**CIS 6930: Introduction to Data Mining**

**Individual Project 2 – Report**

In this project the **Apriori** association rule mining technique is applied to the following datasets.

1. Titanic Dataset
2. Retail Dataset
3. Game of Thrones Dataset

The rule sets for each dataset and the corresponding observations are as follows,

**Titanic Dataset**

1. *All the rules obtained*

**lhs rhs support confidence lift**

1 {} => {Age=Adult} 0.9504771 0.9504771 1.0000000

2 {Class=2nd} => {Age=Adult} 0.1185825 0.9157895 0.9635051

3 {Class=1st} => {Age=Adult} 0.1449341 0.9815385 1.0326798

4 {Sex=Female} => {Age=Adult} 0.1930940 0.9042553 0.9513700

5 {Class=3rd} => {Age=Adult} 0.2848705 0.8881020 0.9343750

6 {Survived=Yes} => {Age=Adult} 0.2971377 0.9198312 0.9677574

7 {Class=Crew} => {Sex=Male} 0.3916402 0.9740113 1.2384742

8 {Class=Crew} => {Age=Adult} 0.4020900 1.0000000 1.0521033

9 {Survived=No} => {Sex=Male} 0.6197183 0.9154362 1.1639949

10 {Survived=No} => {Age=Adult} 0.6533394 0.9651007 1.0153856

11 {Sex=Male} => {Age=Adult} 0.7573830 0.9630272 1.0132040

12 {Sex=Female,Survived=Yes} => {Age=Adult} 0.1435711 0.9186047 0.9664669

13 {Class=3rd,Sex=Male} => {Survived=No} 0.1917310 0.8274510 1.2222950

14 {Class=3rd,Survived=No} => {Age=Adult} 0.2162653 0.9015152 0.9484870

15 {Class=3rd,Sex=Male} => {Age=Adult} 0.2099046 0.9058824 0.9530818

16 {Sex=Male,Survived=Yes} => {Age=Adult} 0.1535666 0.9209809 0.9689670

17 {Class=Crew,Survived=No} => {Sex=Male} 0.3044071 0.9955423 1.2658514

18 {Class=Crew,Survived=No} => {Age=Adult} 0.3057701 1.0000000 1.0521033

19 {Class=Crew,Sex=Male} => {Age=Adult} 0.3916402 1.0000000 1.0521033

20 {Class=Crew,Age=Adult} => {Sex=Male} 0.3916402 0.9740113 1.2384742

21 {Sex=Male,Survived=No} => {Age=Adult} 0.6038164 0.9743402 1.0251065

22 {Age=Adult,Survived=No} => {Sex=Male} 0.6038164 0.9242003 1.1751385

23 {Class=3rd,Sex=Male,Survived=No} => {Age=Adult} 0.1758292 0.9170616 0.9648435

24 {Class=3rd,Age=Adult,Survived=No} => {Sex=Male} 0.1758292 0.8130252 1.0337773

25 {Class=3rd,Sex=Male,Age=Adult} => {Survived=No} 0.1758292 0.8376623 1.2373791

26 {Class=Crew,Sex=Male,Survived=No} => {Age=Adult} 0.3044071 1.0000000 1.0521033

27 {Class=Crew,Age=Adult,Survived=No} => {Sex=Male} 0.3044071 0.9955423 1.2658514

1. *Rules meeting the requested criteria i.e.*

*Support = 0.01 Confidence = 0.90*

**lhs rhs support confidence lift**

1 {Class=2nd,Age=Child} => {Survived=Yes} 0.01090413 1.0000000 3.095640

2 {Class=1st,Sex=Female} => {Survived=Yes} 0.06406179 0.9724138 3.010243

3 {Class=2nd,Sex=Male,Age=Adult} => {Survived=No} 0.06996820 0.9166667 1.354083

4 {Class=1st,Sex=Female,Age=Adult} => {Survived=Yes} 0.06360745 0.9722222 3.009650

1. *Rules after removing redundant rules and listing unique ones*

**lhs rhs support confidence lift**

1 {Class=2nd,Age=Child} => {Survived=Yes} 0.01090413 1.0000000 3.095640

2 {Class=1st,Sex=Female} => {Survived=Yes} 0.06406179 0.9724138 3.010243

3 {Class=2nd,Sex=Male,Age=Adult} => {Survived=No} 0.06996820 0.9166667 1.354083

1. *Remaining rules sorted based on their lifts*

**lhs rhs support confidence lift**

1 {Class=2nd,Age=Child} => {Survived=Yes} 0.01090413 1.0000000 3.095640

2 {Class=1st,Sex=Female} => {Survived=Yes} 0.06406179 0.9724138 3.010243

3 {Class=2nd,Sex=Male,Age=Adult} => {Survived=No} 0.06996820 0.9166667 1.354083

1. *Observations*

The dataset contains four attributes: Class, Age, Sex, Survived. The above generated rules are used to show the Survived=YES or NO in the RHS with respect to the LHS. We find that only 4 rules match our criteria for support and confidence. After removing the redundant rules we are left with only 3 rules. The rules are used to predict whether a particular class and gender of people in the Titanic survive or not. Given the rules generated, it is safe to say that most of the women and children in the 1st and 2nd classes survived but only a few men had survived. Also most of the people in the 3rd class of the ship, regardless of their gender, did not survive.

**Retail Dataset**

1. *All the rules obtained*

Please refer script output

1. *Rules meeting the requested criteria i.e.*

*Support = 0.01 Confidence = 0.90 (Listing only a few as sample rules since the actual number of rules is very large. Refer script output for the full list)*

**lhs rhs support confidence lift**

{Bread,

CannedGoods,

Dairy,

Drink,

FrozenFood,

Fruit,

Tobacco,

Vegetable} => {PersonalCare} 0.0252 0.9473684 1.636498

{Bread,

CannedGoods,

Dairy,

Drink,

FrozenFood,

Fruit,

PaperGoods,

Vegetable} => {PersonalCare} 0.0136 0.9444444 1.631447

{Bread,

Dairy,

Drink,

FrozenFood,

Fruit,

PaperGoods,

Tobacco,

Vegetable} => {PersonalCare} 0.0191 0.9597990 1.657970

{Bread,

CannedGoods,

Dairy,

Drink,

FrozenFood,

Fruit,

PaperGoods,

Tobacco,

Vegetable} => {PersonalCare} 0.0108 0.9391304 1.622267

1. *Rules after removing redundant rules and listing unique ones*

**lhs rhs support confidence lift**

1 {Drink} => {PersonalCare} 0.2921 0.9675389 1.67134

1. *Remaining rules sorted based on their lifts*

**lhs rhs support confidence lift**

1 {Drink} => {PersonalCare} 0.2921 0.9675389 1.67134

1. *Observations*

The retail dataset contains 10000 transactions in a retail store. The rules generated above convey that the list of items of the LHS would result in the purchase of the items in the RHS. We consider only the items “Beverage”, “Meat” and “PersonalCare” as the possible items in the RHS. An interesting result of the association rules mining above is that the rules satisfying the given support and confidence criteria indicate that only “PersonalCare” seems to be the item which is purchased with respect to the items in the LHS. This means that “PersonalCare” is bought more frequently that that of “Beverage” or “Meat” with respect to the item in the LHS.

The final rule set in (iii) above indicates that most people who bought drinks also purchased “PersonalCare” items, as this seems to be the most frequent rule in the dataset

**Game of Thrones Dataset**

1. *All the rules obtained*

Please refer script output

1. *Rules meeting the requested criteria i.e.*

*Support = 0.01 Confidence = 0.90(Listing only a few as sample rules since the actual number of rules is very large. Refer script output for the full list)*

**lhs rhs support confidence lift**

{Gender=F,

Nobility=1,

Book1=0,

Book2=0,

Book3=1,

Book4=1,

Book5=0} => {Survives=1} 0.01308615 1.0000000 1.498366

{Gender=F,

Nobility=1,

Book1=0,

Book2=0,

Book3=0,

Book4=0,

Book5=1} => {Survives=1} 0.01308615 1.0000000 1.498366

{Gender=M,

Nobility=1,

Book1=0,

Book2=0,

Book3=0,

Book4=1,

Book5=0} => {Survives=1} 0.05561614 0.9444444 1.415123

{House=Greyjoy,

Gender=M,

Nobility=1,

Book1=0,

Book2=0,

Book3=0,

Book4=1,

Book5=0} => {Survives=1} 0.02617230 0.9600000 1.438431

1. *Rules after removing redundant rules and listing unique ones*

**lhs rhs support confidence lift**

1 {House=HouseTyrell} => {Survives=1} 0.01090513 0.9090909 1.362151

2 {House=HouseMartell} => {Survives=1} 0.01199564 0.9166667 1.373502

3 {House=Tyrell} => {Survives=1} 0.01526718 0.9333333 1.398475

4 {House=Martell} => {Survives=1} 0.02508179 0.9200000 1.378497

14 {House=Targaryen,Book5=1} => {Survives=1} 0.01308615 0.9230769 1.383107

15 {House=Arryn,Book4=1} => {Survives=1} 0.01635769 0.9375000 1.404718

16 {House=Arryn,Nobility=1} => {Survives=1} 0.01853871 0.9444444 1.415123

17 {House=Arryn,Book3=0} => {Survives=1} 0.01962923 0.9473684 1.419505

18 {House=Arryn,Book1=0} => {Survives=1} 0.01199564 1.0000000 1.498366

24 {House=Greyjoy,Book4=1} => {Survives=1} 0.03925845 0.9230769 1.383107

25 {House=Greyjoy,Book2=0} => {Survives=1} 0.04143948 0.9268293 1.388729

26 {House=Stark,Gender=F} => {Survives=1} 0.02071974 0.9047619 1.355664

27 {House=Stark,Book5=1} => {Survives=1} 0.02944384 1.0000000 1.498366

28 {House=Lannister,Book4=1} => {Survives=1} 0.04252999 0.9285714 1.391340

29 {House=Lannister,Book5=1} => {Survives=1} 0.01199564 1.0000000 1.498366

30 {Gender=F,Book4=1} => {Survives=1} 0.05125409 0.9591837 1.437208

31 {Nobility=1,Book4=1} => {Survives=1} 0.14721919 0.9183673 1.376050

32 {Book3=0,Book4=1} => {Survives=1} 0.15703381 0.9056604 1.357011

46 {House=Tully,Nobility=1,Book2=0} => {Survives=1} 0.01308615 0.9230769 1.383107

47 {House=Tully,Gender=M,Book2=0} => {Survives=1} 0.01308615 0.9230769 1.383107

88 {House=Baratheon,Nobility=0,Book1=0} => {Survives=1} 0.01199564 0.9166667 1.373502

107 {House=Lannister,Nobility=1,Book2=0} => {Survives=1} 0.02726281 0.9615385 1.440737

108 {House=Lannister,Book1=0,Book2=0} => {Survives=1} 0.02508179 0.9583333 1.435934

109 {House=Lannister,Gender=M,Book1=0} => {Survives=1} 0.04252999 0.9069767 1.358983

110 {House=NightsWatch,Book1=1,Book5=1} => {Survives=1} 0.01308615 1.0000000 1.498366

120 {House=None,Gender=F,Nobility=1} => {Survives=1} 0.02181025 0.9523810 1.427015

121 {Gender=F,Book3=1,Book5=1} => {Survives=1} 0.02071974 0.9500000 1.423448

122 {Gender=F,Nobility=1,Book5=1} => {Survives=1} 0.03162486 0.9666667 1.448420

123 {Gender=F,Nobility=1,Book2=0} => {Survives=1} 0.06106870 0.9180328 1.375549

124 {Book1=1,Book3=1,Book5=1} => {Survives=1} 0.04034896 0.9024390 1.352184

125 {Book1=1,Book2=0,Book5=1} => {Survives=1} 0.01308615 1.0000000 1.498366

126 {Book2=1,Book4=1,Book5=1} => {Survives=1} 0.02399128 0.9166667 1.373502

235 {House=Stark,Nobility=1,Book1=0,Book2=0} => {Survives=1} 0.02290076 0.9545455 1.430258

236 {House=Stark,Book1=0,Book2=0,Book3=0} => {Survives=1} 0.01962923 1.0000000 1.498366

237 {House=Stark,Book1=0,Book2=0,Book4=0} => {Survives=1} 0.02071974 0.9047619 1.355664

259 {House=Lannister,Gender=M,Nobility=1,Book3=1} => {Survives=1} 0.02508179 0.9200000 1.378497

261 {House=Lannister,Nobility=1,Book1=0,Book3=0} => {Survives=1} 0.01962923 0.9000000 1.348529

311 {Nobility=0,Book1=1,Book4=0,Book5=1} => {Survives=1} 0.01526718 0.9333333 1.398475

312 {Gender=M,Nobility=0,Book1=1,Book5=1} => {Survives=1} 0.01635769 0.9375000 1.404718

324 {Book1=0,Book2=1,Book3=1,Book4=1} => {Survives=1} 0.02071974 0.9047619 1.355664

342 {Nobility=0,Book2=1,Book3=1,Book5=1} => {Survives=1} 0.02181025 0.9523810 1.427015

343 {Book1=0,Book2=1,Book3=1,Book5=1} => {Survives=1} 0.02181025 0.9090909 1.362151

344 {Nobility=0,Book1=0,Book2=1,Book5=1} => {Survives=1} 0.01308615 0.9230769 1.383107

345 {Gender=M,Nobility=0,Book2=1,Book5=1} => {Survives=1} 0.01962923 0.9000000 1.348529

346 {Nobility=1,Book2=0,Book3=1,Book5=1} => {Survives=1} 0.01308615 1.0000000 1.498366

347 {Nobility=1,Book1=0,Book3=1,Book5=1} => {Survives=1} 0.02071974 0.9047619 1.355664

348 {Nobility=1,Book1=0,Book2=0,Book3=0} => {Survives=1} 0.12322792 0.9112903 1.365446

349 {Book1=0,Book2=0,Book3=0,Book5=0} => {Survives=1} 0.10905125 0.9259259 1.387376

507 {House=None,Gender=F,Book1=0,Book3=0,Book5=0} => {Survives=1} 0.01199564 0.9166667 1.373502

515 {Gender=F,Nobility=0,Book1=0,Book2=1,Book3=0} => {Survives=1} 0.01090513 0.9090909 1.362151

520 {Gender=F,Nobility=0,Book1=0,Book3=0,Book5=0} => {Survives=1} 0.02071974 0.9047619 1.355664

1. *Remaining rules sorted based on their lifts*

**lhs rhs support confidence lift**

18 {House=Arryn,Book1=0} => {Survives=1} 0.01199564 1.0000000 1.498366

27 {House=Stark,Book5=1} => {Survives=1} 0.02944384 1.0000000 1.498366

29 {House=Lannister,Book5=1} => {Survives=1} 0.01199564 1.0000000 1.498366

110 {House=NightsWatch,Book1=1,Book5=1} => {Survives=1} 0.01308615 1.0000000 1.498366

125 {Book1=1,Book2=0,Book5=1} => {Survives=1} 0.01308615 1.0000000 1.498366

236 {House=Stark,Book1=0,Book2=0,Book3=0} => {Survives=1} 0.01962923 1.0000000 1.498366

346 {Nobility=1,Book2=0,Book3=1,Book5=1} => {Survives=1} 0.01308615 1.0000000 1.498366

122 {Gender=F,Nobility=1,Book5=1} => {Survives=1} 0.03162486 0.9666667 1.448420

107 {House=Lannister,Nobility=1,Book2=0} => {Survives=1} 0.02726281 0.9615385 1.440737

30 {Gender=F,Book4=1} => {Survives=1} 0.05125409 0.9591837 1.437208

108 {House=Lannister,Book1=0,Book2=0} => {Survives=1} 0.02508179 0.9583333 1.435934

235 {House=Stark,Nobility=1,Book1=0,Book2=0} => {Survives=1} 0.02290076 0.9545455 1.430258

120 {House=None,Gender=F,Nobility=1} => {Survives=1} 0.02181025 0.9523810 1.427015

342 {Nobility=0,Book2=1,Book3=1,Book5=1} => {Survives=1} 0.02181025 0.9523810 1.427015

121 {Gender=F,Book3=1,Book5=1} => {Survives=1} 0.02071974 0.9500000 1.423448

17 {House=Arryn,Book3=0} => {Survives=1} 0.01962923 0.9473684 1.419505

16 {House=Arryn,Nobility=1} => {Survives=1} 0.01853871 0.9444444 1.415123

15 {House=Arryn,Book4=1} => {Survives=1} 0.01635769 0.9375000 1.404718

312 {Gender=M,Nobility=0,Book1=1,Book5=1} => {Survives=1} 0.01635769 0.9375000 1.404718

3 {House=Tyrell} => {Survives=1} 0.01526718 0.9333333 1.398475

311 {Nobility=0,Book1=1,Book4=0,Book5=1} => {Survives=1} 0.01526718 0.9333333 1.398475

28 {House=Lannister,Book4=1} => {Survives=1} 0.04252999 0.9285714 1.391340

25 {House=Greyjoy,Book2=0} => {Survives=1} 0.04143948 0.9268293 1.388729

349 {Book1=0,Book2=0,Book3=0,Book5=0} => {Survives=1} 0.10905125 0.9259259 1.387376

14 {House=Targaryen,Book5=1} => {Survives=1} 0.01308615 0.9230769 1.383107

24 {House=Greyjoy,Book4=1} => {Survives=1} 0.03925845 0.9230769 1.383107

46 {House=Tully,Nobility=1,Book2=0} => {Survives=1} 0.01308615 0.9230769 1.383107

47 {House=Tully,Gender=M,Book2=0} => {Survives=1} 0.01308615 0.9230769 1.383107

344 {Nobility=0,Book1=0,Book2=1,Book5=1} => {Survives=1} 0.01308615 0.9230769 1.383107

4 {House=Martell} => {Survives=1} 0.02508179 0.9200000 1.378497

259 {House=Lannister,Gender=M,Nobility=1,Book3=1} => {Survives=1} 0.02508179 0.9200000 1.378497

31 {Nobility=1,Book4=1} => {Survives=1} 0.14721919 0.9183673 1.376050

123 {Gender=F,Nobility=1,Book2=0} => {Survives=1} 0.06106870 0.9180328 1.375549

2 {House=HouseMartell} => {Survives=1} 0.01199564 0.9166667 1.373502

88 {House=Baratheon,Nobility=0,Book1=0} => {Survives=1} 0.01199564 0.9166667 1.373502

126 {Book2=1,Book4=1,Book5=1} => {Survives=1} 0.02399128 0.9166667 1.373502

507 {House=None,Gender=F,Book1=0,Book3=0,Book5=0} => {Survives=1} 0.01199564 0.9166667 1.373502

348 {Nobility=1,Book1=0,Book2=0,Book3=0} => {Survives=1} 0.12322792 0.9112903 1.365446

1 {House=HouseTyrell} => {Survives=1} 0.01090513 0.9090909 1.362151

343 {Book1=0,Book2=1,Book3=1,Book5=1} => {Survives=1} 0.02181025 0.9090909 1.362151

515 {Gender=F,Nobility=0,Book1=0,Book2=1,Book3=0} => {Survives=1} 0.01090513 0.9090909 1.362151

109 {House=Lannister,Gender=M,Book1=0} => {Survives=1} 0.04252999 0.9069767 1.358983

32 {Book3=0,Book4=1} => {Survives=1} 0.15703381 0.9056604 1.357011

26 {House=Stark,Gender=F} => {Survives=1} 0.02071974 0.9047619 1.355664

237 {House=Stark,Book1=0,Book2=0,Book4=0} => {Survives=1} 0.02071974 0.9047619 1.355664

324 {Book1=0,Book2=1,Book3=1,Book4=1} => {Survives=1} 0.02071974 0.9047619 1.355664

347 {Nobility=1,Book1=0,Book3=1,Book5=1} => {Survives=1} 0.02071974 0.9047619 1.355664

520 {Gender=F,Nobility=0,Book1=0,Book3=0,Book5=0} => {Survives=1} 0.02071974 0.9047619 1.355664

124 {Book1=1,Book3=1,Book5=1} => {Survives=1} 0.04034896 0.9024390 1.352184

261 {House=Lannister,Nobility=1,Book1=0,Book3=0} => {Survives=1} 0.01962923 0.9000000 1.348529

345 {Gender=M,Nobility=0,Book2=1,Book5=1} => {Survives=1} 0.01962923 0.9000000 1.348529

1. *Observations*

The dataset contains the values for the attributes Survival, Nobility etc., for all of the characters in the Game of Thrones books. It can be seen that the above rules generated after removing redundant rules give the “survives” value for the character as either 1 or 0. Since we have ignored using “Name” in our analysis, the above rules are generated for the other remaining attributes.

The rules which have Survives on the RHS are reported. From the unique rules above we can answer the questions the following questions

Does nobility play any role in survival? What about gender?

Yes nobility plays a role in the survival of a character. Also. Gender too plays a role in the survival of a character. These are evident from the presence of nobility and gender values in the LHS of the above rule set.

In the first five books, Jon Snow lives. Do you find any rule which says otherwise? (like his fate in the Game of Thrones TV-Series)

No. There is no rule which says that Jon Snow will not live. The following rule indicates that given Jon Snow’s attribute values, he is likely to survive.

**110 {House=NightsWatch,Book1=1,Book5=1} => {Survives=1} 0.01308615 1.0000000 1.498366**

Select your own favorite character (someone other than Jon Snow ) and see if you have any rule for his/her survival?

My favorite character is Arya Stark. Given her attribute values, the following three rules say that she will survive which is true.

**27 {House=Stark,Book5=1} => {Survives=1} 0.02944384 1.0000000 1.498366**

**122 {Gender=F,Nobility=1,Book5=1} => {Survives=1} 0.03162486 0.9666667 1.448420**

**26 {House=Stark,Gender=F} => {Survives=1} 0.02071974 0.9047619 1.355664**