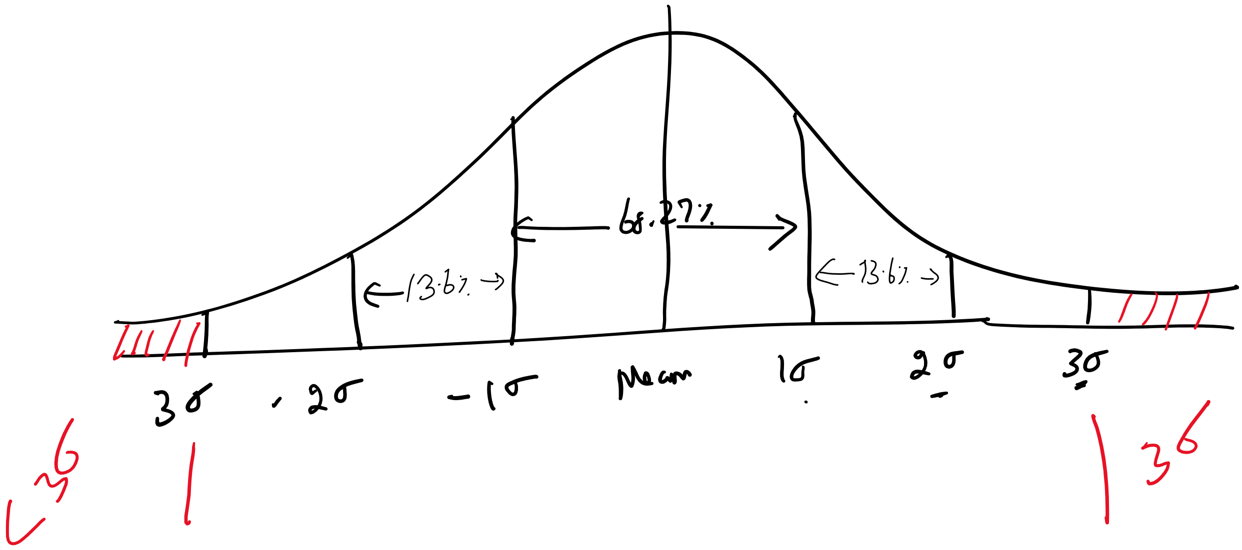
2) Scatter plot

Outlier Detection

Box plot / Inter quartile Range.



Standard deviation.



Naïve Bias

Bayes Theorem

$$P(c/X) = \frac{P(X/c) * P(c)}{P(X)}$$

$P(c)$ → Class prior probability

X → feature

C → class.

$P(c/X)$ ↳ x_1 x_2 x_3

$P(X/c)$ = Probability of features given the class.

$P(C)$ = Probability of class

10 → NS − 8, S − 2 } 10 email

$P(NS) = 8/10 = 0.8$

$P(S) = 2/10 = 0.2$

features

	NS	P(x/c)	spam	P(x/c)
Dear	40	40/100 = 0.4	2	2/50 = 0.04
Thanks	30	30/100	1	1/50 = 0.02
friend	18		2	
offer	2	= 2/100	25	25/50 = 0.5
discount	10	10/100	20	20/50 = 0.4
	100		50	

"Dear Friend"

$P(NS/X) = P(NS) \times P(Dear/NS) \times P(Friend/NS)$

$= 0.8 \times 0.4 \times 0.3$

$= 0.096$

Dear Friend

$P(S/x) = P(S) \times P(Dear/S) \times P(Friend/NS)$

$= 0.2 \times 0.04 \times 0.02$

$= 0.00016$

offer discount → spam.

$= 0.8 \times 0.02 \times 0.1$

$= 0.0016$

$= 0.2 \times 0.5 \times 0.4$ } spam.

$= 0.04$

"good" "morning"

0 0

Probability > 0.5 or 80

Dear Friend is a Non Spam

Assumption.

Features are independent.

Advantage

simple, fast and does surprisingly well

→ we have lot categorical variable. → independent.

DisAdvantage

1. Zero frequency
2. probability estimate is not to be seriously

→ Relative

Data Cleaning

Imputation.

① Delete Imputation.

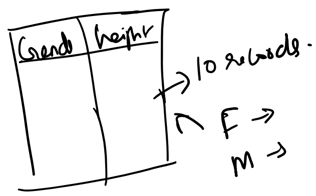
look at the no. of missing values

If the missing is large

② Mean / Median / Mode Imputation: → most common.

↳ Generalized Imputation → complete mean.

↳ Similiar Case Imputation.



③ Predictive Imputation.

① Random Sampling.

Over sampling – the minority

Under Sampling → Majority Class

② SMOTE → over sampling.

↳ Create synthetic data.
→ which similiar to real data
but has some diff.

① Dummy Encoding.

bizId	State
1	Chicago
2	Florida
3	Chicago
4	Florida
5	California

⇒

bizId	city_chicago	City Florida
1	1	0
2	0	1
3	1	0
4	0	1
5	0	0

N → N-1 feature.

Drawback.

If N becomes large it will be difficult to handle.

eg: 30 diff. State needs 29 feature

If we have col. like city with 200 different city then we need 199 features.

Target Encoding.

State	Target	City
Florida	0	0.33
Chicago	1	0.25
Florida	1	0.33
California	0	0.66
Chicago	0	0.25
California	1	0.66
California	1	0.66
Chicago	0	0.25
Chicago	0	0.25
Florida	0	0.33

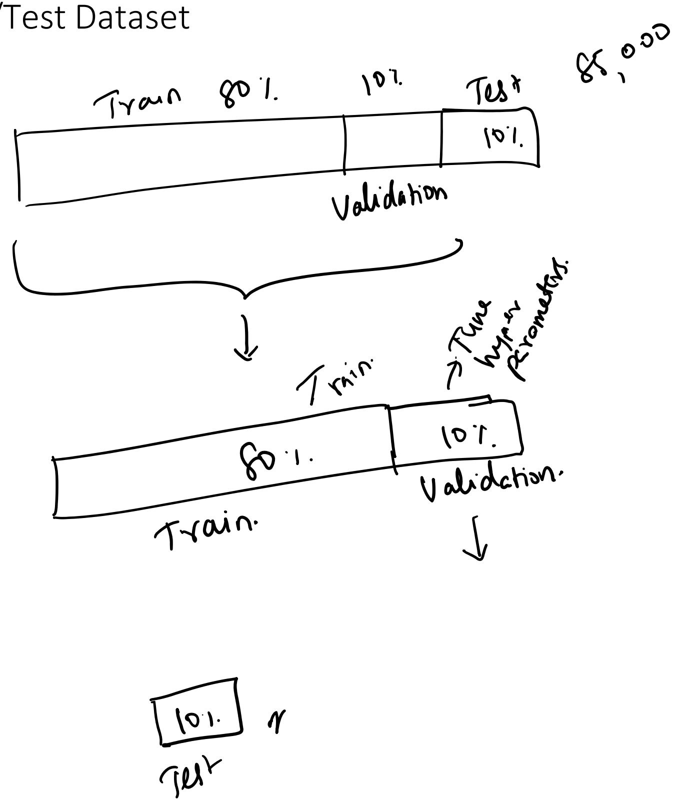
Calculate prior probability from Training data for each city..

State	1 (class-1)	0 (class-0)	P(1)	
Florida	1	2	1/3	= 0.33
Chicago	1	3	1/4	= 0.25
California	2	1	2/3	= 0.66

5 Class → AAI, AAC, INQ, RFA, REV ⇒ P 4 column.

Train/Validation/Test Dataset

Training



Testing

KNN

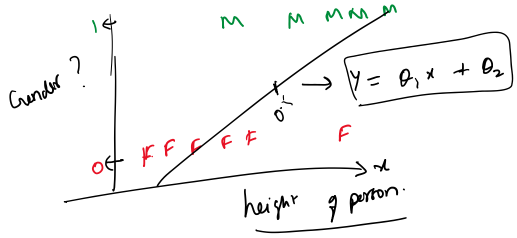
K = where the accuracy high.

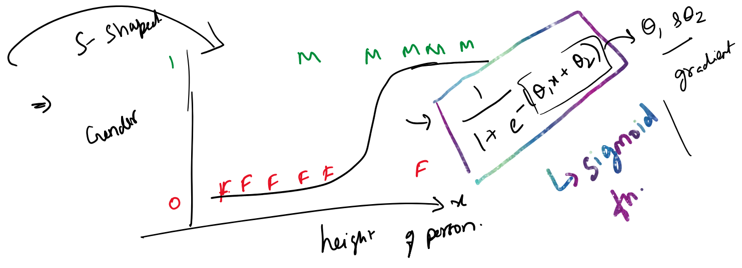
K = 5

K = 3, 4, 5, 6 ...

ACC = 0.7 0.72 0.78 0.6

Logistic Regression



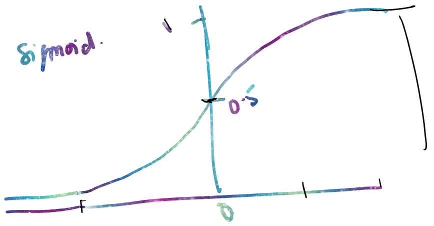


Issue

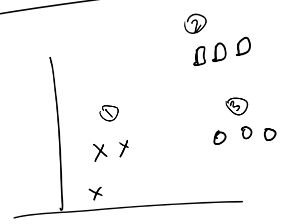
① Value of Y can be > 1 , < 0

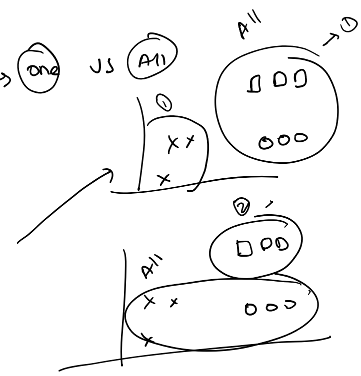
② Difficult to set a threshold

③ Can't deal with more than 2class



Multi class classification → One vs All





Classification Metrics

How are we going to measure the performance of our Model.

1) Confusion Matrix.

Binary Classification. → Spam (VS) Non Spam.
(10 above Spam, 90 above Non Spam)

negative : Nonspam
positive = Spam.

Predicted \ Actual	Spam	Non Spam	
Spam	8	10	18
Non Spam	2	80	82
	10	90	

Predicted \ Actual	Spam	Non spam
Spam	TP	FP
Non Spam	FN	TN

TP = True positive.

FP = false positive

FN = False Negative.

TN = True Negative

Actual 10 | 90

2 Accuracy

$$\frac{\text{No. of corrected prediction.}}{\text{Total no. of record.}} = \frac{88}{100} = 88\%$$

$$= \frac{90}{100} = 90\%$$

Not suitable for Imbalanced data set.

Predicted \ Actual	Spam	Non-spam
Spam	0	0
Non-Spam	10	90

3. Precision → How precise is the prediction.

$$= \frac{TP}{TP+FP} = \frac{8}{8+10} = 0.44$$

Predicted \ Actual	1	0
1	TP	
0	FN	

4. Recall → Fraction of class that is predicted correctly

$$\text{Recall (Spam)} = \frac{TP}{TP+FN} = \frac{8}{8+2} = 0.8$$

5. F1 Score

$$F1 = \frac{2 \times P \times R}{(P+R)} = \frac{2 \times 0.8 \times 0.44}{(0.8 + 0.44)} = 0.56$$

Multi Class Classification

Predicted \ Actual	Spam	Promotion	Social
Spam	4	6	3
Promotional	1	2	0
	1	2	3

$$P_{spam} = \frac{4}{4+6+3} = 0.3 = 4$$

$$\text{Recall}_{spam} = \frac{4}{4+1+1} = 0.66 = 4$$

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

predict			
Promoti	1	2	0
social	1	2	3

spam

$4+1+1$

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

$$F1_{spam} = \frac{2\times 0.3 \times 0.66}{0.3+0.66} = 0.4125$$

11^{th}

$$F1_{promotional} = 0.3$$

$$F_1 \text{ Social} = 0.66$$

F1 MACRO

$$F1 \text{ MACRO} = \frac{F_1 \text{ spam} + F_1 \text{ Promotion} + F_1 \text{ social}}{3}$$

$$= \frac{0.4125 + 0.3 + 0.66}{3} = 0.4575$$

Weighted F_1

W spam W pro W socia

No. of prom

$$= \frac{(6\times F_1 \text{ spam} + 10\times f_1 \text{ promotional} + 9\ F_1 \text{ social})}{25}$$

No. of spam

25 (No. of spam + promot + No. of social)

=

$$F1 \text{ Micro} = \text{Accuracy}$$

$$= \frac{4+2+3}{4+6+3+1+2+0+1+2+3}$$

=