

# Decision Tree



# Course Overview

Term	CDF	GCD	GCDAI	PGPDSAI
Term 1	Data Analytics with Python	Data Analytics with Python	Data Analytics with Python	Data Analytics with Python
Term 2	Data Visualization Techniques	Data Visualization Techniques	Data Visualization Techniques	Data Visualization Techniques
Term 3	EDA & Data Storytelling	EDA & Data Storytelling	EDA & Data Storytelling	EDA & Data Storytelling
		Minor Project	Minor Project	Minor Project
Term 4		Machine Learning Foundation	Machine Learning Foundation	Machine Learning Foundation ✓
Term 5		Machine Learning Intermediate	Machine Learning Intermediate	Machine Learning Intermediate (in progress)
Term 6		Machine Learning Advanced (Mandatory)	Machine Learning Advanced (Mandatory)	Machine Learning Advanced (Mandatory)
		Data Visualization with Tableau (Elective - I)	Data Visualization with Tableau (Elective - I)	Data Visualization with Tableau (Elective - I)
		Data Analytics with R (Elective - II)	Data Analytics with R (Elective - II)	Data Analytics with R (Elective - II)
		Capstone Project	Capstone Project	Capstone Project
Term 7		Bonus: Industrial ML (ML – 4 & 5)	Basics of AI, TensorFlow, and Keras	Basics of AI, TensorFlow, and Keras
Term 8			Deep Learning Foundation	Deep Learning Foundation
Term 9			NPL – I/CV – I	CV – I
Term 10			NLP – II/CV – II	NLP – I
			Capstone Project	Capstone Project
Term 11				CV – II
Term 12				NLP – II
				NLP – III + CV – III
				AutoVision & AutoNLP
				Building AI product

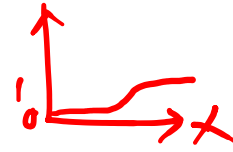
Done: Linear Reg



SLR vs MLR  
 $y = mx + c$

Metrics: ME, MAE, MSE, RMSE, MAPE,  $R^2$ ,  $Adj R^2$

Logistic Reg



$\sigma = [0, 1]$   
 S-shaped  
 log-odd func.

$$\sigma = \frac{1}{1 + e^{-y}} = \frac{1}{1 + e^{-(mx+c)}}$$

Supervised  
 ML Algos  
 (Classification & Regression)  
 One Tree

## Term Context

	Actual	Pred	
FP: Type I error	No	Yes	eg: Security chk at Airport
FN: Type II error	Yes	No	eg: RTPCR

• Decision Tree

• Random Forest (building multiple Trees, default=100)

• Principal Component Analysis

• Naïve Bayes Classifier (Sup. ML)

Math heavy  
 You are here...

Metrics: Confusion Matrix,

Accuracy,

Precision (Min. of FP),

Recall (Min. of FN),

Sensitivity

Specificity

AUC } AUROC  
 ROC }

$F_1$  /  $F_\beta$  score

Classification Report

Unsupervised ML (Feature / Dimensionality Reduction)  
 (only X, no y)  $X_{4000} \xrightarrow{PCA} X_{500}$

# Agenda

- Terminology Related to Trees
- Decision Tree
- Decision Tree Algorithms
- Attribute Selection Measures
- ID3 Algorithm → *Classification*
- Entropy & Information Gain
- Steps to Estimate Entropy & Information Gain
- CART Algorithm
- Gini Index
- Steps to estimate Gini Index
- CART – Regression Example
- Issues with Decision Trees
- Tree Pruning
- Decision Tree Applications



Thanks for watching