Titanic Data Set

Skyler Preston

MIS6240

Data Warehousing

Hands-On Lab 2

4-16-2023

Professor

Dr. Amitava Karmaker

**Dataset Name:** TitanicSurvial

**Format:** Association Rule

**Format:**

A data frame with 1309 observations on the following 4 variables.

survived

no or yes.

sex

female or male

age

in years (and for some children, fractions of a year); age is missing for 263 of the passengers.

passengerClass

1st, 2nd, or 3rd class.

**Details:**

This is part of a larger data set compiled by Thomas Cason. Many additional details are given in the sources cited below.

**Rule Package**: Apriori

**Function:** Remove column 1 as the association rule does not care about the names of the surviving witness.

**Script**: tt=TitanicSurvival [,2:5]

**Rule**: rules=apriori(tt,parameter = list(supp=0.1,conf=0.5,minlen=2))

Text

Description automatically generated

**Object Created**: 54 Objects of Rules

**Sort:** rules=sort(rules, by= "lift")

**Function**: This sorts the descending rules by lift

**Script :** inspect(rules)

Table

Description automatically generated

**Description**: The data above is showing the likelihood of a passenger surviving on Titanic in row 1 for example. If the person is a Female and sits in first class this person would have a 96% confidence of surviving. This goes up by 2.5 times that you would survive. The RHS in row one shows that female in passenger class will survive.

**Rule**:rules=apriori(tt,parameter =list(supp=0.15,conf=0.7,minlen=2))

**Script**: rules=sort(rules, by= "lift")

> inspect(rules)

Text

Description automatically generated

**Description:** Similar to Row 1 with changing out the support to .15 and Conf= 0.7 the association rule changes the dataset to show total of six results. In row one Condition Female with RHS survived the confidence goes 72% and goes up by .9 times for the lift out of 339 data count.

**References:**

https://www.encyclopedia-titanica.org/

F. E. Harrell, Jr. (2001) Regression Modeling Strategies New York: Springer.