Topic: Logistic Regression

Name of Presenter: L. Srinivasa Varadharajan

Date of Class/Presentation: 18/01/2022

Learning Outcomes

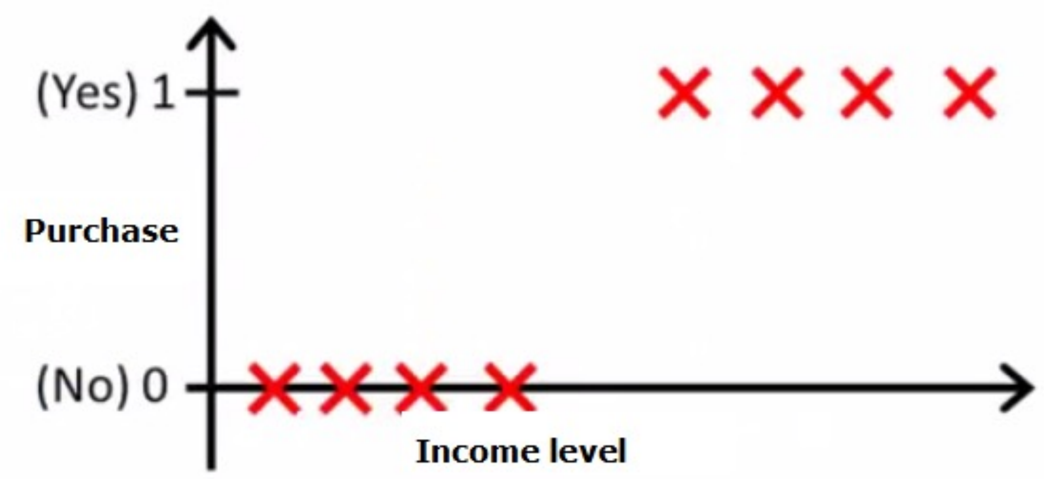
*When revising the material, keep in mind that if you can **confidently** and **fluently** answer the below, you have understood everything that needs to be understood from today's session – these are the expected outcomes from your learning today. First revise (material and videos of the class) and then ask questions.*

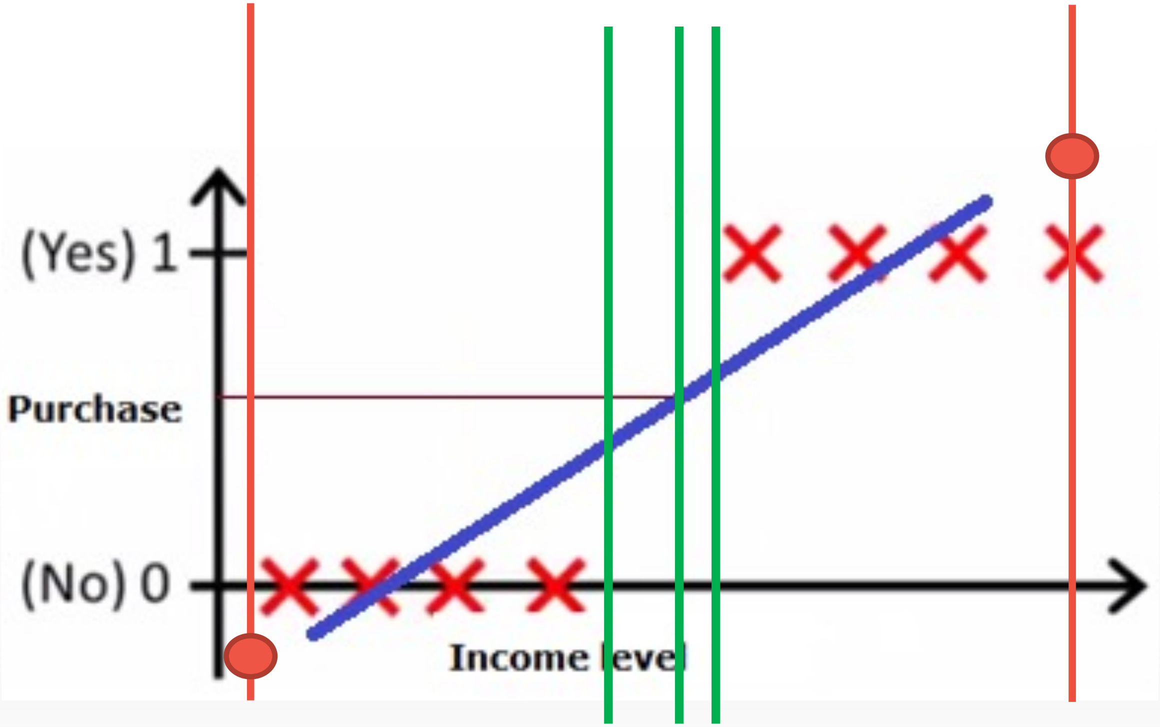
- Explain the terms odds, logit, sigmoid and binary classification
- Build a model for categorizing data in to one of two categories using logarithm of odds
- Describe how to use Deviance to estimate the goodness of classification
- Define the various metrics that are used to study the goodness of classification

Binary Classification

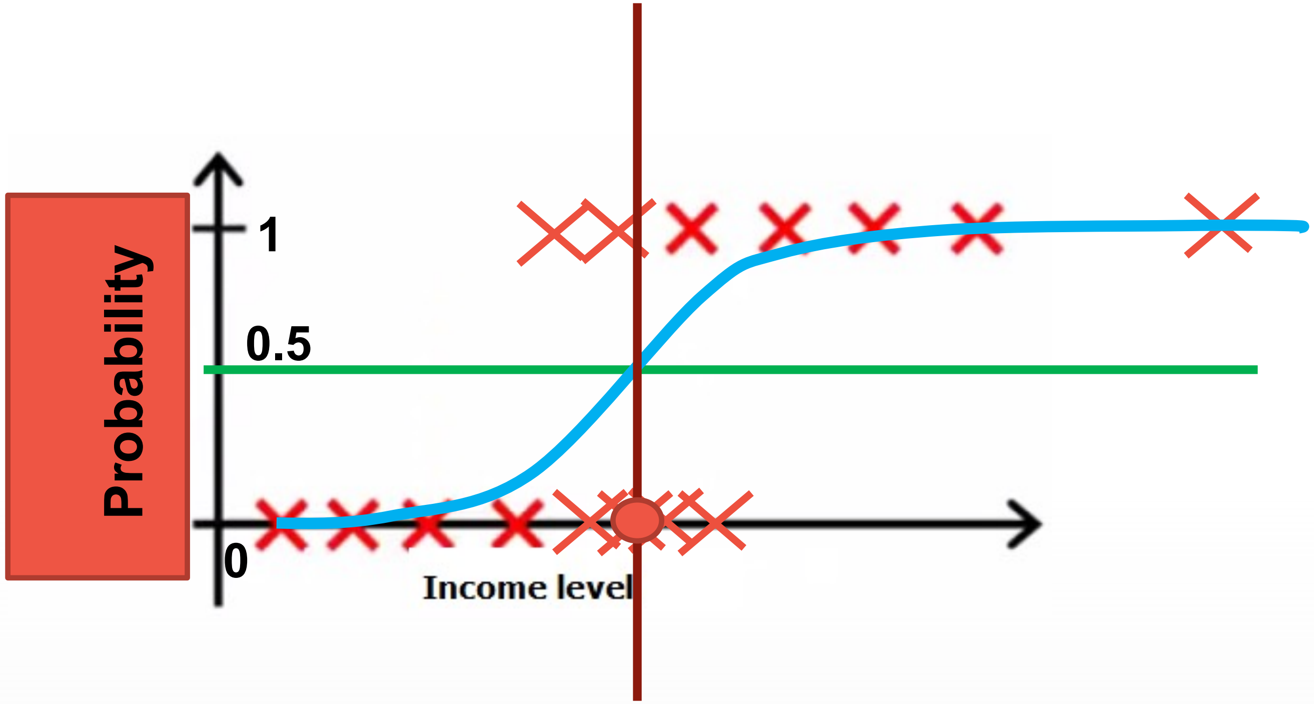
- Outcome variable is either 0 or 1
- Regressor variables are numerical or categorical

Logistic Regression – Binary Classification



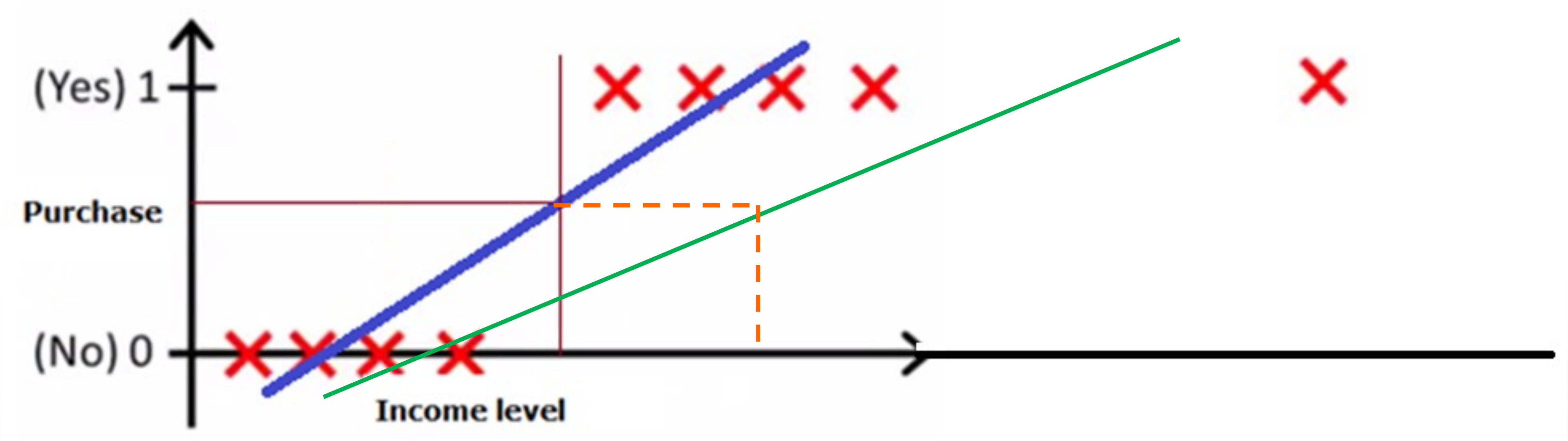


Logistic Regression – Binary Classification

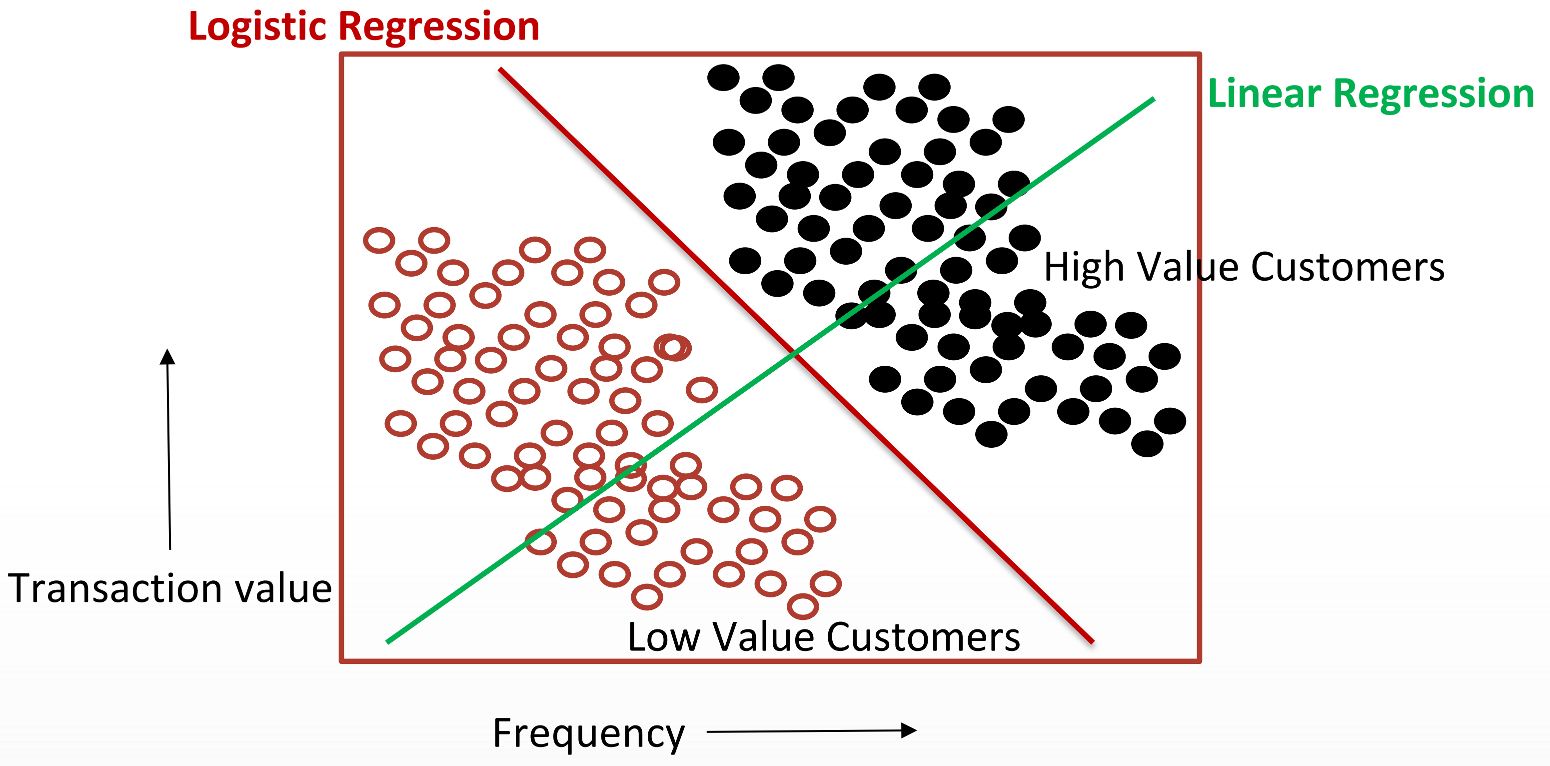


Income	Total Number	Number of people who said Yes	Probability of saying Yes
1000	35	0	0/35 = 0
5000	27	0	0/27 = 0
10000	44	12	12/44 = 0.27
20000	33	20	20/33 = 0.60
50000	17	15	15/17 = 0.88
100000	10	10	10/10 = 1

Logistic Regression – Binary Classification



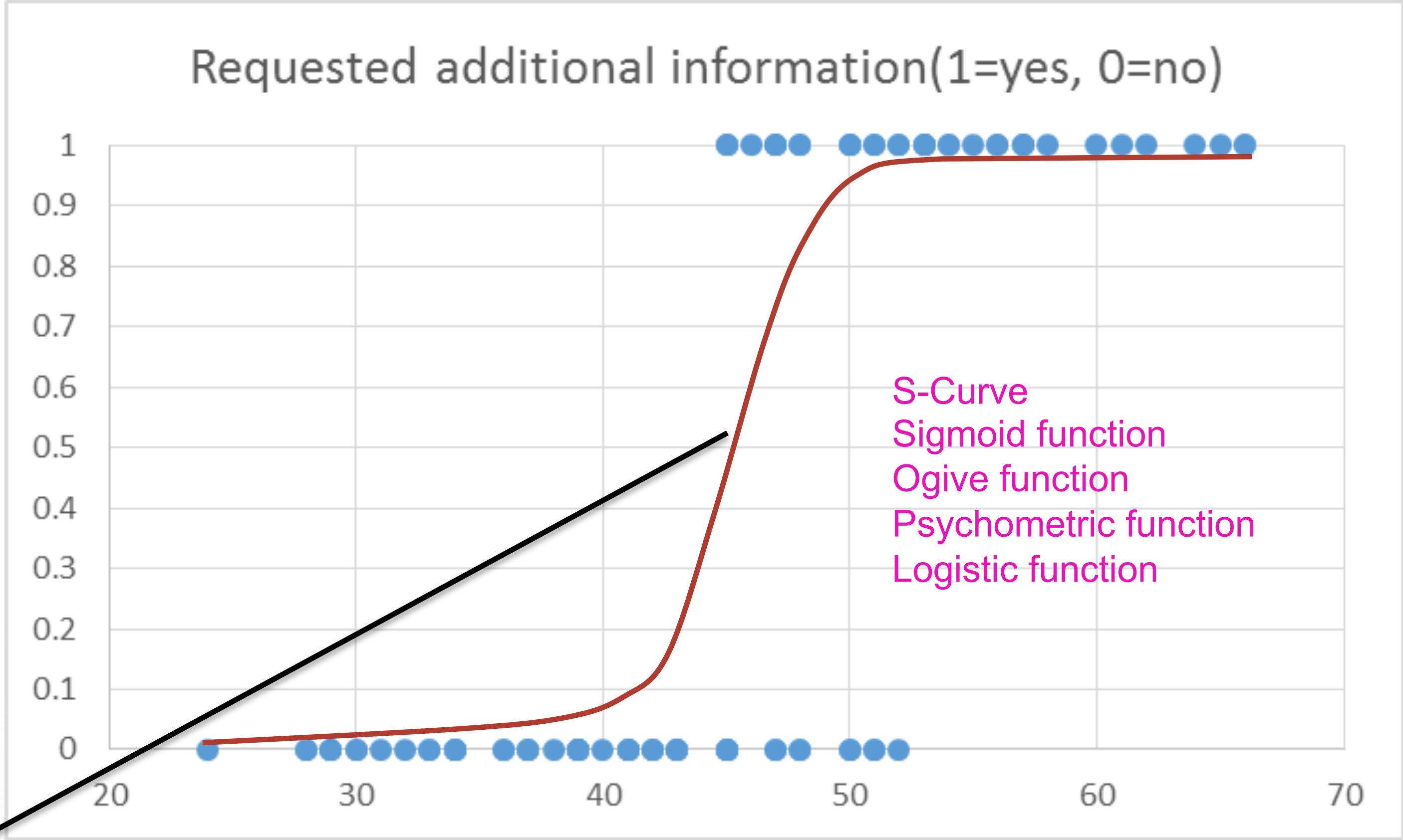
Logistic Regression – Binary Classification



Auto club mailer

An auto club mails a flier to its members offering to send more information regarding a supplemental health insurance plan if the member returns a brief enclosed form.

Can a model be built to predict if a member will return the form or not?



$$f(x) = p = \frac{1}{1+e^{-\mu}} = \frac{e^{\mu}}{1+e^{\mu}}$$

where $\mu = \beta_0 + \beta_1 x_1$ (also known as the systematic or the structural component or linear predictor).

Logistic – Sigmoid – Ogive – Psychometric function

$$f(x) = p = \frac{1}{1 + e^{-\mu}} = \frac{e^{\mu}}{1 + e^{\mu}} = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n}}$$

Odds

$$S = Odds = \frac{p}{1 - p}$$

$$S(1 - p) = p \qquad S - Sp = p$$

$$S = Sp + p \qquad S = (1 + S)p$$

$$p = \frac{S}{1 + S}$$

Odds – Test your understanding

If the probability of winning is 6/12, what are the odds of winning?	1:1 (Note, the probability of losing also is 6/12)
If the odds of winning are 13:2, what is the probability of winning?	13/15
If the odds of winning are 3:8, what is the probability of losing?	8/11
If the probability of losing is 6/8, what are the odds of winning?	2:6 or 1:3

Odds – Test your understanding

Probability (p)	Odds (s)
0	0
0.10	1/9
0.25	1/3
0.5	**1**
0.75	3
0.9	9
0.95	19
1	∞

Odds

$$S = Odds \; = \frac{p}{1-p}$$

$$S = \frac{\dfrac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}}}{1-\dfrac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}}} = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}$$

$$log(S) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

Solve for beta-s

Logit function

Odds – Test your understanding

Probability (p)	**Odds (S)**	**log(S)**
0	0	undefined
0.10	1/9	-2.19
0.25	1/3	-1.09
0.5	**1**	**0**
0.75	3	1.09
0.9	9	2.19
0.95	19	2.94
1	∞	inf

Odds

$$S = Odds = \frac{p}{1-p} \qquad S = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}$$

$$log(S) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

$$log(S) = 0 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$ **Boundary**

$$0 = \beta_0 + \beta_1 x_1 \Rightarrow x_1 = -\frac{\beta_0}{\beta_1}$$

$$0 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \Rightarrow x_2 = -\frac{\beta_0}{\beta_2} - \frac{\beta_1}{\beta_2} x_1 \Rightarrow y = mx + c$$

log Likelihood function

- In Linear Regression, we minimized

$$SSE = \sum (y_i - \hat{y}_i)^2$$

- In Logistic Regression, we maximize log likelihood instead

$$\log(likelihood) = \sum_{i=1}^{n} [y_i \ln \hat{y}_i \; + \; (1 - y_i) \ln(1 - \hat{y}_i)]$$

Cost Function

y_i: $\mathrm{p_i}$: probability of the ith point to be in the category 1

$\hat{y}_i$: predicted probability of the ith point to be in the category 1

$$Get\ \hat{y}_i\ from\ log(S) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

Auto club mailer

Demo in Excel

$$log(S) = -20.41 + 0.43 * Age$$

$$log(Likelihood) = -24.97$$

$$ResidualDeviance = -2 * log(Likelihood) = 49.94$$ $With\ the\ model$

$$Null\ Deviance = 123.16$$ $Without\ the\ model$

$$PseudoR^2 = 1 - \frac{Residual\ Deviance}{Null\ Deviance} = 0.61$$ $McFadden's\ R - squared$

Auto club mailer

$$log(S) = -20.41 + 0.43 * Age$$

$Age = 50$ Will this person respond or not?

$$log(S) = -20.41 + 0.43 * 50 = 0.89$$

$$S = exp(0.89) \ = 2.43$$

$$p = \frac{S}{1+S} = 0.71$$ Will respond

Auto club mailer

$$log(S) = -20.41 + 0.43 * Age$$

Classification boundary

$p = 0.5$ $\qquad$ $S = 1$

$$log(S) = 0$$

$$-20.41 + 0.43 * Age = 0$$

$$Age = 47.9$$

We classify everyone age 47.9 years or older as responder and others as non-responders

INSOFE's Vision

The BEST GLOBAL DESTINATION for individuals and organizations to learn and adopt disruptive technologies for solving business and society's challenges.

INSOFE - HYDERABAD
2nd Floor, Jyothi Imperial, Vamsiram Builders
Janardana Hills, Gachibowli
Hyderabad – 500032
+91 93199 77257

INSOFE - BENGALURU
Floors 1-3, L77, 15th Cross Road
Sector 6, HSR Layout
Bengaluru - 560102
+91 93199 77267

INSOFE - MUMBAI
4th Floor - A Wing, Spaces - Kanaki
Andheri-Kurla Road, Chakala
Andheri East, Mumbai - 400093
+91 93199 77269

Email: info@insofe.edu.in **Website:** www.insofe.edu.in

Follow us on Social Media: /insofeglobal/ /school/insofe/ /INSOFEedu /insofe_global/ /InsofeVideos