Assignment 3

	713=1911111111	
9.1	Compare decision to	esidosification owith
	K-NN clossification	eridosification Devith
\Rightarrow	Device Jan Lakes	
6	make a deal Ning	o'Decison
	noise Classification	minarothree
0	Does not require	1)22 Readize Hand Indiner
n alara (1 LSpecifico foriginal	zivatraja da politica
Ode	2) Paralogo topinings	2) 0 111
Ĺ	- Julian called liame alice	2) Builds a model based
1	a why called lazy algorithm	a) a maringson
6 mer	3) 12040 WAGE DIMONS DE	3) once tonining is completed
	available for making	Haraining data me not
	predictions.	: Méeded, isibara
	4) Most interpretable	Moderates interpretable
	algorithm.	since we leanser decision
20.	e e Highly indespeeded	making in Afree format.
	3) Takes much time in	3). Once Hamingis completed
000100	decision making a since	prodiction time is legs.
0.5	and have to straverse	no surt por door for
	the whole datased to	
. 2	it collected the distance.	10 ch same 7701 (2)
1	with each point.	to be marine middle
	6) Since no training is	6) Require Mithal training
habal :	regulared hence it	lation exists in add decides
		node and branches.
~^	7) Suitable For small size	7) Need Sein Sell Carrell
	tolatabaset lima	
	A CO.	ag large size dataset.
		2420400
	,	

Segression classification, sold sold and all logistic Regisersion logistic Regression Logistic regression is 1) A decision trèe is a learning algorithm that is commonly used used for classification in mel tall model and bising Hasks: where the ad spredictionationes based goalis to model prom inputudatas predict probability that .2001-10:home Olgon letters given class or not 2) Coefficients com be 2) Highly interpreted, as his a mos sinterpreted to under- thie trees structure 1051 istand the influence ocleanly shows the deciden of each feature on a mode out weach mode the outcomes. Lossion slowy and 3) Less prone to over 3) Brine to sovorfitting. fitting comported to (4) It is not majorly offected 4) Italis majorly affected 29 passby bnoises by noise. 110 5) 11 Requires a large is 5) Can be dition of on . La senough e trainings small training ret. dataset.

Assignment 3



9.3 List down the attribute is election measures jused by TD3 algorithm to Tookshoult a decision tree is lot light in (27: 1. A HAMbute (TSelection Measures line 1) hini Index 2) Information Painos nothermodate (3 3) = Entropy 110 bother sit TE 4) Gain . Ratrophoo ad 1-de bampoon · Lariupan nothermodul 20 hours att? 1) dans Tradex organ mobiles of is atingle All attributes rare lassumed to be continous valued. It is assumed that there exist general possible split values for teach Gini Index method en be modified For nategorical attributes. + Gini index is used in Classification and Regression of Tree (CART). radaid all interest Esta dataset To contains Example from n classes, gini index, ghilT) is defined arshowing the second of the seco of an Alogala despoint gut as plinichesons In the above equation I represents the relative frequency of class jagnat.



After scollitting To into thousubsets Til index of split data is to noising gini: 30(TD) our Mottginberti) ottuder gini (TE) 2) Information aginal IG) sometime (=) In this method all arttributes are assumed to be categorical. 100 10 The amount of information required to decide if a random record in is belongs suprimos to A bom Wais de fined togicipto 11A & It is assumed that I(ppn) ottom (ppn) togis (ppn) /pnove bordiber set up Boodsalogx (ho) his Political in the Political o companies used in Classification of The higher the Tage the better the a attribute in its attractifying the texaple sailab of (17) into , wabit inip i zassola n Entropy measures this impurity or disorder in the dataset. It is a measure of the uncertainity in the dataset's classification. Enthopy (s) = = 5 171/092 Pi





· data is more in party of a set of The higher outhe entropy it the more uncertain 4) Gain Ration => Rain ingto (15: on extension inf IG that dakes into account the intringic information of a split , penalizing addribute that with smallproperty offedata, bataglas e the civi index ranges from a to 0.5 10 Cain Ratio(A) Frankation (Rain (SIA) 12010 alerto a de enois Splita Totomation (A) D. 4 : Explain the properties to f aini Index The Gini index is a measure used in decision trees to evaluate quality of the aplit.

It quantifies the degree of impurity or disorder in a dataset. The aini is used in Classification and Regression Tree (CART). If a dataset T contains example from A classes, gini index, gini(T) is defined as-Gini (T) = 1 - 2 (Pj)2



relative frequency of class is inst. To coith sizes bull and Nami, giri index of split data is in 1016 OF gini (prit) on the plan gini (TI) of the gini (TE) host abudishbo prisilonad , filga - 20 do The attributer withol smallest + ginight (T) is selected to applit the moderne die The aini index ranges from 0 to 0.B. (b) On dicater perfect (publity of meaning all the (in) inodances l'abelong to a single class. + 0.5 represents maximum impurity, indicating that the inotances are uniformigialistable ocross all classes. - The Gini index is a measure escal in decitilge att fo philosop aboutous of the aplit It quantities the degree of impunity and storages anistros T topolos a 3T Si (T) inip, xabri inip, 29220





0 0.3 Dictues in brief pruning in decision tree A sach esta porcentlo hard sentonal sint? (= Principle in a decición betreellie a etechnique used to reduce the size of the adecision tree by removing sections of tree that provide 11+4/e: +60 no additional predictive The main good of prining mis to prevent overfittings and improve models ability to generalizes to sunseen dators sactula theorision trees can grow very large and complex, capturing noises and loutliers in the Healing data, which leads to over fitting. - Bruning helper this mitigate by simplifying the Hopen hart savanced anto savanos (= By "removing unnecessary branchers i pruning reduces the complexity of the model, leading to better performance on new, unseen there are two main types of prining in decision tree !-1) Bre-prunning. This involves halding the growth of the tree, before it becomes too complex. It sets a limit on tree depth, maximum gain required for a aplit.

ماع ده	Distoct - prunning ra Jaird ai source a. D
	=) This Involves fixed allowing the tree to
andia.	grow hilly and theres pounting estack unnere-
10 (80 00)	esant branches. + + subse de base
host	sasta do anoitanon poíviemen ud sasta
7~13	esets to socitions operates added a society
	to the opining of eckniques! arebitand
0	= cost complexity prunning non ant
	subtree primarily-spaced on the accuracy
	and complexity inthem selects the subtree
	1000 ithis the endowned by feet sion printings
L	data, estit leade to even fitting.
adt on	192): 10 Reducted & Bornor Minninglad minust
-	= Removes the boranches that do not
pola	Significantly affect the overall accuracy.
اعدام	Removes the branches that do not coursely accuracy.
اعدام	Removes the branches that do not not significantly proffee et wither overall vaccuracy. 12bara 3 th 2 dividance at 200 base.
اعدام	Removes the branches that do not not significantly proffee et wither overall vaccuracy. 12bara 3 th 2 dividance at 200 base.
اعدام	Removes the branches that do not consider the ighnificantly maffect with en overall vacuumocy. 12 born 3 dt 3 dividences at 2000 box 200 10.00 paget sion 180 to 300 and 100 to
اعدام	Removes the branches that do not not significantly proffee et wither overall vaccuracy. 12bara 3 th 2 dividance at 200 base.
اعدام	Removes the branches that do not configurate interest vacuumocy. Isbara att an divisiones att assubase. Tally con a service of the service
Point - Point	Removes the bounches that do not consider the ighther overall vaccuracy. Isbam at an dividence at souther overall vaccuracy. Isbam at an dividence at souther overall vaccuracy. Isbam at an dividence at souther overall vaccuracy. Isbam at an allow the souther overall vaccuracy. Isbam at an allow the souther overall vaccuracy. Isbam at an allow the souther overall vaccuracy.
point of the point	Removes the branches that do not consider the significantly proffee of wither overall faculary. 12 born at 2 dividence at 200 born. 12 born at 2 dividence at 200 born. 12 born at 2 dividence at 200 born. 12 born at 200 born. 13 born at 200 born. 14 do thouse at enthold 200 born. 14 do thouse at enthold 200 born.
point of the point	Removes the branches that do not the ighnificantly professionally professionerally professi
point of the series of the ser	Removes the branches that do not consider the significantly proffee of wither overall faculary. 12 born at 2 dividence at 200 born. 12 born at 2 dividence at 200 born. 12 born at 2 dividence at 200 born. 12 born at 200 born. 13 born at 200 born. 14 do thouse at enthold 200 born. 14 do thouse at enthold 200 born.
point of the series of the ser	Removes the branches that do not the ighthifteantly proffer ethytherioverally accuracy. Isbara att 3/ Hiralamo att 200 bar. 1200 10.20 10 and Acoto a co and to a company of the company
point of the series of the ser	Removes the boundes that do not one ighnificantly parthering and souther overall facturacy. Isbam att In dividence and souther or souther. Isbam att In dividence and souther of the control of the con