Model Performance Analysis

Machine Learning

- **Classification**
  - ↳ Yes / No
    - Diabetes
    - No Diabetes
- **Regression**
  - ↳ continuous value (House Price Prediction)

Classification Problem

↳ **Confusion Metrics**

Spam ↓ / Actual →	1	0 (Not spam — Important mail)
1 (spam)	TP	FP
0 (not spam)	FN	TN

Predicted ↳

Dataset (Diabetes Prediction)

x_1 x_2 x_3 x_4 - - - x_{10} — I/p features

Predicted $\hat{y}$	Actual y
1	1
0	0
1	0
0	1

↳ Accuracy = $\frac{TP + TN}{TP + FP + FN + TN}$

Evaluation Metrics

Imbalanced Data
↳ 100 Records

Healthcare
- 90 Records — 0
- 10 Records — 1

$\Rightarrow$ 90%

Accuracy is not at all a good measure

↳ Precision = $\frac{TP}{TP + FP}$ (Predicted Data) $\longrightarrow$ Spam Classification

FP ↓↓

↳ Recall = $\frac{TP}{TP + FN}$ (Actual Data) $\longrightarrow$ Diabetes Prediction

FN ↓↓

Healthcare Dataset
↳ classification task

↳ f1 score = $\frac{Precision * Recall}{Precision + Recall}$