LAB EVAL 102117154  
  
PREDICTIVE ANALYSIS

import os

for dirname, \_, filenames in os.walk('/kaggle/input'):

for filename in filenames:

print(os.path.join(dirname, filename))



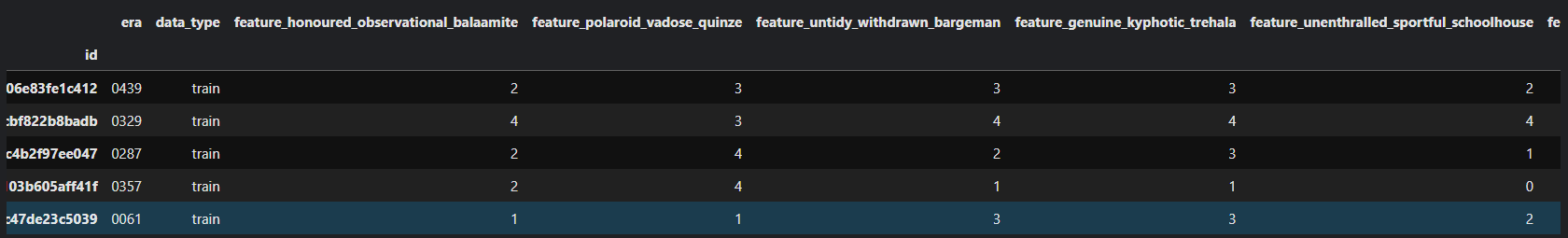
import numpy as np

import pandas as pd

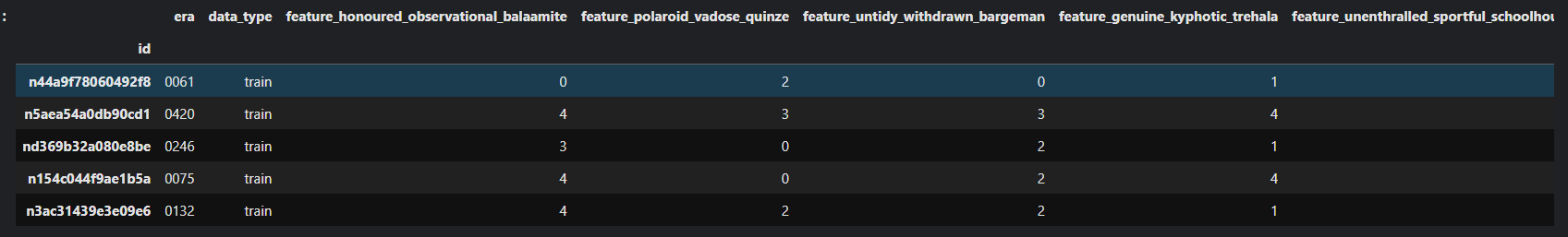
tantrain = pd.read\_parquet('/kaggle/input/2024ucs654labeval1004/Lab Eval/train\_data.parquet')

tantest = pd.read\_parquet('/kaggle/input/2024ucs654labeval1004/Lab Eval/test\_data.parquet')

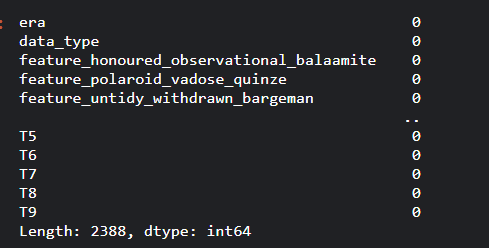
tantrain.head()



tantest.head()



tantrain.isnull().sum()



tantrain.shape



random\_seed = 42

sample\_size=1930000

np.random.seed(random\_seed)

tantrain= tantrain.sample(n=sample\_size)

y\_tantrain = tantrain['T4']

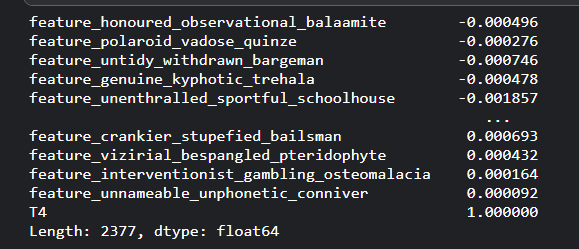
tantrain=tantrain.drop(columns=['era', 'data\_type', 'T0', 'T1', 'T2', 'T3','T5', 'T6', 'T7', 'T8', 'T9'])

X\_tantrain = tantrain

del tantrain

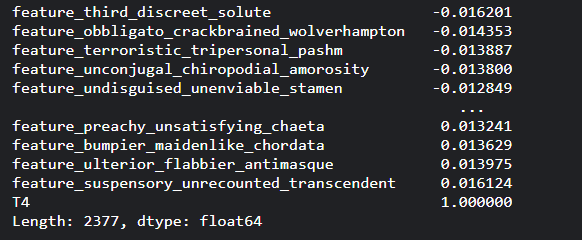
corr\_values = X\_tantrain.apply(lambda x: x.corr(y\_tantrain))

print(corr\_values)



corr\_values = corr\_values.sort\_values()

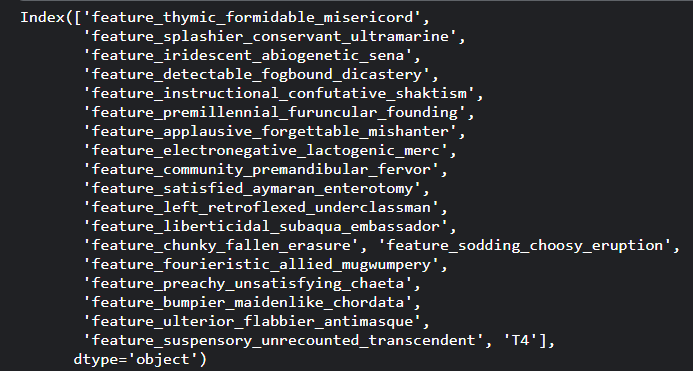
print(corr\_values)



cortan = corr\_values[corr\_values<=-0.01]

cor2tan = corr\_values[corr\_values>=0.01]

print(cor2tan.index)



finalcolumns=['feature\_third\_discreet\_solute',

'feature\_obbligato\_crackbrained\_wolverhampton',

'feature\_undisguised\_unenviable\_stamen',

'feature\_terroristic\_tripersonal\_pashm',

'feature\_unconjugal\_chiropodial\_amorosity',

'feature\_undrilled\_wheezier\_countermand',

'feature\_encysted\_conventionalized\_dematerialization',

'feature\_unbarking\_apolitical\_hibernian',

'feature\_surrogate\_unmalleable\_tasset', 'feature\_wetter\_unbaffled\_loma',

'feature\_unscriptural\_coconut\_trisulphide',

'feature\_optical\_kempt\_aisle', 'feature\_fanfold\_tartarian\_diamondback',

'feature\_elmier\_unidentifiable\_broccoli',

'feature\_eruciform\_novice\_thanker', 'feature\_zincky\_unseemly\_butt',

'feature\_multipolar\_syncopated\_ambrotype',

'feature\_addressable\_intransitive\_reconnoitrer',

'feature\_lemuroid\_unwishful\_mannequin',

'feature\_unreproving\_capsian\_decolourization',

'feature\_bursarial\_southmost\_kaduna',

'feature\_goyish\_riparian\_recipient', 'feature\_unpreached\_pickiest\_lint',

'feature\_amitotic\_gonadial\_submediant',

'feature\_domanial\_shellproof\_rationing',

'feature\_subfusc\_furriest\_nervule',

'feature\_herniated\_exasperate\_victorian',

'feature\_setose\_processed\_crevice',

'feature\_gandhian\_discretional\_cricoid',

'feature\_associate\_unproper\_gridder',

'feature\_laziest\_saronic\_hornbeam', 'feature\_milkier\_gassy\_pincushion',

'feature\_shrinelike\_introverted\_eagre',

'feature\_smuggest\_galvanic\_memorial',

'feature\_toed\_accusatory\_zoologist', 'feature\_kirtled\_cockiest\_etaerio',

'feature\_fearsome\_merry\_bluewing', 'feature\_scissile\_dejected\_kainite',

'feature\_incertain\_catchable\_zibet', 'feature\_synodal\_feisty\_weave',

'feature\_anencephalic\_unattempted\_pschent',

'feature\_shrinelike\_unreplaceable\_nitrogenization',

'feature\_bamboo\_nosier\_phil', 'feature\_litigant\_unsizable\_rhebok',

'feature\_sensitive\_incendiary\_heraclid',

'feature\_fungible\_allotted\_deterioration',

'feature\_idled\_unwieldy\_improvement',

'feature\_deposed\_toughish\_bribery', 'feature\_cupular\_porky\_catafalque','feature\_preterite\_antediluvian\_parasailing',

'feature\_unpitied\_jingoist\_pyretology',

'feature\_hippiatric\_tinctorial\_slowpoke',

'feature\_swelled\_jugate\_haystack', 'feature\_bifocal\_disposable\_clacton',

'feature\_thymic\_formidable\_misericord',

'feature\_electronegative\_lactogenic\_merc',

'feature\_cloaked\_taillike\_usurpation',

'feature\_nonnegotiable\_errant\_soya', 'feature\_sodding\_choosy\_eruption',

'feature\_fumed\_pivotal\_oscine', 'feature\_unconfinable\_snuffly\_cupid',

'feature\_detectable\_fogbound\_dicastery',

'feature\_phrenetic\_visitorial\_entrenchment',

'feature\_subatomic\_raffish\_hexagram',

'feature\_fishable\_ascendible\_micky',

'feature\_manufactured\_nodal\_seeking',

'feature\_splashier\_conservant\_ultramarine',

'feature\_premillennial\_furuncular\_founding',

'feature\_instructional\_confutative\_shaktism',

'feature\_community\_premandibular\_fervor',

'feature\_left\_retroflexed\_underclassman',

'feature\_satisfied\_aymaran\_enterotomy',

'feature\_liberticidal\_subaqua\_embassador',

'feature\_fourieristic\_allied\_mugwumpery',

'feature\_chunky\_fallen\_erasure', 'feature\_preachy\_unsatisfying\_chaeta',

'feature\_bumpier\_maidenlike\_chordata',

'feature\_ulterior\_flabbier\_antimasque',

'feature\_suspensory\_unrecounted\_transcendent']

X\_tantrain=X\_tantrain[finalcolumns]

from sklearn.model\_selection import train\_test\_split

X\_tantrain\_splitting, X\_tantest\_splitting, y\_tantrain\_splitting, y\_tantest\_splitting = train\_test\_split(X\_tantrain, y\_tantrain, test\_size=0.2, random\_state=42)

print(cortan.index.shape)



import xgboost as xgb

dubtrain = xgb.DMatrix(X\_tantrain\_splitting, label=y\_tantrain\_splitting)

dubtest = xgb.DMatrix(X\_tantest\_splitting, label=y\_tantest\_splitting)

params = {

'max\_depth': 20,

'learning\_rate': 0.07,

'objective': 'reg:squarederror',

'eval\_metric': 'rmse',

'n\_estimators' : 250

}

tan\_reg= xgb.XGBRegressor(\*\*params)

tan\_reg.fit(X\_tantrain\_splitting, y\_tantrain\_splitting)

y\_tanprediction = tan\_reg.predict(X\_tantest\_splitting)

from sklearn.metrics import r2\_score

r2finalscoretan = r2\_score(y\_tantest\_splitting, y\_tanprediction)

print("Final R-square Score:", r2finalscoretan)



final\_tan\_test\_data = tantest.drop(columns=['era', 'data\_type'])

del tantest

final\_tan\_test\_data=final\_tan\_test\_data[finalcolumns]

mypred = tan\_reg.predict(final\_tan\_test\_data)

mysub = pd.DataFrame({'ID': final\_tan\_test\_data.index, 'Target': np.round(mypred, 2)})

mysub.to\_csv('submission\_final.csv', index=False)

mysub.head()

