

```
4468. import numpy as np
from sklearn.metrics import accuracy_score, mean_squared_error, confusion_matrix, classification_report, auc, roc_auc_score
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier, GradientBoostingClassifier
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.feature_selection import exhaustive_feature_selector, SequentialFeatureSelector
from sklearn.model_selection import train_test_split, cross_val_score

In [470]: pd.pandas.set_option('display.max_columns',None)

In [471]: data= pd.read_csv('adult.csv')

In [472]: data.head()
```

	age	workclass	fnlwgt	education	education.num	marital.status	occupation	relationship	race	sex	capital.gain	capital.loss	hours.per.week
0	90	?	77053	HS-grad		9	Widowed	?	Not-in-family	White	Female	0	4356
1	82	Private	128270	HS-grad		9	Widowed	Exec-managerial	Not-in-family	White	Female	0	4356
2	66	?	186061	Some-college		10	Widowed	?	Unmarried	Black	Female	0	4356
3	54	Private	140359	7th-8th		4	Divorced	Machine-op-inspct	Unmarried	White	Female	0	3900
4	41	Private	264663	Some-college		10	Separated	Prof-specialty	Own-child	White	Female	0	3900

```
In [473]: data.tail()
```

	age	workclass	fnlwgt	education	education.num	marital.status	occupation	relationship	race	sex	capital.gain	capital.loss	hours.per.week
32556	22	Private	310152	Some-college		10	Never-married	Protective-serv	Not-in-family	White	Male	0	0
32557	27	Private	257302	Assoc-acdm		12	Married-civ-spouse	Tech-support	Wife	White	Female	0	0
32558	40	Private	154374	HS-grad		9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	0	0
32559	58	Private	151910	HS-grad		9	Widowed	Adm-clerical	Unmarried	White	Female	0	0
32560	22	Private	201490	HS-grad		9	Never-married	Adm-clerical	Own-child	White	Male	0	0

```
In [474]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32561 entries, 0 to 32560
Data columns (total 15 columns):
age                32561 non-null int64
workclass          32561 non-null object
fnlwgt             32561 non-null int64
education          32561 non-null object
education.num      32561 non-null int64
marital.status     32561 non-null object
occupation         32561 non-null object
relationship       32561 non-null object
race              32561 non-null object
sex               32561 non-null object
capital.gain       32561 non-null float64
capital.loss       32561 non-null float64
hours.per.week     32561 non-null int64
native.country     32561 non-null object
income            32561 non-null object
dtypes: int64(6), float64(2), object(7)
memory usage: 2.6+ MB
```

```
In [475]: data.isnull().sum()
```

```
age                0
workclass          0
fnlwgt             0
education          0
education.num      0
marital.status     0
occupation         0
relationship       0
race              0
sex               0
capital.gain       0
capital.loss       0
hours.per.week     0
native.country     0
income            0
dtype: object
```

```
In [476]: round(data.describe(),1)
```

	age	fnlwgt	education.num	capital.gain	capital.loss	hours.per.week
count	32561.0	32561.0	32561.0	32561.0	32561.0	32561.0
mean	38.6	189778.4	10.1	1077.6	87.3	40.4
std	13.6	105550.0	2.6	7385.3	403.0	12.3
min	17.0	12285.0	1.0	0.0	0.0	1.0
25%	28.0	117827.0	9.0	0.0	0.0	40.0
50%	37.0	178356.0	10.0	0.0	0.0	40.0
75%	48.0	237051.0	12.0	0.0	0.0	45.0
max	90.0	1484705.0	16.0	99999.0	4356.0	99.0

```
In [477]: data.dtypes
```

```
age                int64
workclass          object
fnlwgt             int64
education          object
education.num      int64
marital.status     object
occupation         object
relationship       object
race              object
sex               object
capital.gain       float64
capital.loss       float64
hours.per.week     int64
native.country     object
income            object
dtype: object
```

```
data['income'].value_counts()
```

```
<50K    24720
>50K    1842
Name: income, dtype: int64
```

```
d = {'<50K':0, '>50K':1}
data['income']=data['income'].map(d)
```

```
data['workclass'].value_counts()
```

```
Private      22696
Self-emp-not-inc 2941
Local-gov    2093
Data column(s) total 15 column(s)
State-gov    1836
Self-emp-inc 1116
Federal-gov  960
Without-pay  14
Never-worked 23
Name: workclass, dtype: int64
```

```
print('Private :- ',data[data['workclass']=='Private']['income'].mean())
print('Self-emp-not-inc :- ',data[data['workclass']=='Self-emp-not-inc']['income'].mean())
print('Local-gov :- ',data[data['workclass']=='Local-gov']['income'].mean())
print('State-gov :- ',data[data['workclass']=='State-gov']['income'].mean())
print('Self-emp-inc :- ',data[data['workclass']=='Self-emp-inc']['income'].mean())
print('Federal-gov :- ',data[data['workclass']=='Federal-gov']['income'].mean())
print('Never-worked :- ',data[data['workclass']=='Never-worked']['income'].mean())
print('Without-pay :- ',data[data['workclass']=='Without-pay']['income'].mean())

Private :- 0.2186728939209017
Self-emp-not-inc :- 0.2849271940181031
Local-gov :- 0.29479216435738176
? :- 0.1040305018932661
State-gov :- 0.27195685670261943
Self-emp-inc :- 0.5573467670250896
Federal-gov :- 0.3864983333333333
Never-worked :- 0.0
Without-pay :- 0.0
```

```
array=np.array([data[data['workclass']=='Private']['income'].mean(),data[data['workclass']=='Self-emp-not-inc']
```

```
array.argsort()
```

```
array([7, 8, 3, 0, 4, 1, 2, 6, 5], dtype=int32)
```

```
di = ['Private':3,'Self-emp-not-inc':5,'Local-gov':6,'?':2,'State-gov':4,'Self-emp-inc':8,'Federal-gov':7,'Never-worked':9,'Without-pay':10]
data['workclass']=data['workclass'].map(di)
```

```
data['education'].value_counts()
```

```
HS-grad      10501
Some-college 7291
Bachelors    5335
Masters       1723
Assoc-voc    1382
11th         168
Assoc-acdm   1067
10th         933
7th-8th      646
Prof-school  576
9th          534
12th         433
Doctorate    413
5th-6th      333
1st-4th      198
Preschool   51
Name: education, dtype: int64
```

```
edu = ['Preschool':0,'1st-4th':1,'5th-6th':2,'7th-8th':3,'9th':4,'10th':5,'11th':6,'12th':7,'HS-grad':8,'Some-college':9,'Bachelors':10,'Masters':11,'Assoc-voc':12,'Assoc-acdm':13,'10th':14,'11th':15,'12th':16,'13th':17,'14th':18,'15th':19,'16th':20,'17th':21,'18th':22,'19th':23,'20th':24,'21st':25,'22nd':26,'23rd':27,'24th':28,'25th':29,'26th':30,'27th':31,'28th':32,'29th':33,'30th':34,'31st':35,'32nd':36,'33rd':37,'34th':38,'35th':39,'36th':40,'37th':41,'38th':42,'39th':43,'40th':44,'41st':45,'42nd':46,'43rd':47,'44th':48,'45th':49,'46th':50,'47th':51,'48th':52,'49th':53,'50th':54,'51st':55,'52nd':56,'53rd':57,'54th':58,'55th':59,'56th':60,'57th':61,'58th':62,'59th':63,'60th':64,'61st':65,'62nd':66,'63rd':67,'64th':68,'65th':69,'66th':70,'67th':71,'68th':72,'69th':73,'70th':74,'71st':75,'72nd':76,'73rd':77,'74th':78,'75th':79,'76th':80,'77th':81,'78th':82,'79th':83,'80th':84,'81st':85,'82nd':86,'83rd':87,'84th':88,'85th':89,'86th':90,'87th':91,'88th':92,'89th':93,'90th':94,'91st':95,'92nd':96,'93rd':97,'94th':98,'95th':99,'96th':100,'97th':101,'98th':102,'99th':103,'100th':104,'101st':105,'102nd':106,'103rd':107,'104th':108,'105th':109,'106th':110,'107th':111,'108th':112,'109th':113,'110th':114,'111th':115,'112th':116,'113th':117,'114th':118,'115th':119,'116th':120,'117th':121,'118th':122,'119th':123,'120th':124,'121st':125,'122nd':126,'123rd':127,'124th':128,'125th':129,'126th':130,'127th':131,'128th':132,'129th':133,'130th':134,'131st':135,'132nd':136,'133rd':137,'134th':138,'135th':139,'136th':140,'137th':141,'138th':142,'139th':143,'140th':144,'141st':145,'142nd':146,'143rd':147,'144th':148,'145th':149,'146th':150,'147th':151,'148th':152,'149th':153,'150th':154,'151st':155,'152nd':156,'153rd':157,'154th':158,'155th':159,'156th':160,'157th':161,'158th':162,'159th':163,'160th':164,'161st':165,'162nd':166,'163rd':167,'164th':168,'165th':169,'166th':170,'167th':171,'168th':172,'169th':173,'170th':174,'171st':175,'172nd':176,'173rd':177,'174th':178,'175th':179,'176th':180,'177th':181,'178th':182,'179th':183,'180th':184,'181st':185,'182nd':186,'183rd':187,'184th':188,'185th':189,'186th':190,'187th':191,'188th':192,'189th':193,'190th':194,'191st':195,'192nd':196,'193rd':197,'194th':198,'195th':199,'196th':200,'197th':201,'198th':202,'199th':203,'200th':204,'201st':205,'202nd':206,'203rd':207,'204th':208,'205th':209,'206th':210,'207th':211,'208th':212,'209th':213,'210th':214,'211st':215,'212nd':216,'213rd':217,'214th':218,'215th':219,'216th':220,'217th':221,'218th':222,'219th':223,'220th':224,'221st':225,'222nd':226,'223rd':227,'224th':228,'225th':229,'226th':230,'227th':231,'228th':232,'229th':233,'230th':234,'231st':235,'232nd':236,'233rd':237,'234th':238,'235th':239,'236th':240,'237th':241,'238th':242,'239th':243,'240th':244,'241st':245,'242nd':246,'243rd':247,'244th':248,'245th':249,'246th':250,'247th':251,'248th':252,'249th':253,'250th':254,'251st':255,'252nd':256,'253rd':257,'254th':258,'255th':259,'256th':260,'257th':261,'258th':262,'259th':263,'260th':264,'261st':265,'262nd':266,'263rd':267,'264th':268,'265th':269,'266th':270,'267th':271,'268th':272,'269th':273,'270th':274,'271st':275,'272nd':276,'273rd':277,'274th':278,'275th':279,'276th':280,'277th':281,'278th':282,'279th':283,'280th':284,'281st':285,'282nd':286,'283rd':287,'284th':288,'285th':289,'286th':290,'287th':291,'288th':292,'289th':293,'290th':294,'291st':295,'292nd':296,'293rd':297,'294th':298,'295th':299,'296th':300,'297th':301,'298th':302,'299th':303,'300th':304,'301st':305,'302nd':306,'303rd':307,'304th':308,'305th':309,'306th':310,'307th':311,'308th':312,'309th':313,'310th':314,'311st':315,'312nd':316,'313rd':317,'314th':318,'315th':319,'316th':320,'317th':321,'318th':322,'319th':323,'320th':324,'321st':325,'322nd':326,'323rd':327,'324th':328,'325th':329,'326th':330,'327th':331,'328th':332,'329th':333,'330th':334,'331st':335,'332nd':336,'333rd':337,'334th':338,'335th':339,'336th':340,'337th':341,'338th':342,'339th':343,'340th':344,'341st':345,'342nd':346,'343rd':347,'344th':348,'345th':349,'346th':350,'347th':351,'348th':352,'349th':353,'350th':354,'351st':355,'352nd':356,'353rd':357,'354th':358,'355th':359,'356th':360,'357th':361,'358th':362,'359th':363,'360th':364,'361st':365,'362nd':366,'363rd':367,'364th':368,'365th':369,'366th':370,'367th':371,'368th':372,'369th':373,'370th':374,'371st':375,'372nd':376,'373rd':377,'374th':378,'375th':379,'376th':380,'377th':381,'378th':382,'379th':383,'380th':384,'381st':385,'382nd':386,'383rd':387,'384th':388,'385th':389,'386th':390,'387th':391,'388th':392,'389th':393,'390th':394,'391st':395,'392nd':396,'393rd':397,'394th':398,'395th':399,'396th':400,'397th':401,'398th':402,'399th':403,'400th':404,'401st':405,'402nd':406,'403rd':407,'404th':408,'405th':409,'406th':410,'407th':411,'408th':412,'409th':413,'410th':414,'411st':415,'412nd':416,'413rd':417,'414th':418,'415th':419,'416th':420,'417th':421,'418th':422,'419th':423,'420th':424,'421st':425,'422nd':426,'423rd':427,'424th':428,'425th':429,'426th':430,'427th':431,'428th':432,'429th':433,'430th':434,'431st':435,'432nd':436,'433rd':437,'434th':438,'435th':439,'436th':440,'437th':441,'438th':442,'439th':443,'440th':444,'441st':445,'442nd':446,'443rd':447,'444th':448,'445th':449,'446th':450,'447th':451,'448th':452,'449th':453,'450th':454,'451st':455,'452nd':456,'453rd':457,'454th':458,'455th':459,'456th':460,'457th':461,'458th':462,'459th':463,'460th':464,'461st':465,'462nd':466,'463rd':467,'464th':468,'465th':469,'466th':470,'467th':471,'468th':472,'469th':473,'470th':474,'471st':475,'472nd':476,'473rd':477,'474th':478,'475th':479,'476th':480,'477th':481,'478th':482,'479th':483,'480th':484,'481st':485,'482nd':486,'483rd':487,'484th':488,'485th':489,'486th':490,'487th':491,'488th':492,'489th':493,'490th':494,'491st':495,'492nd':496,'493rd':497,'494th':498,'495th':499,'496th':500,'497th':501,'498th':502,'499th':503,'500th':504,'501st':505,'502nd':506,'503rd':507,'504th':508,'505th':509,'506th':510,'507th':511,'508th':512,'509th':513,'510th':514,'511st':515,'512nd':516,'513rd':517,'514th':518,'515th':519,'516th':520,'517th':521,'518th':522,'519th':523,'520th':524,'521st':525,'522nd':526,'523rd':527,'524th':528,'525th':529,'526th':530,'527th':531,'528th':532,'529th':533,'530th':534,'531st':535,'532nd':536,'533rd':537,'534th':538,'535th':539,'536th':540,'537th':541,'538th':542,'539th':543,'540th':544,'541st':545,'542nd':546,'543rd':547,'544th':548,'545th':549,'546th':550,'547th':551,'548th':552,'549th':553,'550th':554,'551st':555,'552nd':556,'553rd':557,'554th':558,'555th':559,'556th':560,'557th':561,'558th':562,'559th':563,'560th':564,'561st':565,'562nd':566,'563rd':567,'564th':568,'565th':569,'566th':570,'567th':571,'568th':572,'569th':573,'570th':574,'571st':575,'572nd':576,'573rd':577,'574th':578,'575th':579,'576th':580,'577th':581,'578th':582,'579th':583,'580th':584,'581st':585,'582nd':586,'583rd':587,'584th':588,'585th':589,'586th':590,'587th':591,'588th':592,'589th':593,'590th':594,'591st':595,'592nd':596,'593rd':597,'594th':598,'595th':599,'596th':600,'597th':601,'598th':602,'599th':603,'600th':604,'601st':605,'602nd':606,'603rd':607,'604th':608,'605th':609,'606th':610,'607th':611,'608th':612,'609th':613,'610th':614,'611st':615,'612nd':616,'613rd':617,'614th':618,'615th':619,'616th':620,'617th':621,'618th':622,'619th':623,'620th':624,'621st':625,'622nd':626,'623rd':627,'624th':628,'625th':629,'626th':630,'627th':631,'628th':632,'629th':633,'630th':634,'631st':635,'632nd':636,'633rd':637,'634th':638,'635th':639,'636th':640,'637th':641,'638th':642,'639th':643,'640th':644,'641st':645,'642nd':646,'643rd':647,'644th':648,'645th':649,'646th':650,'647th':651,'648th':652,'649th':653,'650th':654,'651st':655,'652nd':656,'653rd':657,'654th':658,'655th':659,'656th':660,'657th':661,'658th':662,'659th':663,'660th':664,'661st':665,'662nd':666,'663rd':667,'664th':668,'665th':669,'666th':670,'667th':671,'668th':672,'669th':673,'670th':674,'671st':675,'672nd':676,'673rd':677,'674th':678,'675th':679,'676th':680,'677th':681,'678th':682,'679th':683,'680th':684,'681st':685,'682nd':686,'683rd':687,'684th':688,'685th':689,'686th':690,'687th':691,'688th':692,'689th':693,'690th':694,'691st':695,'692nd':696,'693rd':697,'694th':698,'695th':699,'696th':700,'697th':701,'698th':702,'699th':703,'700th':704,'701st':705,'702nd':706,'703rd':707,'704th':708,'705th':709,'706th':710,'707th':711,'708th':712,'709th':713,'710th':714,'711st':715,'712nd':716,'713rd':717,'714th':718,'715th':719,'716th':720,'717th':721,'718th':722,'719th':723,'720th':724,'721st':725,'722nd':726,'723rd':727,'724th':728,'725th':729,'726th':730,'727th':731,'728th':732,'729th':733,'730th':734,'731st':735,'732nd':736,'733rd':737,'734th':738,'735th':739,'736th':740,'737th':741,'738th':742,'739th':743,'740th':744,'741st':745,'742nd':746,'743rd':747,'744th':748,'745th':749,'746th':750,'747th':751,'748th':752,'749th':753,'750th':754,'751st':755,'752nd':756,'753rd':757,'754th':758,'755th':759,'756th':760,'757th':761,'758th':762,'759th':763,'760th':764,'761st':765,'762nd':766,'763rd':767,'764th':768,'765th':769,'766th':770,'767th':771,'768th':772,'769th':773,'770th':774,'771st':775,'772nd':776,'773rd':777,'774th':778,'775th':779,'776th':780,'777th':781,'778th':782,'779th':783,'780th':784,'781st':785,'782nd':786,'783rd':787,'784th':788,'785th':789,'786th':790,'787th':791,'788th':792,'789th':793,'790th':794,'791st':795,'792nd':796,'793rd':797,'794th':798,'795th':799,'796th':800,'797th':801,'798th':802,'799th':803,'800th':804,'801st':805,'802nd':806,'803rd':807,'804th':808,'805th':809,'806th':810,'807th':811,'808th':812,'809th':813,'810th':814,'811st':815,'812nd':816,'813rd':817,'814th':818,'815th':819,'816th':820,'817th':821,'818th':822,'819th':823,'820th':824,'821st':825,'822nd':826,'823rd':827,'824th':828,'825th':829,'826th':830,'827th':831,'828th':832,'829th':833,'830th':834,'831st':835,'832nd':836,'833rd':837,'834th':838,'835th':839,'836th':840,'837th':841,'838th':842,'839th':843,'840th':844,'841st':845,'842nd':846,'843rd':847,'844th':848,'845th':849,'846th':850,'847th':851,'848th':852,'849th':853,'850th':854,'851st':855,'852nd':856,'853rd':857,'854th':858,'855th':859,'856th':860,'857th':861,'858th':862,'859th':863,'860th':864,'861st':865,'862nd':866,'863rd':867,'864th':868,'865th':869,'866th':870,'867th':871,'868th':872,'869th':873,'870th':874,'871st':875,'872nd':876,'873rd':877,'874th':878,'875th':879,'876th':880,'877th':881,'878th':882,'879th':883,'880th':884,'881st':885,'882nd':886,'883rd':887,'884th':888,'885th':889,'886th':890,'887th':891,'888th':892,'889th':893,'890th':894,'891st':895,'892nd':896,'893rd':897,'894th':898,'895th':899,'896th':900,'897th':901,'898th':902,'899th':903,'900th':904,'901st':905,'902nd':906,'903rd':907,'904th':908,'905th':909,'906th':910,'907th':911,'908th':912,'909th':913,'910th':914,'911st':915,'912nd':916,'913rd':917,'914th':918,'915th':919,'916th':920,'917th':921,'918th':922,'919th':923,'920th':924,'921st':925,'922nd':926,'923rd':927,'924th':928,'925th':929,'926th':930,'927th':931,'928th':932,'929th':933,'930th':934,'931st':935,'932nd':936,'933rd':937,'934th':938,'935th':939,'936th':940,'937th':941,'938th':942,'939th':943,'940th':944,'941st':945,'942nd':946,'943rd':947,'944th':948,'945th':949,'946th':950,'947th':951,'948th':952,'949th':953,'950th':954,'951st':955,'952nd':956,'953rd':957,'954th':958,'955th':959,'956th':960,'957th':961,'958th':962,'959th':963,'960th':964,'961st':965,'962nd':966,'963rd':967,'964th':968,'965th':969,'966th':970,'967th':971,'968th':972,'969th':973,'970th':974,'971st':975,'972nd':976,'973rd':977,'974th':978,'975th':979,'976th':980,'977th':981,'978th':982,'979th':983,'980th':984,'981st':985,'982nd':986,'983rd':987,'984th':988,'985th':989,'986th':990,'987th':991,'988th':992,'989th':993,'990th':994,'991st':995,'992nd':996,'993rd':997,'994th':998,'995th':999,'996th':1000,'997th':1001,'998th':1002,'999th':1003
```