Θεμα 8

1. The apriori algorithm applied to the fertility dataset has generated a total of 496 rules. These rules represent associations between various antecedent conditions, which are factors related to a man's sperm quality, and the consequent diagnosis of the sperm being altered or normal (O or N). Each rule provides insights into the relationships and patterns observed in the dataset, indicating under which conditions a man's sperm is more likely to be diagnosed as altered. The antecedent conditions specified in the rules capture the factors influencing fertility, and the consequent reflects the ultimate diagnosis based on those conditions.
2. Firstly, a minimum support threshold of 0.02 is chosen. This means that only associations or item-sets appearing in at least 2% of the transactions will be considered significant, indicating their frequency in the dataset. Next, a confidence level of 1 is selected. A confidence of 1 implies that only rules with 100% confidence will be included in the results. This signifies a strong relationship between the antecedent conditions and the consequent outcome. In other words, the presence of the antecedent conditions ensures the presence of the specified consequent, which, in this case, is a diagnosis of "altered." Additionally, a constraint is applied to the right-hand side of the rules, specifying that only those with "Diagnosis=altered" should be considered. This focuses the analysis on rules that directly lead to an altered diagnosis, providing targeted insights into factors associated with this particular outcome in male fertility. After applying these conditions, the algorithm returns 26 rules. These 26 rules reflect the strongest associations between specific antecedent conditions and the consequent diagnosis of "altered." Each rule encapsulates a pattern or relationship observed in the dataset, shedding light on potential factors influencing the alteration in sperm quality.
3. Next, we eliminate based on criteria of being a super rule. A rule is considered a super rule if there exists another rule with a larger or equal lift and identical antecedent and consequent conditions. In simpler terms, if one rule entirely encompasses another in terms of conditions and has an equal or greater lift value, the latter is deemed a super rule and can be excluded from the final set of rules. Lift quantifies the strength of a rule by calculating the ratio of the observed support of the rule to the expected support if the antecedent and consequent were independent. A lift value greater than 1 indicates that the rule has a positive influence, suggesting that the occurrence of the antecedent conditions increases the likelihood of the consequent. Conversely, a lift less than 1 implies a negative or no association. By sorting the rules based on lift, we prioritize those with higher lift values, emphasizing stronger and more meaningful associations. After the elimination process, the code identifies and retains five rules that stand out as significant patterns within the dataset. These remaining rules are characterized by strong and non-redundant associations within the dataset, and their higher lift values, highlighting robust associations between specific antecedent conditions and the consequent diagnosis of "altered" in male fertility.