titanic

April 17, 2024

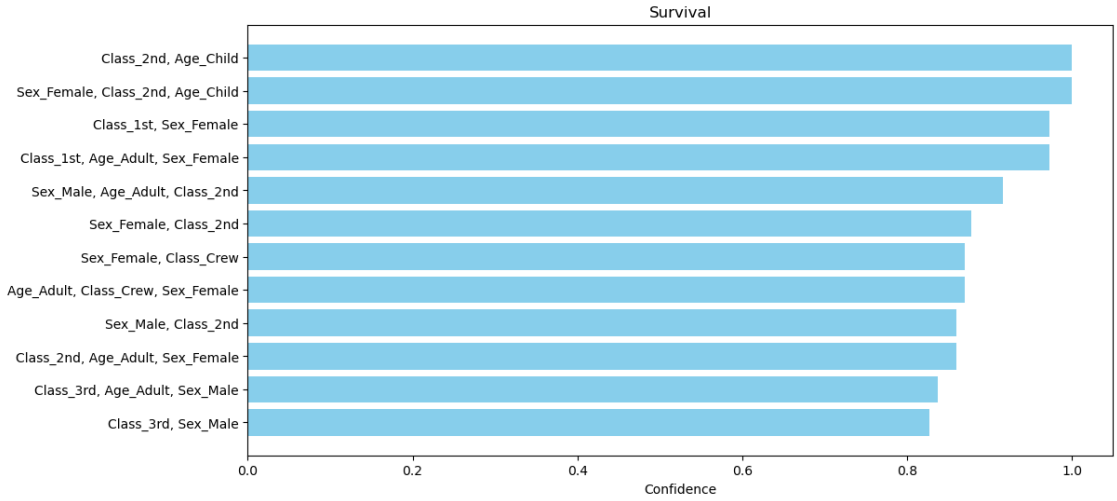
0.1 Titanic with Apriori

[23]:
```
import pandas as pd # type: ignore
import numpy as np # type: ignore
from mlxtend.frequent_patterns import apriori, association_rules
import matplotlib.pyplot as plt

df = pd.read_csv('titanic.csv')
df = df.drop(df.columns[0], axis=1)

data = pd.get_dummies(df, columns=['Class', 'Sex', 'Age', 'Survived'])

#print(data.head())

freq = apriori(data, min_support=0.005, use_colnames=True, verbose=1)

#print(freq.head(20))

rules = association_rules(freq, metric="confidence", min_threshold=0.8)
rules = rules.sort_values('confidence', ascending=False)
print(rules)

# Filtrowanie reguł dotyczących przeżywalności
survival_rules = rules[rules['consequents'].apply(lambda x: ('Survived_Yes' in␣
 ↪str(x) or 'Survived_No' in str(x)) and 'Age_Adult' not in str(x))]

# Sortowanie reguł wg ufności
survival_rules = survival_rules.sort_values('confidence', ascending=False)

# Wyświetlenie najciekawszych reguł
print(survival_rules)

# Utworzenie wykresu słupkowego
plt.figure(figsize=(12, 6))
plt.barh(range(len(survival_rules)), survival_rules['confidence'],␣
 ↪color='skyblue')
plt.yticks(range(len(survival_rules)), survival_rules['antecedents'].
 ↪apply(lambda x: ', '.join(list(x))))
```

```
plt.xlabel('Confidence')
plt.title('Survival')
plt.gca().invert_yaxis()  # Odwróć oś Y, aby najważniejsze reguły były na górze
plt.show()
```

```
Processing 40 combinations | Sampling itemset size 54
                                antecedents     consequents  antecedent support  \
65  (Survived_No, Sex_Female, Age_Child)     (Class_3rd)            0.007724
22                (Class_2nd, Age_Child)  (Survived_Yes)            0.010904
57    (Sex_Male, Survived_No, Class_2nd)     (Age_Adult)            0.069968
70    (Sex_Male, Survived_No, Age_Child)     (Class_3rd)            0.015902
56    (Sex_Female, Class_2nd, Age_Child)  (Survived_Yes)            0.005906
..                                    …            …                 …
62  (Class_3rd, Survived_No, Sex_Female)     (Age_Adult)            0.048160
67      (Class_3rd, Age_Adult, Sex_Male)   (Survived_No)            0.209905
26                 (Class_3rd, Sex_Male)   (Survived_No)            0.231713
63  (Survived_No, Age_Adult, Sex_Female)     (Class_3rd)            0.049523
66   (Class_3rd, Survived_No, Age_Adult)      (Sex_Male)            0.216265

    consequent support   support  confidence      lift  leverage  conviction  \
65            0.320763  0.007724    1.000000  3.117564  0.005246         inf
22            0.323035  0.010904    1.000000  3.095640  0.007382         inf
57            0.950477  0.069968    1.000000  1.052103  0.003465         inf
70            0.320763  0.015902    1.000000  3.117564  0.010801         inf
56            0.323035  0.005906    1.000000  3.095640  0.003998         inf
..                 …         …           …         …         …           …
62            0.950477  0.040436    0.839623  0.883370 -0.005339    0.308790
67            0.676965  0.175829    0.837662  1.237379  0.033731    1.989896
26            0.676965  0.191731    0.827451  1.222295  0.034870    1.872135
63            0.320763  0.040436    0.816514  2.545534  0.024551    3.701840
66            0.786461  0.175829    0.813025  1.033777  0.005745    1.142075

    zhangs_metric
65       0.684524
22       0.684428
57       0.053249
70       0.690212
56       0.680987
..            …
62      -0.121813
67       0.242806
26       0.236717
63       0.638790
66       0.041690

[80 rows x 10 columns]
                           antecedents     consequents  antecedent support  \
```

```
22                (Class_2nd, Age_Child)  (Survived_Yes)           0.010904
56   (Sex_Female, Class_2nd, Age_Child)  (Survived_Yes)           0.005906
11              (Class_1st, Sex_Female)  (Survived_Yes)           0.065879
46   (Class_1st, Age_Adult, Sex_Female)  (Survived_Yes)           0.065425
58     (Sex_Male, Age_Adult, Class_2nd)   (Survived_No)           0.076329
17              (Sex_Female, Class_2nd)  (Survived_Yes)           0.048160
31             (Sex_Female, Class_Crew)  (Survived_Yes)           0.010450
72  (Age_Adult, Class_Crew, Sex_Female)  (Survived_Yes)           0.010450
19                (Sex_Male, Class_2nd)   (Survived_No)           0.081327
55   (Class_2nd, Age_Adult, Sex_Female)  (Survived_Yes)           0.042254
67     (Class_3rd, Age_Adult, Sex_Male)   (Survived_No)           0.209905
26                (Class_3rd, Sex_Male)   (Survived_No)           0.231713

    consequent support   support  confidence      lift  leverage  conviction  \
22            0.323035  0.010904    1.000000  3.095640  0.007382         inf
56            0.323035  0.005906    1.000000  3.095640  0.003998         inf
11            0.323035  0.064062    0.972414  3.010243  0.042781   24.539982
46            0.323035  0.063607    0.972222  3.009650  0.042473   24.370741
58            0.676965  0.069968    0.916667  1.354083  0.018296    3.876420
17            0.323035  0.042254    0.877358  2.715986  0.026696    5.519869
31            0.323035  0.009087    0.869565  2.691861  0.005711    5.190065
72            0.323035  0.009087    0.869565  2.691861  0.005711    5.190065
19            0.676965  0.069968    0.860335  1.270871  0.014913    2.312930
55            0.323035  0.036347    0.860215  2.662916  0.022698    4.842904
67            0.676965  0.175829    0.837662  1.237379  0.033731    1.989896
26            0.676965  0.191731    0.827451  1.222295  0.034870    1.872135

    zhangs_metric
22       0.684428
56       0.680987
11       0.714898
46       0.714480
58       0.283102
17       0.663777
31       0.635147
72       0.635147
19       0.232006
55       0.652022
67       0.242806
26       0.236717
```


Survival
Class_2nd, Age_Child
Sex_Female, Class_2nd, Age_Child
Class_1st, Sex_Female
Class_1st, Age_Adult, Sex_Female
Sex_Male, Age_Adult, Class_2nd
Sex_Female, Class_2nd
Sex_Female, Class_Crew
Age_Adult, Class_Crew, Sex_Female
Sex_Male, Class_2nd
Class_2nd, Age_Adult, Sex_Female
Class_3rd, Age_Adult, Sex_Male
Class_3rd, Sex_Male
0.0
0.2
0.4
0.6
0.8
1.0
Confidence