jaof9mly2

April 29, 2024

[ ]:
```
def fpgrowthFromFile(fname, minSupRatio, minConf):
    itemSetList, frequency = getFromFile(fname)
    minSup = len(itemSetList) * minSupRatio
    fpTree, headerTable = constructTree(itemSetList, frequency, minSup)

    freqItems = []
    mineTree(headerTable, minSup, set(), freqItems)
    rules = associationRule(freqItems, itemSetList, minConf)
    return freqItems, rules
```

[ ]:
```
df = pd.read_csv('Market_Basket_Optimisation.csv',sep=',',header=None)
df.fillna("", inplace=True)
df.head()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:

	0	1	2	3	4	\
0	shrimp	almonds	avocado	vegetables mix	green grapes	
1	burgers	meatballs	eggs			
2	chutney					
3	turkey	avocado				
4	mineral water	milk	energy bar	whole wheat rice	green tea	

	5	6	7	8	9	\
0	whole weat flour	yams	cottage cheese	energy drink	tomato juice	
1						
2						
3						
4						

	10	11	12	13	14	15	\
0	low fat yogurt	green tea	honey	salad	mineral water	salmon	
1							

```
2
3
4
```

```
                  16                17       18         19
0  antioxydant juice  frozen smoothie  spinach  olive oil
1
2
3
4
```

[ ]:
```
items = set()
for col in df:
    unique_values = df[col].unique()
    for val in unique_values:
        if val != "":
            items.add(val)

print(items)
```

```
{'frozen smoothie', 'yogurt cake', 'vegetables mix', 'low fat yogurt',
'ketchup', 'fromage blanc', 'cake', 'light cream', 'cream', 'chocolate', 'mint',
'black tea', 'light mayo', 'chocolate bread', 'sparkling water', 'salmon',
'green beans', 'fresh bread', 'energy bar', 'oatmeal', 'french fries',
'chicken', 'mushroom cream sauce', 'white wine', 'strong cheese', 'fresh tuna',
'salad', 'bramble', 'green grapes', 'green tea', 'avocado', 'parmesan cheese',
'cereals', 'butter', 'mineral water', 'spaghetti', 'protein bar', 'salt', 'pet
food', 'pancakes', 'napkins', 'milk', 'frozen vegetables', 'eggplant', 'soup',
'whole weat flour', 'burgers', 'tomato sauce', 'honey', 'strawberries', 'nonfat
milk', 'asparagus', 'oil', 'tomato juice', 'spinach', 'shampoo', 'carrots',
'dessert wine', 'rice', 'champagne', 'olive oil', 'gluten free bar',
'magazines', 'whole wheat pasta', 'clothes accessories', 'hand protein bar',
'pickles', 'corn', 'muffins', 'body spray', 'turkey', 'cauliflower', 'cooking
oil', 'ham', 'flax seed', 'cottage cheese', 'tea', 'eggs', 'almonds', 'melons',
'zucchini', 'antioxydant juice', 'yams', 'candy bars', 'chutney', 'extra dark
chocolate', 'pepper', 'grated cheese', 'gums', 'pasta', 'hot dogs', 'soda',
'shallot', 'energy drink', 'babies food', 'mint green tea', 'brownies', 'cider',
'tomatoes', 'red wine', 'ground beef', 'chili', 'french wine', 'bacon',
'cookies', 'toothpaste', 'whole wheat rice', 'sandwich', 'bug spray',
'meatballs', ' asparagus', 'burger sauce', 'barbecue sauce', 'mashed potato',
'shrimp', 'herb & pepper', 'escalope', 'water spray', 'blueberries',
'mayonnaise'}
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
```

```
  and should_run_async(code)
```

[ ]:
```
itemset = set(items)
encoded_vals = []
for index, row in df.iterrows():
    rowset = set(row)
    labels = {}
    uncommons = list(itemset - rowset)
    commons = list(itemset.intersection(rowset))
    for uc in uncommons:
        labels[uc] = 0
    for com in commons:
        labels[com] = 1
    encoded_vals.append(labels)
encoded_vals[0]
ohe_df = pd.DataFrame(encoded_vals)
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:
```
fp = fpgrowth(ohe_df, min_support=0.02, use_colnames=True)
fp
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
/usr/local/lib/python3.10/dist-
packages/mlxtend/frequent_patterns/fpcommon.py:110: DeprecationWarning:
DataFrames with non-bool types result in worse computationalperformance and
their support might be discontinued in the future.Please use a DataFrame with
bool type
  warnings.warn(
```

[ ]:

	support	itemsets
0	0.238368	(mineral water)
1	0.132116	(green tea)
2	0.076523	(low fat yogurt)
3	0.071457	(shrimp)
4	0.065858	(olive oil)
..	…	…
98	0.040928	(mineral water, ground beef)

```
99   0.039195      (spaghetti, ground beef)
100  0.021997           (ground beef, milk)
101  0.023064      (chocolate, ground beef)
102  0.027463         (mineral water, cake)

[103 rows x 2 columns]
```

[ ]:
```
association_rules(fp,metric="confidence",min_threshold=0.1)
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:
```
         antecedents      consequents  antecedent support  consequent support  \
0   (mineral water)      (green tea)            0.238368            0.132116
1       (green tea)  (mineral water)            0.132116            0.238368
2       (spaghetti)      (green tea)            0.174110            0.132116
3       (green tea)      (spaghetti)            0.132116            0.174110
4    (french fries)      (green tea)            0.170911            0.132116
..              …              …                   …                   …
89           (milk)    (ground beef)            0.129583            0.098254
90      (chocolate)    (ground beef)            0.163845            0.098254
91    (ground beef)      (chocolate)            0.098254            0.163845
92  (mineral water)           (cake)            0.238368            0.081056
93           (cake)  (mineral water)            0.081056            0.238368

     support  confidence      lift  leverage  conviction  zhangs_metric
0   0.031063    0.130313  0.986357 -0.000430    0.997927      -0.017837
1   0.031063    0.235116  0.986357 -0.000430    0.995748      -0.015688
2   0.026530    0.152374  1.153335  0.003527    1.023900       0.160977
3   0.026530    0.200807  1.153335  0.003527    1.033405       0.153188
4   0.028530    0.166927  1.263488  0.005950    1.041786       0.251529
..       …           …       …         …           …              …
89  0.021997    0.169753  1.727704  0.009265    1.086118       0.483903
90  0.023064    0.140765  1.432669  0.006965    1.049476       0.361180
91  0.023064    0.234735  1.432669  0.006965    1.092635       0.334908
92  0.027463    0.115213  1.421397  0.008142    1.038604       0.389252
93  0.027463    0.338816  1.421397  0.008142    1.151921       0.322617

[94 rows x 10 columns]
```

[ ]:
```

```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
```

```
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-12-6db5754b978d> in <cell line: 2>()
      1 minTransaction = 300
----> 2 totalTransactions = len(basket_sets.index)
      3 min_support_calc = minTransaction/totalTransactions
      4
      5 print('number of baskets for analysis is', totalTransactions)

NameError: name 'basket_sets' is not defined
```

Decision Tree classification

[ ]:
```
df1=pd.read_csv("diabetes_dataset.csv")
df1.head()
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	\
0	6	148	72	35	0	33.6	
1	1	85	66	29	0	26.6	
2	8	183	64	0	0	23.3	
3	1	89	66	23	94	28.1	
4	0	137	40	35	168	43.1	

	DiabetesPedigreeFunction	Age	Outcome
0	0.627	50	1
1	0.351	31	0
2	0.672	32	1
3	0.167	21	0
4	2.288	33	1

[ ]:
```
x=df1.drop(['Outcome'], axis=1)
x
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
```

```
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:
```
     Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  \
0              6      148             72             35        0  33.6
1              1       85             66             29        0  26.6
2              8      183             64              0        0  23.3
3              1       89             66             23       94  28.1
4              0      137             40             35      168  43.1
..           …      …              …              …       …   …
763           10      101             76             48      180  32.9
764            2      122             70             27        0  36.8
765            5      121             72             23      112  26.2
766            1      126             60              0        0  30.1
767            1       93             70             31        0  30.4

     DiabetesPedigreeFunction  Age
0                       0.627   50
1                       0.351   31
2                       0.672   32
3                       0.167   21
4                       2.288   33
..                        …  …
763                     0.171   63
764                     0.340   27
765                     0.245   30
766                     0.349   47
767                     0.315   23

[768 rows x 8 columns]
```

[ ]:
```
#target variable
y=df1.Outcome
y
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:
```
0      1
1      0
2      1
3      0
4      1
```

```
      ..
763    0
764    0
765    0
766    1
767    0
Name: Outcome, Length: 768, dtype: int64
```

[ ]:
```
from sklearn.tree import DecisionTreeClassifier # Import Decision Tree␣
 ↪Classifier
from sklearn.model_selection import train_test_split # Import train_test_split␣
 ↪function
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2,␣
 ↪random_state=1)
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:
```
# Create Decision Tree classifer object
model = DecisionTreeClassifier()

# Train Decision Tree Classifer
model = model.fit(x_train,y_train)

#Predict the response for test dataset
y_pred = model.predict(x_test)
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
/usr/local/lib/python3.10/dist-packages/pandas/core/dtypes/cast.py:1641:
DeprecationWarning: np.find_common_type is deprecated.  Please use
`np.result_type` or `np.promote_types`.
See https://numpy.org/devdocs/release/1.25.0-notes.html and the docs for more
information.  (Deprecated NumPy 1.25)
  return np.find_common_type(types, [])
/usr/local/lib/python3.10/dist-packages/pandas/core/dtypes/cast.py:1641:
DeprecationWarning: np.find_common_type is deprecated.  Please use
`np.result_type` or `np.promote_types`.
See https://numpy.org/devdocs/release/1.25.0-notes.html and the docs for more
```

```
information.  (Deprecated NumPy 1.25)
  return np.find_common_type(types, [])
```

[ ]:
```
#Evaluation using Accuracy score
from sklearn import metrics #Import scikit-learn metrics module for accuracy␣
 ↪calculation
print("Accuracy:",metrics.accuracy_score(y_test, y_pred)*100)
```

```
Accuracy: 67.53246753246754

/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:
```
#Evaluation using Confusion matrix
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,y_pred)
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:
```
array([[75, 24],
       [26, 29]])
```

[ ]:
```
print("Accuracy:",((82+27)/154))
```

```
Accuracy: 0.7077922077922078

/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:
```
#Evaluation using Classification report
from sklearn.metrics import classification_report
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.74      0.76      0.75        99
```

```
           1       0.55      0.53      0.54        55

    accuracy                           0.68       154
   macro avg       0.64      0.64      0.64       154
weighted avg       0.67      0.68      0.67       154
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:
```
#checking prediction value
model.predict([[6,148,72,35,0,33.6,0.627,50]])
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does
not have valid feature names, but DecisionTreeClassifier was fitted with feature
names
  warnings.warn(
```

[ ]:
```
array([1])
```

[ ]:
```
#Import modules for Visualizing Decision trees
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:
```
import six
import sys
sys.modules['sklearn.externals.six']=six
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
```

```
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:
```
features=x.columns
features
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]:
```
Index(['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin',
       'BMI', 'DiabetesPedigreeFunction', 'Age'],
      dtype='object')
```

[ ]:
```
dot_data = StringIO()
export_graphviz(model, out_file=dot_data,filled=True,␣
 ↪rounded=True,special_characters=True,feature_names =␣
 ↪features,class_names=['0','1'])
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('diabetes_set.png')
Image(graph.create_png())
```

[ ]:
```
# Create Decision Tree classifer object
model = DecisionTreeClassifier(criterion="entropy", max_depth=3)

# Train Decision Tree Classifer
model = model.fit(x_train,y_train)

#Predict the response for test dataset
y_pred = model.predict(x_test)

# Model Accuracy
print("Accuracy:",metrics.accuracy_score(y_test, y_pred)*100)
```

[ ]:
```
from sklearn.externals.six import StringIO
from IPython.display import Image
from sklearn.tree import export_graphviz
import pydotplus
dot_data = StringIO()
export_graphviz(model, out_file=dot_data,filled=True,␣
 ↪rounded=True,special_characters=True, feature_names =␣
 ↪features,class_names=['0','1'])
```

```
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('diabetes_set.png')
Image(graph.create_png())
```

```
/usr/local/lib/python3.10/dist-packages/ipykernel/ipkernel.py:283:
DeprecationWarning: `should_run_async` will not call `transform_cell`
automatically in the future. Please pass the result to `transformed_cell`
argument and any exception that happen during thetransform in
`preprocessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```

[ ]: