Shay Snyder

CSCI 4957-901

Dr. Husari

Exercise 10

Today, I am going to apply my data mining skills, specifically the Apriori algorithm, to the ‘TitanicData.csv’ dataset with the goal of identifying which groups of people were more likely to survive the horrific incident . This data represents 2201 of the real Titanic passengers. It contains various attributes about every passenger such as their state of survival, their age, their class aboard the ship, and their sex.

To determine which groups of individuals are more likely to survive, I will apply the Apriori algorithm with support of 0.005, confidence 0.8, and minimum item length must be 2. I will remove the redundant rules and report all the rules sorted by lift. The first run of the Apriori algorithm produced 27 rules.

A close up of text on a white background

Description automatically generated

Since, we have an abundance of rules, lets apply our requirements for a given rule. These minimum requirements are item length equal to 2, support 0.005, and confidence 0.8. This calculation results in 12 rules.

A close up of text on a white background

Description automatically generated

From these results, we can see that a vast majority of the removed rules violated the minimum length requirement of 2. The resulting rules are those who hold some precedence to our study. Let’s sort these rules based on lift.

A close up of text on a white background

Description automatically generated

From the results of our study, we can draw various conclusions. It is visually evident that a strong, positive lift correlates with confidence. In fact, 100% of the tested children in 2nd class were able to survive the accident. We can infer that children were much more likely to get a seat upon a raft, as their life was deemed most important. Closely trailing children were their mothers and other females as having the highest chance of survival. Even though this data is very informative, we are unable to account for a large proportion of the population. However, we can conclude that females and children were given priority when boarding the boats. In conclusion, the most important factors contributing to an individual’s survival, was their age and sex.