ino any fea The	2.Its very easier to im 3.Fewer feature leads 4.Feature selection re 5.Training time for a r 6.Models with the few eat is Filter Method er method is belonging	aplement the machine is to enhance the gener educes data redudancy model with fewer featurer features are less pod ?  g to the feature selections	ures is significantly lower. rone to errors. on method that select fea	educed features.	
Ur In t	volve any machine lear machine learning modure selection pipeline re are two kind of feat  1.Univariate 2.Multi-Variant  iivariant ne univariate we are se	rning algorithm that's del. And another adva so this kind of best sc ture seection method i	it is a pretty fast. Feature ntages of the fiter method reening method to get su in filtering method those	selected by using with the filted is they are very fast. Filter meabset of the features.  The are the seatures. Thats it kinds	r method can be use as an in ethod is a generally first step
fea The fea info  M In i tho	ure. In the univariant for univariate filter method ures are then selected remation and variance ulti-Variant.  In multi-variate feature see subset features and tivariant filter method	feature selection we can ods are types of the mand. Diffrent type of the rand of the feature.  Selection method we will method from which we	an select the features base ethod where individual fe anking criteria are used fo take information of featur ve select known as "pears	s without taking the information of the Fisher score, Mutual I atures are ranked according to refer the univariat filter methods. For the detailed on correletion on correletion on the data since they mutual	nformation Gain etc.  The specific criteria. The topor an example: Fisher Score,  of find out the best feature out
Ur • • •	Constant feature Rem Quasi Constant Remo Duplicate Feature Rem Instant Feature  Import the all like Import numpy as np	noval oval moval <b>Removal</b>			
i i :	mport pandas as pomport matplotlib.pmport seaborn as smatplotlib inline  com sklearn.model_ com sklearn.metric com sklearn.ensemb com sklearn.featur  load the datasets a_rows for the sim	selection import es import accuracy ele import RandomF ee_selection impor	_score orestClassifier <b>t</b> VarianceThreshold		
# # # d d 0 1 2 3	we are using the resince, tree based means standardizate tata.head()  ID var3 var15 imp_  1 2 23  3 2 34  4 2 23  8 2 37	ent_var16_ult1 imp_op  0.0  0.0  0.0  0.0  0.0	sifier :- it does no ider scales of vario tion not required in  o_var39_comer_ult1 imp_op  0.0  0.0  0.0  195.0	the tree based model.  p_var39_comer_ult3 imp_op_var4  0.0  0.0  0.0  195.0	40_comer_ult1 imp_op_var40_c 0.0 0.0 0.0 0.0 0.0
x y x	10 2 39  ws × 371 columns  = data.drop('TARGedata['TARGET'] shape, y.shape  2000, 370), (2000,  er we have to apply the id the overfitting.	,))	o the training and testing	0.0 we have to apply the some op	0.0 eration on the test dataset pa
# # c c	nstant Features Renere we have to fithresold variance onstant_filter = Vonstant_filter.fit rianceThreshold(the	emoval  irst use the varia  is always work on  VarianceThreshold( c(x_train)  hreshold=0)	nce thresold ,fit an the training datase threshold=0)	st_size=.2,train_size=.8  d then use transform.  ts.  after removing the cons	
# 26 # C C # [F F F	if invert the above postant_list = [no postant_list]	on_constant featur  ve code we will ge  ot temp for temp i	t the constant feature of the constant filter.ge	t_support()]	
F F F F F F F F F F F F F F F F F F F	alse,				
TTFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	alse, alse, alse, rue, rue, alse,				
F F T T F F F T T	alse, alse, alse, alse, rue, rue, alse, alse, alse, alse, alse, alse, alse, alse, alse,				
	alse, rue, rue,				
TFFFFFFFTTT	rue, rue, alse, alse, alse, alse, alse, alse, alse, alse, arue, rue, rue,				
FFFFFFFFTTFFFTT	alse, alse, alse, alse, alse, alse, alse, alse, alse, rue, alse, alse, alse, alse, alse, alse, alse, alse,				
	alse,				
FTTFFFFTTFF	alse, alse, rue, rue, alse, alse, alse, alse, alse, alse, alse, rue, rue, alse, alse,				
	alse,				
FFFTTTTFFF	alse, alse, alse, alse, alse, rue, rue, rue, alse, alse, alse, alse,				
FFFFTTTFFFFF	alse, alse, alse, alse, alse, rue, rue, rue, alse, alse, alse, alse, alse,				
F F F F F T T F F	alse, alse, alse, rue, alse, alse, alse, alse, alse, alse, alse, rue, rue, alse,				
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	alse, alse, alse, alse, alse, rue, alse,				
TFFFFTFFFFFFFF	rue, rue, alse, alse, alse, rue, rue, alse, alse, rue, alse, alse, alse, alse,				
FTTFFTTFTTFT	rue, alse, rue, alse, alse, alse, alse, rue, alse, alse, rue, rue, rue, rue, alse, rue,				
TFFFTFFTTFTT	rue, rue, alse, alse, alse, rue, alse, alse, rue, rue, rue, rue, rue, rue,				
TETTTETTETT	rue, alse, rue, rue, rue, alse, alse, alse, rue, alse, rue, alse, rue, rue,				
FFTFTTTTFFT	alse, alse, alse, rue, alse, rue, rue, rue, alse, alse, alse, alse, rue,				
TTTTTTFFFFF	rue, rue, rue, rue, rue, rue, rue, alse, rue, alse, alse, alse, alse,				
FFFTTFFTFTTT	alse, alse, alse, alse, rue, rue, alse, alse, rue, alse, rue, rue, alse, rue,				
F F F F F F F F F F F F F F F F F F F	alse,				
FTFFFFFFFF	alse, rue, alse, alse, rue, alse,				
FFFFFFTTTTT	alse, alse, alse, alse, alse, alse, alse, rue, alse, rue, rue, rue, rue, rue, rue,				
FFFTFTTTTTT	alse, alse, alse, rue, alse, rue, rue, rue, rue, rue, rue, rue, ru				
FFFFFFFFFF	alse,				
FFFFFFFFFF	alse, arue,				
TFFTTFTTT	rue, rue, alse, rue, alse, alse, rue, rue, alse, rue, rue, rue, rue, rue,				
F F F F X	'ind_var18_0' 'ind_var27', 	_list] , 'ind_var2', 'ind ', 'ind_var18', 'i		r28_0', 'ind_var28',	
# * * *	'saldo_medio_ 'saldo_medio_ 'saldo_medio_ 'saldo_medio_ dtype='object'  now lets go ahead remove all the con to avoid the overf train_filter = con test_filter = con so,lets check about	var13_medio_ult3' var29_hace2', 'sa var33_hace2', 'sa var33_ult1', 'sal ', length=109)  transform the x_t nstant feature spa  fitting we had use onstant_filter.tran at the features av	ce.  d the the transform  nsform(x_train)  sform(x_test)  alibilities in the x	_hace3', 3', 3', ],  on-coonstant feature spa  method #we appllied on t  _train_filter,y_tain_fil	raing dataset as well o
# # (()	train_filter.shap  now can see that to  whereas, the previous  1600, 261), (400,  nasi Constant Form  asi constant menas:-	there is a constantious traing datase  261), (1600, 370)  eature Remova	.shape,x_train.shape  t feature are remove  t had 370 features o		test datasets by using features are removed co
Dut The /ar set A fo sor	out and such are not re is no rule as to wah ance thresold is that methey are 99% similar. The salway nehow from the mean nehow near to the con	really usefull for making the thresologies the sologies of similarity as content a single valuation that means variance wastant.	g predictions .This feature ld we select for the quasi . that means if we select t ue which is almost constan	es create unnecessarily overhead constant feature but in general he 0.01 value of the variance t int but there is very a small chan k kind feature we say as quasi o	ad on the machine learning n I and widely we use the 0.01 hat time we remove the all fe nce that feature is shifted var
#	uasi_constant_filt uasi_constant_filt rianceThreshold(th	cer = VarianceThre	shold(threshold=0.01		
q Va q	ure 261 to the x_train_	_filter i.e.261-223 = 38  er = quasi_constant er = quasi_constant er.shape , x_test  223))	-	test_filter)	
q q q q q q q q q q q q q q q q q q q	test_quasi_filter train_quasi_filte 1600, 223), (400, s is the way we had rer		they are duplicate.That me	•	del by increasing the training
qqq22222222222222222222222222222222222	test_quasi_filter train_quasi_filter 1600, 223), (400, is the way we had rer plicate Feature the two features are exact it increases the overhead overall performance we oblem is that the sklern aspose technique, so less	es Removal acly same that means the ead of the machine lead would be reduce in case or python doesn't have the see the place of the column aguasi_filter.T	arning model.In that way is if our datasets having some any liberary to remove	duplicate of the feature datase	ts.50 for that we have to use
qq	test_quasi_filter train_quasi_filter train_quasi_filte  1600, 223), (400,  is the way we had rer  plicate Feature the two features are exact it increases the overhel overall performance we to blem is that the sklern the spose technique, so lest train_T = x_train test_T = x_train test_T = x_test_  by using the trans type(x_train_T)  mpy.ndarray  so, we need to convect train_T = pd.Data test_T = pd.Data	es Removal acly same that means to ead of the machine lead would be reduce in case or python doesnt have the place of the column and according to the place of the column according to the place of the place of the column according to the place of the column according to the place of the column according to the place of the place of the column according to the column according t	arning model.In that way is a if our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format		ts.50 for that we have to use
q q 222 q f wax x (() ihi Di thi uut trocarara x x # # t t nuu # x x # # d d d 0 1 2 2 3 4 4 4 4 4 0 0 1 2 2 3 4 4 4 4 4 6 1 2 2 3 4 4 4 4 4 6 1 2 2 3 4 4 4 4 4 6 1 2 2 3 4 4 4 4 4 6 1 2 2 3 4 4 4 4 4 6 1 2 2 3 4 4 4 4 4 6 1 2 2 3 4 4 4 4 4 6 1 2 2 3 4 4 4 4 4 6 1 2 2 3 4 4 4 4 4 6 1 2 2 3 4 4 4 4 4 6 1 2 2 3 4 4 4 4 4 6 1 2 2 3 4 4 4 4 4 6 1 2 2 3 4 4 4 4 6 4 6 1 2 2 3 4 4 4 4 4 6 6 1 2 2 3 4 4 4 4 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6	test_quasi_filter train_quasi_filter train_quasi_filter 1600, 223), (400,  is the way we had rer  plicate Feature the two features are exact it increases the overhological performance we have been spose technique, so less that the sklern is pose technique, so less that the sklern is pose technique, so less that the sklern is pose-rows will get the train_T = x_train_test_T = x_test_  by using the transformation for the strain_T = pd. Data for we have the strain_T = pd. Data for we want to check train_T. shape, x how can see that the strain_T. shape, x how can see that the strain_T. duplicated_feature is the sis the way we deplicated_feature is plicated_feature.  False False False False	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an end of the earth and are the machine (x_train_T)  Frame (x_train_T)  Frame (x_train_T)  Extended the rows the column and the end of the	arning model. In that way is a if our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.		heck by using the dupl
q 22 wa a x x () ( hi ) thut the real ra x x # t nu # x x # d d 01234 212222222224 # # f	test_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter 1600, 223), (400,  is the way we had rer  plicate Feature the two features are examit increases the overher overall performance we have been spose technique, so less the spose technique, so less train_T = x_train_test  test_T = x_train_test  test_T = x_train_test  test_T = pd.Dataf  test_T = pd.Dataf test_	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is a if our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the duplicated if not then in
Q 22 f we a x x x () (ii)	test_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter 1600, 223), (400,  is the way we had rer  plicate Feature the two features are examited increases the overhead overall performance we have been shown to be the second to convert the second the second to convert the	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	theck by using the duplicated if not then it
Q 2 f ea x x x () hi	test_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_duasi_filter train_duasi_filter train_test_reature the two features are examit increases the overhow overall performance were to blem is that the sklern train_test_reature, train_test_reatu	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the duplated of not then it
V 2 fe a xxx () in D ft ut rom	test_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_test_reature the two features the way we had rer the two features are examit increases the overher to overall performance we have been state that the sklern train_test_reature in the sklern train_test_reature t	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the duplated of not then it
The state of the s	test_quasi_filter train_quasi_filter train_quasi_filter 1600, 223), (400,  is the way we had rer  plicate Feature the two features are examit increases the overher overall performance we specified the sklern train_test_test_now will get the train_test_test_now using the trains train_test_test_now using train_test_now using train_test_n	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the duplicated if not then in
V 2 fee xxx () in D that to m #xx #x# () #x 3 #dd 01234 22222E ##ff TITITITITITITITITITITITITITITITITITIT	test_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter 1600, 223), (400,  is the way we had rer  plicate Feature the two features are exa- it increases the overhow overall performance we to blem is that the sklern the spose technique, so left train_T = x_train test_T = x_test_ the spose technique, so left train_T = x_train test_T = x_test_ the spose train_T)  mpy.ndarray  so, we need to convent train_T = pd.Dataf tif we want to check train_T.shape , x- to check how many fifter train_T.duplicate  this is the way we train	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the duplicated if not then in
V 2 fee xxx () in    The contract of the contr	test_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter 1600, 223), (400,  is the way we had rer  plicate Feature the two features are exa- it increases the overhe overall performance we to blem is that the sklern spose technique, so let train_T = x_train test_T = x_train test_T = x_test_  by using the trans train_T = pd.Data test_T = pd.Data test_T = pd.Data if we want to check train_T.shape , x thow can see that the train_T.shape , x thow can see that the train_T.duplicate  if we want to check train_T.duplicate  false Talse False False False False Talse False False False Talse Talse Talse False False Talse	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the duplicated if not then in
V Q 2 fee xxx () hi D fluid rain xx #x# () #x 2 #dd 01234 222222 # #ff [TITITITITITITITITITITITITITITITITITITI	test_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_duasi_filter train_duasi_fi	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the duplar
V 2 fe xxx () in D that to the man with the strain of the contraction	test_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the duplar
O 2 fee xxx () in the first transfer of the fee fee fee fee for the fee first	test_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_duasi_filter train_text_record trai	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the dupl
O 2 fee xxx () in the first transfer of transfer of the first transfer of transfer o	test_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_decorpts the way we had rer the two features are exa- train_	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the duplicated if not then in
The state of the s	test quasi filter train the way we had rer the work features are exa to increase the worker to we features are exa to increase the overher to we features are exa to increase the overher train T = x train trest T = pd. Data firm man train T = pd. Data firm man train train train T = pd. Data firm man train train train T = pd. Data firm man train train train T = pd. Data firm man train train train T = pd. Data firm man train train train T = pd. Data firm man train train train T = pd. Data firm man train train T = pd. Da	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	heck by using the dupl
The state of the s	test quasi filter train the way we had rer train the way we had rer train the skiem train	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	heck by using the dupl
The state of the s	test quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_test way we had rer train_test train_test we can be considered to consider the considered train_test and tra	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	heck by using the dupl
The state of the s	test quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_test_rest_rest_rest_rest_rest_rest_rest	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the dupl
TO A TO A STATE OF THE STATE OF	test quasi filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_quasi_filter train_train_quasi_filter train_tr	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the duplar
The state of the s	rest quasi filter train_quasi, filter train_cuasi, filter train_cu	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the dupl
TO SEE THE TENEFTE	test quasi filter train_quasi, filter train_trai	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	arning model. In that way is eif our datasets having some any liberary to remove an and vice versa.  aframe become array.  dataframe format  and columns  umns and column become array.  cated in the datafraticated feature if the cated()	me rows  me for that we need to contain the second to the	Theck by using the duplicated if not then it
To be the second of the second	test quasifilet  train_quasi, (400,  train_quasi, (400,  train_train_quasi, (400,  train_t	es Removal  acly same that means to ead of the machine lead would be reduce in case or python doesnt have the see the place of the column and an accordance of the column and accordance of the column accordance	to_keep].T  to_keep].T  dataframe format  and columns  umns and column becon  cated in the datafra  icated feature if the  cated()  e non-duplicated feature if eature if the  cated in duplicated feature if the duplicated f	me rows  me for that we need to contain the second to the	theck by using the dup.