project

April 19, 2020

[122]:
```
import numpy as np
import pandas as pd
from sklearn import preprocessing

df = pd.read_csv('training.csv') # use test which is smaller for development
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9500 entries, 0 to 9499
Data columns (total 3 columns):
article_number    9500 non-null int64
article_words     9500 non-null object
topic             9500 non-null object
dtypes: int64(1), object(2)
memory usage: 222.8+ KB
```

[123]:
```
df.head()
```

[123]:

	article_number	article_words	topic
0	1	open,absent,cent,cent,cent,stock,inflow,rate,k…	FOREX MARKETS
1	2	morn,stead,end,end,day,day,day,patch,patch,pat…	MONEY MARKETS
2	3	socc,socc,world,world,recent,law,fifa,fifa,fif…	SPORTS
3	4	open,forint,forint,forint,forint,cent,cent,ste…	FOREX MARKETS
4	5	morn,complet,weekend,minut,minut,minut,arrow,d…	IRRELEVANT

[124]:
```
input_cols = ['article_words']
out_cols = ['topic']
X = df[input_cols]
y = df[out_cols]
split = int(X.shape[0] * 0.9)
#X_t = X[:split]
```

```
#y_t = y[:split]
#X_v = X[split:]
#y_v = y[split:]
```

[154]:
```
TOP = 5
words = set() # every word in the doc
top_words = set() # a set of TOP # words from each article
features = [] # array of array of top word
for i in range (X.shape[0]):
    tally = dict()
    # this for loop get all words and count their frequency of each article,
    for word in X['article_words'][i].split(','):
        words.add(word)
        if word not in tally.keys():
            tally[word] = 1
        else:
            tally[word] +=1
    #print(tally)
    sorted_tally = sorted(tally.items(), key=lambda kv: kv[1], reverse=True)
    #print(sorted_tally[:TOP], y['topic'][i])

    # this loop puts top words to the list of features
    loc_feat = []
    for j in range(TOP):
        top_words.add(sorted_tally[j][0])
        loc_feat.append(sorted_tally[j][0])
    features.append(loc_feat)

print(len(words))
#print(words)
print(len(top_words))
#print(top_words)
#print(features)
```

```
35823
6486
```

[155]:
```
index = dict()
count = 0
for word in top_words:
    index[word] = count
    count+=1
# print(index)
```

[ ]:
```
```

[156]:
```
bool_features = []
for i in range (len(features)):
    bool_f = [0] * len(top_words)
    for t in range (TOP):
        bool_f[index[features[i][t]]] = 1
    bool_features.append(tuple(bool_f))
df_cleaned = pd.DataFrame(bool_features, columns = list(top_words))
# print(df_cleaned.head())
# df_cleaned.head().to_csv('tmp.csv') # write to file to validate output
```

[157]:
```
from sklearn.preprocessing import OneHotEncoder
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import roc_auc_score
from sklearn.metrics import precision_recall_fscore_support

enc = OneHotEncoder(handle_unknown='ignore')
enc.fit(y)
y_trans = enc.transform(y).toarray()

START = 3
END = 4
#y_pred = 0
def optimal_min_leaf(tx, ty, true_y, pred_xarg):
    scores = []
    metrics = []
    for k in range(START, END):
        dtc = DecisionTreeClassifier(min_samples_leaf=k)
        dtc.fit(tx, ty)
        y_pred = dtc.predict_proba(pred_xarg)
        print(len(y_pred))
        score = roc_auc_score(true_y,y_pred)
        metric = precision_recall_fscore_support(true_y,y_pred, average='micro')
        scores.append(score)
        metrics.append(metric)
        print(dtc.min_samples_leaf,score)
    return START+np.argmax(scores), scores, metrics


#op_min_leaf, test_scores, metrics = optimal_min_leaf(df_cleaned[:split],␣
 ↪y_trans[:split], y_trans[split:],df_cleaned[split:])
#print('op_min_samples_leaf =', op_min_leaf)
#print('score =', test_scores)
#print('precision recall f1 =', metrics)
```

[158]:
```
dtc = DecisionTreeClassifier(min_samples_leaf=5)
dtc.fit(df_cleaned[:split], y_trans[:split])
y_pred = dtc.predict_proba(df_cleaned[split:])
```

[166]:
```
#print(len(y_trans[split:]),y_trans[split:])
y_p_sum = []
for k in range (len(y_pred[0])):
    y_p_sum.append([]);

#print(type(y_pred[0][:,1]))
#print(len(y_pred))
#print(y_trans[split:])

#convert y_pred[11][950] to y_p_sum[950][11] where 950 is number of sample and␣
 ↪11 is the type of article

for i in range(len(y_pred)):
    for k in range (len(y_pred[i])):
        #print(y_pred[i][:,1][k])
        y_p_sum[k].append(y_pred[i][:,1][k])

count = 0
for i in range(len(y_p_sum)):
    index = np.argmax(y_p_sum[i])
    index_t = np.argmax(y_trans[split+i])
    #print(i, index)
    #print(i," true ",index_t)
    if index == index_t:
        count+=1;
print(count, len(y_p_sum), count/len(y_p_sum))

metric = precision_recall_fscore_support(y_trans[split:] ,y_p_sum,␣
 ↪average='micro')
# Classification metrics can't handle a mix of multilabel-indicator and␣
 ↪continuous-multioutput targets
```

```
669 950 0.7042105263157895
```

```
      ␣
↪---------------------------------------------------------------------------

       ValueError                                Traceback (most recent call␣
↪last)

       <ipython-input-166-1dfccab7892b> in <module>
        25 print(count, len(y_p_sum), count/len(y_p_sum))
        26
   ---> 27 metric = precision_recall_fscore_support(y_trans[split:] ,y_p_sum,␣
↪average='micro')
```

```
      F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py in␣
↪precision_recall_fscore_support(y_true, y_pred, beta, labels, pos_label,␣
↪average, warn_for, sample_weight)
     1413         raise ValueError("beta should be >0 in the F-beta score")
     1414     labels = _check_set_wise_labels(y_true, y_pred, average, labels,
  -> 1415                                     pos_label)
     1416
     1417     # Calculate tp_sum, pred_sum, true_sum ###


      F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py in␣
↪_check_set_wise_labels(y_true, y_pred, average, labels, pos_label)
     1237                          str(average_options))
     1238
  -> 1239     y_type, y_true, y_pred = _check_targets(y_true, y_pred)
     1240     present_labels = unique_labels(y_true, y_pred)
     1241     if average == 'binary':


      F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py in␣
↪_check_targets(y_true, y_pred)
       79     if len(y_type) > 1:
       80         raise ValueError("Classification metrics can't handle a mix␣
↪of {0} "
  ---> 81                          "and {1} targets".format(type_true,␣
↪type_pred))
       82
       83     # We can't have more than one value on y_type => The set is no␣
↪more needed


      ValueError: Classification metrics can't handle a mix of␣
↪multilabel-indicator and continuous-multioutput targets
```

[165]:
```
from sklearn.model_selection import cross_val_score
from sklearn.naive_bayes import MultinomialNB, BernoulliNB

def Model_Score (X, y, method, k):
    clf = method

    accuracy_scores = cross_val_score(clf, X, y, cv=k, scoring="accuracy")
    precision_scores = cross_val_score(clf, X, y, cv=k,␣
 ↪scoring="precision_macro")
    recall_scores = cross_val_score(clf, X, y, cv=k, scoring="recall_macro")
    f1_scores = cross_val_score(clf, X, y, cv=k, scoring="f1_macro")
```

```
    return np.mean(accuracy_scores), np.mean(precision_scores), np.
↪mean(recall_scores), np.mean(f1_scores)

bernoulliNB_accuracy, bernoulliNB_precision, bernoulliNB_recall, bernoulliNB_f1␣
↪= Model_Score(df_cleaned, y, BernoulliNB(), 10)
print("==========================================================================================")
print("Without doing any data cleaning, the score of bernoulliNB,\naccuracy:  "␣
↪+ str(bernoulliNB_accuracy) +
     "\nprecision: " + str(bernoulliNB_precision) + "\nrecall:    " +␣
↪str(bernoulliNB_recall) + "\nf1:        " +
     str(bernoulliNB_f1))
```

```
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
```

```
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
```

```
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
```

```
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
```

```
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
```

```
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
```

```
  y = column_or_1d(y, warn=True)
```

```
================================================================================
====
Without doing any data cleaning, the score of bernoulliNB,
accuracy:  0.7019830331685644
precision: 0.23682494058828438
recall:    0.2372229847006142
f1:        0.21575967250809636
```

```
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
```

[161]:

```
multinomialNB_accuracy, multinomialNB_precision, multinomialNB_recall,␣
 ↪multinomialNB_f1 = Model_Score(df_cleaned, y, MultinomialNB(), 10)
print("Without doing any data cleaning, the score of multinomialNB,\naccuracy: ␣
 ↪" + str(multinomialNB_accuracy) +
      "\nprecision: " + str(multinomialNB_precision) + "\nrecall:    " +␣
 ↪str(multinomialNB_recall) + "\nf1:        " +
      str(multinomialNB_f1))
```

```
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
```

```
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
```

```
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
```

```
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
```

```
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)
F:\Anaconda3\lib\site-packages\sklearn\metrics\classification.py:1437:
UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels
```

```
with no predicted samples.
  'precision', 'predicted', average, warn_for)
F:\Anaconda3\lib\site-packages\sklearn\utils\validation.py:724:
DataConversionWarning: A column-vector y was passed when a 1d array was
expected. Please change the shape of y to (n_samples, ), for example using
ravel().
  y = column_or_1d(y, warn=True)

Without doing any data cleaning, the score of multinomialNB,
accuracy:  0.7122979293413069
precision: 0.6061310587103884
recall:    0.3993599979849251
f1:        0.43496036008611555
```

[163]:
```
#this would take over 5 min to run. interrupted.
# dt_accuracy, dt_precision, dt_recall, dt_f1 = Model_Score(df_cleaned, y,␣
 ↪DecisionTreeClassifier(), 10)
# print("Without doing any data cleaning, the score of multinomialNB,\naccuracy:
 ↪   " + str(dt_accuracy) +
#     "\nprecision: " + str(dt_precision) + "\nrecall:    " + str(dt_recall) +␣
 ↪"\nf1:        " +
#     str(dt_f1))
```

[ ]:
```

```