Chittagong University of Engineering & Technology

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CT-2

Association Rule Mining Using Apriori Algorithm for Drepression Analysis

COURSE NO: CSE-463

COURSE TITLE: Machine Learning

Submitted By:

Name: Md. Raufur Rahman

ID: 1604012

Level-Term: 4-1

Submitted To:

DR. MD. IQBAL HASAN SARKAR

Assistant professor, Dept of CSE,CUET

Remarks

Problem Definition:

Association rule mining using Apriori algorithm for depression analysis.

Sample Dataset:

This depression dataset consists of 20 features. Shape of this dataset is 130rows, 20columns. The features are current academic year, age, gender, relationship status, financial state, sleeping time, understanding with family, current feelings, satisfied with result, feeling pressure, cop up with environment, location, happy with living place, inferiority complex, satisfied with meal etc. Every input value will be converted into numeric value to process.The following view represents the dataset

	Which year are you in?	How are you feeling right now?	On a scale of 1-100, how would you express this feeling?	Gender	Age	Your location ?	Relationship status	Are you happy with your financial state?	How much have you succeeded to cope up with the environment of your educational institution?	Understanding with your family members?	Are you feeling pressure in your study or work right now?	Are you satisfied with your academic result?	Are you happy with your living place?	Who supports you when you are not succeeding in your academic life?	Hav yo use an soci med withi th last hours
0	4th year	Normal	50	Male	23	Home	Single	Yes	4	Normal	No	Yes	No	Family	Ye
1	4th year	Normal	40	Male	23	Home	Single	Yes	4	Normal	Yes	Yes	No	Family	Ye
2	4th year	Bad	50	Male	23	Home	Single	Yes	4	Normal	Yes	Yes	No	Family	Ye
3	4th year	Normal	45	Male	23	Home	Single	Yes	4	Normal	No	Yes	No	Family	Ye
4	4th year	Normal	50	Male	23	Home	Single	Yes	4	Normal	Yes	Yes	Yes	Family	Ye
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
125	4th year	Very good	91	Male	23	Department	Single	Yes	4	Good	No	No	Yes	Family	Ye
126	4th year	Very good	100	Male	23	Home	Single	Yes	3	Normal	No	Yes	Yes	Friends	Ye
127	4th year	Normal	69	Male	23	Playground	Single	No	5	Good	Yes	Yes	Yes	Family	Ye

Implementation:

To analysis we use apriori alorithm. Using this we can find find frequent dataset and assocaition rules.

```python
# Association Rule Mining Using Apriori Algorithm

# Import necessary libraries
import pandas as pd
import numpy as np
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules

data = pd.read_csv('Copy of Depression and Happiness Factor
Analysis.csv')
data_new = data.iloc[:, 1:].copy() # copy all data excluding 1st column
# print(data_new)
# print(data_new.isnull())

# Preprocessing
info = data_new.to_numpy().tolist() # convert data to list
# print(info)

# Convert all values to string format
for values in info:
for i in range(len(values)):
s = ''
if type(values[i]) ≠ str:
s = str(values[i])
values[i] = s


tr = TransactionEncoder()
# Encodeing data to boolean values of array
t_info = tr.fit(info).transform(info)
#print(t_info)
```

```python
df = pd.DataFrame(t_info, columns=tr.columns_) # create a dataframe

#print(df)

# Using Apriori algo to create frequent items
frequent_items = apriori(df, min_support=0.60, use_colnames=True)
# print(frequent_items)

# Association Rule Mining
result = association_rules(
frequent_items, metric='confidence', min_threshold=.60)
new_result = result.loc[:, ['antecedents',
'consequents', 'support', 'confidence']]

print(new_result)
```

Output:

frequent_items

	support	itemsets
0	1.000000	(23)
1	1.000000	(4th year)
2	0.600000	(Family)
3	1.000000	(Male)
4	0.946154	(No)
...	...	...
122	0.600000	(23, 4th year, Yes, Family, Single, No)
123	0.946154	(23, 4th year, Yes, Male, Single, No)
124	0.600000	(23, Yes, Family, Male, Single, No)
125	0.600000	(4th year, Yes, Family, Male, Single, No)
126	0.600000	(23, 4th year, Yes, Family, Male, Single, No)

127 rows × 2 columns

```
new_result
```

	antecedents	consequents	support	confidence
0	(23)	(4th year)	1.0	1.000000
1	(4th year)	(23)	1.0	1.000000
2	(23)	(Family)	0.6	0.600000
3	(Family)	(23)	0.6	1.000000
4	(23)	(Male)	1.0	1.000000
...	...	...	...	...
1927	(Yes)	(23, 4th year, Family, Male, Single, No)	0.6	0.600000
1928	(Family)	(23, 4th year, Yes, Male, Single, No)	0.6	1.000000
1929	(Male)	(23, 4th year, Yes, Family, Single, No)	0.6	0.600000
1930	(Single)	(23, 4th year, Yes, Family, Male, No)	0.6	0.600000
1931	(No)	(23, 4th year, Yes, Family, Male, Single)	0.6	0.634146

1932 rows × 4 columns

Discussion:

To implement this algorithm we use minimum support value 0.6. By implementing apriori alorithm we get frequent itemsets which has 127 items. To get association rules we use minimum confidence value 0.6. By using this itemsets we get 1932 association rules with support and confidence. From this we can see maximum value of confidence is 1 and minimum value is .6341. Maximum value of support is 1 and minimum value is 0.6