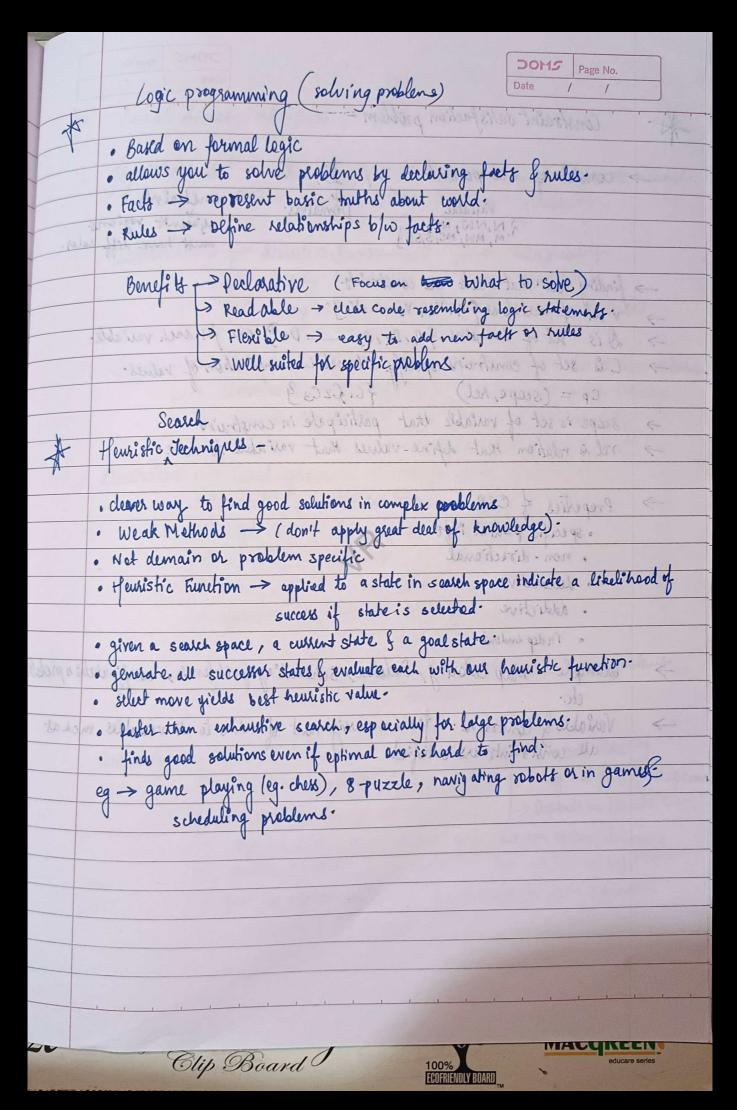
V-1 (Data Science Honoras). handling massive data Automoting tasks > Improving accuracy > Working 24/7 Innovation across fields chanced decision making Personalized experiences Trigumented human capablities. Addressing global thatleges engloration & discovery Applications -> Image & speech recognition > Recognitionmendation system
> Fraud Petection self driving cars > custome service > Health care > Mainfacturing Agriculture &



Local Search Jechniques -. Algorithms that iteratively explore solution space by making incremental changes to current solution, aiming to improve gradually. Characteristics > Afercitive Improvement > Adaptability > Greedy Approach > Scalability May not guaratee optimality > Hewistic based. Types -> Hill climbing Applications -> Joanel Saleman Broble > stmulated anthealing > Graph coloring. > Scheduling. elected - . data streams - high speed of dealt firming Limitations -> Local optima de stab la primaria. > sensitivity to intial solution. Difficulty in Escaping Local eptima. > limited Explorations. an get stuck in tocal optima emperated. < plants sueffectivenessed depends on problemation. Greedy search > simple & intuitive algo used to solve eptimization problems.

Hope of finding optimum. Characteristices > Local optimization dimitations > No Backtracking > Simpl Implementations > efficiency strongs strongs > Depends on Heuristras > simple sefficient > suboptimal sofre some > suboptimal sofre solutions > good for approx. sol? > problem dependent · bright -> occurate & metanoline. Olip Board

	DOMS Page No.
	Big Data Learning (V2) Date / / Honors Data Science
	My North Data Science.
	1 Hours 15 Million Way of many
*	Characteristics of Big data-
	I change to extent solutions aiming to improve gradually.
	big data → large vel of data available at various sources in varying
	degree of complexity, generated at different speed ie. velocities &
	big data > large vel. of data available at various sources in varying degree of complexity, generated at different speed ie. velocities & technologies, processing modules.
7	May not quarter extended & youther box
	volume > m'y data characterized > enourmous value > petapitity  large data needs specialized trels & technologies to store process  observe & tracks data from various sources.
uran field	· large data needs specialized tools & technologies to store process
- FILM	· observe & tracks data from various sources - analyze
	Commerce Hyprithms Chadulin
•	velocity -> . data streams -> high speed & dealt timing
	· processing of data > streamed data > real time results > turk
	· insight generated > returnt factionatel.
	· processing of data > streamed data > real time results > factionable.  · insights generated > returnent factionable.  · speed of generation of data.
	Printled Santolanski.
•	variety -> heterogeneous sources-
	marine of active -> smetrill a unstricted
	SOUTH MARKET LAND
*tonib	value → · business value → Bia Data.
	Manufact to sales
•	value → · businus value → Big Data.
100	· generate some out of
Marine approximately	entights > formal decisions and decisions
	Datimize and the
December of	> gain competative edge in marketplice.
- Jan	veractly > incomistance.
	veracity - in commistency / wneetainity in data  led & challeges in data quality & Reliability  · Insights -> accurate & trustworthy.
	ted & challenges in data quality & Reliability
	· Insights -> accurate & trustweethy.
790	

Date / /

Jypes of Data o feating patterns for unstructure Seni - Streeted Structure No predefined format. Some internal streeties predefined schema (like tables) · Fredict culput based on apatholine · disconers relación pu Log file, JSON, XNL. Databass, spredsheets. Jext documents, images, videos primo dendos solos · DOESNIT FICEING · Easier than unstructured · Requires specialized technique · Easiest (squ, statistics) · Analysis file system, clands toage Relational databases. file systems, storage cloud stolage. NO SQL delabases. spreadsheets. Also called clarification. · K-Mean chareing, recognition. configuration, log analysis deturned. Jext analytics, ing Transactions, reporting Vsu ML No (needs processing) Naybe (after processing) Yes (directly) Jable . statished technique used to extimate ret by a dependent visited ene of more independent variable. · Ideatiff general brend blus variables . Use model to predict dependent ray ble based on value of its independent

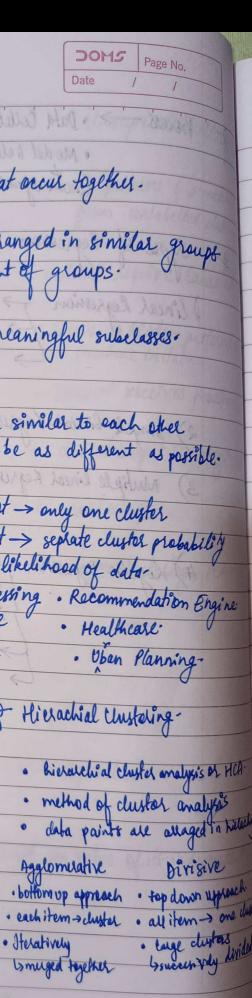
oup Doard

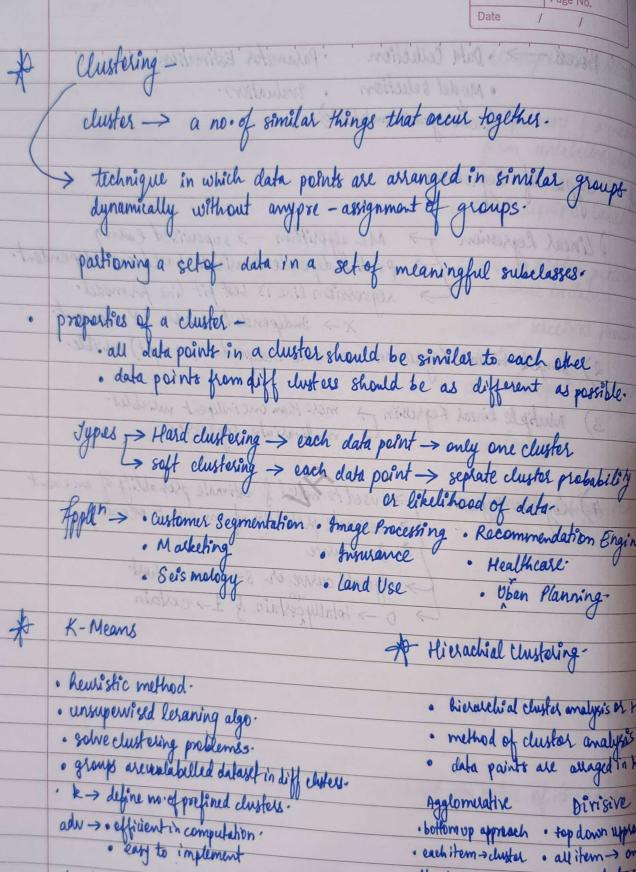
100% ECOFRIENDLY BOAR

* At	Process -> Data Collection Parameter Estimation	· AL
	· Model Selection · Evaluation.	1
	(Unas regression, muliple reto).	1
		)
*	Types of Legression.	1
- 0	The second secon	
	) lineal Regression -> ML algorithm -> supervised lowng	
	) Lineal Regression > ML algorithm -> supervised lorning  > predict dependent variable based on independent  > regression line is best fit line for model.	lent.
	> regression line is best fit line for model.	
	X -> Independent variable I -> OWPUT	3
	2) simple linear Legsussion - one independent (os input) variable	
- 93	. data paints from all dust as straid be as different as seeil	
	3) Mutiple lineal Legrusion -> more than one indepent valiables.	
	3) Mutiple lineal Legission - more than one indepent valiables.  multivasi ute regression.	
1.3	1 soft obushishing > wash date point -> septente courte proposit	160
	4) Rogistic Regression - used to model of estimate probability of an ever	nt .
an in	4) Rogistic Regression - used to model of estimate probablity of an event occurring	4 1
	- Pos acure	111
	> sigmed curve or s-curve in short	
	> 0 → totally costain & 1 → certain	1 10
	K-Means = 1 Histochial Uniferial	-
314.74	. Apartistic mathead historial of historial charles are along	
Signi	and solved to harden rethold be there and	
W 273	- some dustrial publimes	
	Inin alexandrelled defend it is followed in a gordon after north	
	. he ships no depetient shapes belones yound . to down	
-100	moliple. solution molidas. " medily mas de propilla sonale	
Letter Land	" Sport . purtingly	
	discours who seems about the man selfer who see I	
The second second		

· decorpt with noisy date.

or sicilable to the device that were





disadu- . Applicable only when man is defined

· I souble with noisy data.

· not suitable to discoverchasters -

· need to specify K.

Data d'imensionality no. of attributes or features vsed to represent each data point in a dataset. Jupe -> High -> many features relative to no of observations. Impact ->. curse of dimensionality Dimensionality feduction Seaml seletion · overfitting · Interpretability · Computational complexity Techniques > PCA (Principal Component halyers) - + distributed Stoshastic Neighbr Embedding > linear discriminant analysis > Anto encoders. Apply -> Image & Vide o Processing > Text Mining & NEPERSON TO DANG SONT TOPPUS. G Financial Modeling > Built for distributed computing on big data Spark Programming Model > Lazy evaluation of transformations of actions-F> Fault talluance through lineage info. > In memory processing for high performance. Disadu mona) - Min Benefits >> Complexity Steribility Scalability >> Resource Management Overhead >> Interactivity High Performance > Overhead in memory Processing -> Integration Fault Tellehench > Debugging & Monitoring > Inore approachable Ease of Use for programers Unified Framework