"Importing required libraries"

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns; sns.set()

from sklearn.datasets import fetch\_20newsgroups

data= fetch\_20newsgroups()

print(data.target\_names)

"Defining categories"

categories=['alt.atheism', 'comp.graphics', 'comp.os.ms-windows.misc', 'comp.sys.ibm.pc.hardware', 'comp.sys.mac.hardware', 'comp.windows.x', 'misc.forsale', 'rec.autos', 'rec.motorcycles', 'rec.sport.baseball', 'rec.sport.hockey', 'sci.crypt', 'sci.electronics', 'sci.med', 'sci.space', 'soc.religion.christian', 'talk.politics.guns', 'talk.politics.mideast', 'talk.politics.misc', 'talk.religion.misc']

"Train data on these categories"

train=fetch\_20newsgroups(subset='train',categories=categories)

"Test data for these categories"

test=fetch\_20newsgroups(subset='test',categories=categories)

"Print training data"

print(train.data[5])

"Importing required libraries"

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.naive\_bayes import MultinomialNB

from sklearn.pipeline import make\_pipeline

"Creating model based on Multinomial Naive Bayes"

model=make\_pipeline(TfidfVectorizer(),MultinomialNB())

"Training the model with the train data"

model.fit(train.data,train.target)

"Creating labels for the test data"

labels=model.predict(test.data)

"Creating confusion matrix and heatmap"

from sklearn.metrics import confusion\_matrix

mat=confusion\_matrix(test.target,labels)

sns.heatmap(mat.T,square=True, annot=True ,fmt='d',cbar=False ,xticklabels=train.target\_names,yticklabels=train.target\_names)

"Plotting Heatmap of Confusion Matrix"

plt.xlabel('true label')

plt.ylabel('predicted label');

"Predicting category on new data based on trained model"

def predict\_category(s, train=train, model=model):

pred=model.predict([s])

return train.target\_names[pred[0]]