

CB.EN.U4CSE17059-Lab4ML&DMAssignmentAugust18

August 18, 2020

```
In [1]: import pandas as pd
import numpy as np
```

```
In [2]: df = pd.read_csv('GroceryStoreDataSet.csv',header = None)
```

```
In [3]: df[0] = df[0].apply(lambda x: x.split(','))
```

```
In [4]: df.head()
```

```
Out[4]:
```

| | 0 |
|---|------------------------------------|
| 0 | [MILK, BREAD, BISCUIT] |
| 1 | [BREAD, MILK, BISCUIT, CORNFLAKES] |
| 2 | [BREAD, TEA, BOURNVITA] |
| 3 | [JAM, MAGGI, BREAD, MILK] |
| 4 | [MAGGI, TEA, BISCUIT] |

```
In [5]: a = []
for x in df[0]:
    a.append(x)
df = a
```

```
In [6]: import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

```
In [7]: import pandas as pd
from mlxtend.preprocessing import TransactionEncoder

te = TransactionEncoder()
te_ary = te.fit(df).transform(df)
df = pd.DataFrame(te_ary, columns=te.columns_)
df
```

```
Out[7]:
```

| | BISCUIT | BOURNVITA | BREAD | COCK | COFFEE | CORNFLAKES | JAM | MAGGI | MILK | \ |
|---|---------|-----------|-------|-------|--------|------------|-------|-------|-------|---|
| 0 | True | False | True | False | False | False | False | False | True | |
| 1 | True | False | True | False | False | True | False | False | True | |
| 2 | False | True | True | False | False | False | False | False | False | |

| | | | | | | | | | |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 3 | False | False | True | False | False | False | True | True | True |
| 4 | True | False | False | False | False | False | False | True | False |
| 5 | False | True | True | False | False | False | False | False | False |
| 6 | False | False | False | False | False | True | False | True | False |
| 7 | True | False | True | False | False | False | False | True | False |
| 8 | False | False | True | False | False | False | True | True | False |
| 9 | False | False | True | False | False | False | False | False | True |
| 10 | True | False | False | True | True | True | False | False | False |
| 11 | True | False | False | True | True | True | False | False | False |
| 12 | False | True | False | False | True | False | False | False | False |
| 13 | False | False | True | True | True | False | False | False | False |
| 14 | True | False | True | False | False | False | False | False | False |
| 15 | False | False | False | False | True | True | False | False | False |
| 16 | False | True | True | False | False | False | False | False | False |
| 17 | False | False | True | False | True | False | False | False | False |
| 18 | False | False | True | False | True | False | False | False | False |
| 19 | False | False | False | False | True | True | False | False | True |

| | SUGER | TEA |
|----|-------|-------|
| 0 | False | False |
| 1 | False | False |
| 2 | False | True |
| 3 | False | False |
| 4 | False | True |
| 5 | False | True |
| 6 | False | True |
| 7 | False | True |
| 8 | False | True |
| 9 | False | False |
| 10 | False | False |
| 11 | False | False |
| 12 | True | False |
| 13 | False | False |
| 14 | True | False |
| 15 | True | False |
| 16 | True | False |
| 17 | True | False |
| 18 | True | False |
| 19 | False | True |

```
In [8]: from mlxtend.frequent_patterns import apriori

apriori(df, min_support=0.2, use_colnames=True)
```

```
Out[8]:
```

| | support | itemsets |
|---|---------|-------------|
| 0 | 0.35 | (BISCUIT) |
| 1 | 0.20 | (BOURNVITA) |
| 2 | 0.65 | (BREAD) |

| | | |
|----|------|----------------------|
| 3 | 0.40 | (COFFEE) |
| 4 | 0.30 | (CORNFLAKES) |
| 5 | 0.25 | (MAGGI) |
| 6 | 0.25 | (MILK) |
| 7 | 0.30 | (SUGER) |
| 8 | 0.35 | (TEA) |
| 9 | 0.20 | (BISCUIT, BREAD) |
| 10 | 0.20 | (MILK, BREAD) |
| 11 | 0.20 | (SUGER, BREAD) |
| 12 | 0.20 | (TEA, BREAD) |
| 13 | 0.20 | (CORNFLAKES, COFFEE) |
| 14 | 0.20 | (COFFEE, SUGER) |
| 15 | 0.20 | (MAGGI, TEA) |

```
In [9]: frequent_itemsets = apriori(df, min_support=0.2, use_colnames=True)
frequent_itemsets['length'] = frequent_itemsets['itemsets'].apply(lambda x: len(x))
frequent_itemsets
```

```
Out[9]:
```

| | support | itemsets | length |
|----|---------|----------------------|--------|
| 0 | 0.35 | (BISCUIT) | 1 |
| 1 | 0.20 | (BOURNVITA) | 1 |
| 2 | 0.65 | (BREAD) | 1 |
| 3 | 0.40 | (COFFEE) | 1 |
| 4 | 0.30 | (CORNFLAKES) | 1 |
| 5 | 0.25 | (MAGGI) | 1 |
| 6 | 0.25 | (MILK) | 1 |
| 7 | 0.30 | (SUGER) | 1 |
| 8 | 0.35 | (TEA) | 1 |
| 9 | 0.20 | (BISCUIT, BREAD) | 2 |
| 10 | 0.20 | (MILK, BREAD) | 2 |
| 11 | 0.20 | (SUGER, BREAD) | 2 |
| 12 | 0.20 | (TEA, BREAD) | 2 |
| 13 | 0.20 | (CORNFLAKES, COFFEE) | 2 |
| 14 | 0.20 | (COFFEE, SUGER) | 2 |
| 15 | 0.20 | (MAGGI, TEA) | 2 |

```
In [10]: frequent_itemsets[ (frequent_itemsets['length'] == 1) &
(frequent_itemsets['support'] >= 0.) ]
```

```
Out[10]:
```

| | support | itemsets | length |
|---|---------|--------------|--------|
| 0 | 0.35 | (BISCUIT) | 1 |
| 1 | 0.20 | (BOURNVITA) | 1 |
| 2 | 0.65 | (BREAD) | 1 |
| 3 | 0.40 | (COFFEE) | 1 |
| 4 | 0.30 | (CORNFLAKES) | 1 |
| 5 | 0.25 | (MAGGI) | 1 |
| 6 | 0.25 | (MILK) | 1 |
| 7 | 0.30 | (SUGER) | 1 |
| 8 | 0.35 | (TEA) | 1 |

```
In [12]: frequent_itemsets[ (frequent_itemsets['length'] == 2) &
                             (frequent_itemsets['support'] >= 0.2) ]
```

```
Out[12]:
```

| | support | itemsets | length |
|----|---------|----------------------|--------|
| 9 | 0.2 | (BISCUIT, BREAD) | 2 |
| 10 | 0.2 | (MILK, BREAD) | 2 |
| 11 | 0.2 | (SUGER, BREAD) | 2 |
| 12 | 0.2 | (TEA, BREAD) | 2 |
| 13 | 0.2 | (CORNFLAKES, COFFEE) | 2 |
| 14 | 0.2 | (COFFEE, SUGER) | 2 |
| 15 | 0.2 | (MAGGI, TEA) | 2 |

```
In [13]: frequent_itemsets[ frequent_itemsets['itemsets'] == {'BREAD', 'MILK'} ]
```

```
Out[13]:
```

| | support | itemsets | length |
|----|---------|---------------|--------|
| 10 | 0.2 | (MILK, BREAD) | 2 |

```
In [ ]:
```