

```
from sklearn.model_selection import train_test_split

import numpy as np
from sklearn.preprocessing import StandardScaler
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2
from sklearn.metrics import roc_auc_score
from sklearn.linear_model import LogisticRegression
import pickle
import matplotlib.pyplot as plt

In [8]: def select_best(index_X_dep,X_train,X_test)
    test = SelectKBest(score_func=chi2, k=n)
    fit=train.fit(index_X_dep,X_train)
    select_features = fit.transform(index_X_dep)
    return select_features

def split_scaler(index_X_dep,X_train,X_test,y_train,y_test = train_test_split(index_X_dep,X_train,y_train,size = 0.25, random_state = 0))
X_train,X_test,y_train,y_test = train_test_split(index_X_dep,X_train,y_train,size = 0.25, random_state = 0)
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
return X_train,X_test,y_train,y_test

def cm_prediction(classifier,X_test):
y_pred = classifier.predict(X_test)

# Making the Confusion Matrix
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test,y_pred)

from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
fcm = confusion_matrix(y_test,y_pred)

Accuracy=accuracy_score(y_test,y_pred)
report=classification_report(y_test,y_pred)
return classifier,Accuracy,report,X_test,y_test,cm

In [13]: def logistic(X_train,y_train,X_test):
# Fitting K-NN to the Training set
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state = 0)
classifier.fit(X_train,y_train)
classifier.Accuracy,report,X_test,y_test,cm=cm_prediction(classifier,X_test)
return classifier,Accuracy,report,X_test,y_test,cm

def svm_linear(X_train,y_train,X_test):
from sklearn.svm import SVC
classifier = SVC(kernel = 'linear', random_state = 0)
classifier.fit(X_train,y_train)
classifier.Accuracy,report,X_test,y_test,cm=cm_prediction(classifier,X_test)
return classifier,Accuracy,report,X_test,y_test,cm

def svm_RL(X_train,y_train,X_test):
from sklearn.svm import SVC
classifier = SVC(kernel = 'rbf', random_state = 0)
classifier.fit(X_train,y_train)
classifier.Accuracy,report,X_test,y_test,cm=cm_prediction(classifier,X_test)
return classifier,Accuracy,report,X_test,y_test,cm

def naive(X_train,y_train,X_test):
# Fitting K-NN to the Training set
from sklearn.naive_bayes import GaussianNB
classifier = GaussianNB()
classifier.fit(X_train,y_train)
classifier.Accuracy,report,X_test,y_test,cm=cm_prediction(classifier,X_test)
return classifier,Accuracy,report,X_test,y_test,cm

def knn(X_train,y_train,X_test):
# Fitting K-NN to the Training set
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)
classifier.fit(X_train,y_train)
classifier.Accuracy,report,X_test,y_test,cm=cm_prediction(classifier,X_test)
return classifier,Accuracy,report,X_test,y_test,cm

def Decision(X_train,y_train,X_test):
# Fitting K-NN to the Training set
from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier(criterion = 'entropy', random_state = 0)
classifier.fit(X_train,y_train)
classifier.Accuracy,report,X_test,y_test,cm=cm_prediction(classifier,X_test)
return classifier,Accuracy,report,X_test,y_test,cm

def Random(X_train,y_train,X_test):
# Fitting K-NN to the Training set
from sklearn.ensemble import RandomForestClassifier
classifier = RandomForestClassifier(n_estimators = 10, criterion = 'entropy', random_state = 0)
classifier.fit(X_train,y_train)
classifier.Accuracy,report,X_test,y_test,cm=cm_prediction(classifier,X_test)
return classifier,Accuracy,report,X_test,y_test,cm

def select_Classification(aoclog,aocsvm,aocsvml,aocknn,aocnav,aocdes,aocrf):
dataframe=pd.DataFrame(index=['category'],columns=['Logistic','SVM','SVML','KNN','Naive','Decision','Random'])
for number,index in enumerate(dataframe.index):
dataframe['Logistic'][index]=aoclog[number]
dataframe['SVM'][index]=aocsvm[number]
dataframe['SVML'][index]=aocsvml[number]
dataframe['KNN'][index]=aocknn[number]
dataframe['Naive'][index]=aocnav[number]
dataframe['Decision'][index]=aocdes[number]
dataframe['Random'][index]=aocrf[number]
return dataframe

In [41]: dataset=pd.read_csv("ad_click_dataset.csv")
dataset

Out[41]: Id full_name age gender device_type ad_position browsing_history time_of_day click
0 670 User670 22.0 NaN Desktop Top Shopping Afternoon 1
1 3044 User3044 NaN Male Desktop Top NaN NaN 1
2 5912 User5912 41.0 Non-Binary NaN Side Education Night 1
3 5418 User5418 34.0 Male NaN NaN Entertainment Evening 1
4 9452 User9452 39.0 Non-Binary NaN NaN Social Media Morning 0
... ..
9995 8510 User8510 NaN NaN Mobile Top Education NaN 0
9996 7843 User7843 NaN Female Desktop Bottom Entertainment NaN 0
9997 3914 User3914 NaN Male Mobile Side NaN Morning 0
9998 7924 User7924 NaN NaN Desktop NaN Shopping Morning 1
9999 3056 User3056 44.0 Male Tablet Top Social Media Morning 0
10000 rows x 9 columns

In [51]: dataset=pd.get_dummies(dataset,drop_first=True)
dataset.dropna(inplace=True)
dataset

Out[51]: Id age click full_name_User100 full_name_User1000 full_name_User10000 full_name_User1001 full_name_User1002 full_name_User10011 full_name_User10012 ... device_type_Tablet ad_position_Side ad_position_Top browsing_History_Entertainment browsing_History_News browsing_History_Other
0 670 22.0 1 False False False False False False ... False False True False False
1 3044 34.0 1 False False False False False False ... False True False False False
2 5912 41.0 1 False False False False False False ... False True False False False
3 5418 34.0 1 False False False False False False ... False False False True False
4 9452 39.0 0 False False False False False False ... False False False False False
6 7808 26.0 1 False False False False False False ... False False True False False
... ..
9987 2876 23.0 1 False False False False False False ... True False False False False
9988 2713 52.0 0 False False False False False False ... False True False False False
9990 9540 64.0 0 False False False False False False ... False True False False False
9993 503 43.0 1 False False False False False False ... False False False False False
9999 3056 44.0 0 False False False False False False ... True False True False False
5234 rows x 4015 columns

In [61]: dataset.columns

Out[61]: Index(['Id', 'age', 'click', 'full_name_User100', 'full_name_User1000', 'full_name_User10000', 'full_name_User1001', 'full_name_User1002', 'full_name_User10011', 'full_name_User10012', ..., 'device_type_Tablet', 'ad_position_Side', 'ad_position_Top', 'browsing_History_Entertainment', 'browsing_History_News', 'browsing_History_Other', 'time_of_Day_Morning', 'time_of_Day_Evening', 'time_of_Day_Night'], dtype='object', length=4015)

In [71]: dataset["click"].value_counts()

Out[71]: click
1 3591
0 1483
Name: count, dtype: int64

In [91]: index_X_dataset[['Id','age','full_name_User100','full_name_User1000','full_name_User10000','full_name_User1001','full_name_User1002','full_name_User10011','full_name_User10012','device_type_Tablet','ad_position_Side','ad_position_Top'],'dep_Y_dataase=['click']

In [117]: libsvm=select_kbest(index_X_dep,X_dep,Y,S)
aoclog=[]
aocsvm=[]
aocsvml=[]
aocknn=[]
aocnav=[]
aocdes=[]
aocrf=[]
libsvm

Out[117]: array([[6.70e+02, 2.20e+01, 0.00e+00, 0.00e+00, 0.00e+00, 0.00e+00],
       [5.91e+03, 4.10e+01, 0.00e+00, 0.00e+00, 0.00e+00, 0.00e+00],
       [5.41e+03, 3.40e+01, 1.00e+00, 1.00e+00, 0.00e+00, 0.00e+00],
       ...,
       [9.54e+03, 6.40e+01, 0.00e+00, 0.00e+00, 0.00e+00, 0.00e+00],
       [9.00e+02, 4.30e+01, 0.00e+00, 0.00e+00, 0.00e+00, 0.00e+00],
       [9.35e+03, 4.40e+01, 0.00e+00, 0.00e+00, 0.00e+00, 0.00e+00]])

In [181]: X_train,X_test,y_train,y_test=np.split_scalar(libsvm,dep_Y)
classifier,Accuracy,report,X_test,y_test,cm=logistic(X_train,y_train,X_test)
aoclog.append(Accuracy)
classifier,Accuracy,report,X_test,y_test,cm=svm_linear(X_train,y_train,X_test)
aocsvm.append(Accuracy)
classifier,Accuracy,report,X_test,y_test,cm=svm_RL(X_train,y_train,X_test)
aocsvml.append(Accuracy)
classifier,Accuracy,report,X_test,y_test,cm=knn(X_train,y_train,X_test)
aocknn.append(Accuracy)
classifier,Accuracy,report,X_test,y_test,cm=naive(X_train,y_train,X_test)
aocnav.append(Accuracy)
classifier,Accuracy,report,X_test,y_test,cm=Decision(X_train,y_train,X_test)
aocdes.append(Accuracy)
classifier,Accuracy,report,X_test,y_test,cm=random(X_train,y_train,X_test)
aocrf.append(Accuracy)

C:\Users\Downieash\AppData\Local\Temp\ipykernel_9300\2796833036.py:65: FutureWarning: ChainedAssignmentError: behaviour will change in pandas 3.0!
You are setting values through chained assignment. Currently this works in certain cases, but when using Copy-on-Write (which will become the default behaviour in pandas 3.0) this will never work to update the original DataFrame or Series, because the intermediate object on which we are setting values will behave as a copy.
A typical example is when you are setting values in a column of a DataFrame, like:
df["col"][row_indexer] = value
Use "df.loc[row_indexer, "col"] = values" instead, to perform the assignment in a single step and ensure this keeps updating the original "df".
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
dataframe['Logistic'][index]=aoclog[number]
C:\Users\Downieash\AppData\Local\Temp\ipykernel_9300\2796833036.py:66: FutureWarning: ChainedAssignmentError: behaviour will change in pandas 3.0!
You are setting values through chained assignment. Currently this works in certain cases, but when using Copy-on-Write (which will become the default behaviour in pandas 3.0) this will never work to update the original DataFrame or Series, because the intermediate object on which we are setting values will behave as a copy.
A typical example is when you are setting values in a column of a DataFrame, like:
df["col"][row_indexer] = value
Use "df.loc[row_indexer, "col"] = values" instead, to perform the assignment in a single step and ensure this keeps updating the original "df".
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
dataframe['SVM'][index]=aocsvm[number]
C:\Users\Downieash\AppData\Local\Temp\ipykernel_9300\2796833036.py:68: FutureWarning: ChainedAssignmentError: behaviour will change in pandas 3.0!
You are setting values through chained assignment. Currently this works in certain cases, but when using Copy-on-Write (which will become the default behavior in pandas 3.0) this will never work to update the original DataFrame or Series, because the intermediate object on which we are setting values will behave as a copy.
A typical example is when you are setting values in a column of a DataFrame, like:
df["col"][row_indexer] = value
Use "df.loc[row_indexer, "col"] = values" instead, to perform the assignment in a single step and ensure this keeps updating the original "df".
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
dataframe['SVML'][index]=aocsvml[number]
C:\Users\Downieash\AppData\Local\Temp\ipykernel_9300\2796833036.py:69: FutureWarning: ChainedAssignmentError: behaviour will change in pandas 3.0!
You are setting values through chained assignment. Currently this works in certain cases, but when using Copy-on-Write (which will become the default behaviour in pandas 3.0) this will never work to update the original DataFrame or Series, because the intermediate object on which we are setting values will behave as a copy.
A typical example is when you are setting values in a column of a DataFrame, like:
df["col"][row_indexer] = value
Use "df.loc[row_indexer, "col"] = values" instead, to perform the assignment in a single step and ensure this keeps updating the original "df".
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
dataframe['KNN'][index]=aocknn[number]
C:\Users\Downieash\AppData\Local\Temp\ipykernel_9300\2796833036.py:70: FutureWarning: ChainedAssignmentError: behaviour will change in pandas 3.0!
You are setting values through chained assignment. Currently this works in certain cases, but when using Copy-on-Write (which will become the default behaviour in pandas 3.0) this will never work to update the original DataFrame or Series, because the intermediate object on which we are setting values will behave as a copy.
A typical example is when you are setting values in a column of a DataFrame, like:
df["col"][row_indexer] = value
Use "df.loc[row_indexer, "col"] = values" instead, to perform the assignment in a single step and ensure this keeps updating the original "df".
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
dataframe['Naive'][index]=aocnav[number]
C:\Users\Downieash\AppData\Local\Temp\ipykernel_9300\2796833036.py:71: FutureWarning: ChainedAssignmentError: behaviour will change in pandas 3.0!
You are setting values through chained assignment. Currently this works in certain cases, but when using Copy-on-Write (which will become the default behaviour in pandas 3.0) this
```


