



Mining for Insights: An Evaluation of Association Rules for Household Water Perception Data

T. Satheeskumar *, T. Satheeskumar **, P. Selvarajan***

*University of Vavuniya, Vavuniya, Sri Lanka, thevaka_sathees@vau.ac.lk

**Dept. of National Community Water Supply, Jaffna, Sri Lanka, stsatheesjii19@gmail.com

*** University of Vavuniya, Vavuniya, Sri Lanka, psevarajan@vau.ac.lk

Keywords: Apriori Algorithm; Data Mining; Drinking Water; Public Health; Rule Mining

Abstract: Understanding household perceptions of water quality and safety is essential for effective water management and public health planning. This study analyses household water perception data from Sri Lanka's Northern Province using Association Rule Mining (ARM). Data from 101 households covering twelve categorical perception variables were examined using the Apriori algorithm. Frequent itemsets and association rules were evaluated using support, confidence, and lift measures. The analysis identified consistent perception patterns, particularly linking perceived water quality and supply reliability with household precautionary practices. Several strong, high-confidence rules indicated non-random relationships among perception variables. Overall, the findings demonstrate that ARM is a transparent and effective method for uncovering latent perception structures in household survey data and provides useful insights to inform water management and public health interventions.

Introduction:

Public health organizations commonly use Knowledge, Attitudes, and Practices surveys to assess household perceptions of drinking water safety. However, such high-dimensional categorical data are difficult to analyse using conventional descriptive methods alone. Recent advances in data mining have shown that Association Rule Mining (ARM) is effective in uncovering hidden patterns in perception-based survey data (Sarker, 2021). Applying ARM to household water perception data enables the identification of dominant perception structures that may not be evident through traditional analytical approaches.

Methodology:

The present study analyzed 101 valid household responses, comprising 12 categorical perception variables related to water quality, availability, safety, and household management practices. Given the categorical nature of the data, Association Rule Mining was selected as an appropriate analytical framework. All perception variables were transformed into a transactional format using one-hot encoding prior to analysis. The Apriori algorithm was then employed to extract frequent itemsets and association rules. This approaches have been widely validated for survey and perception data analysis due to their interpretability and robustness (Al-Maolegi & Arkok, 2020; Wu &

Zhang, 2021). Rule strength was evaluated using support, confidence, and lift measures, which are standard metrics in association rule analysis.

Findings/Results:

The Apriori algorithm revealed consistent household perception patterns, with individual variables showing ~14% support and multi-variable combinations ~13%. Several rules achieved 1.0 confidence with high lift, indicating strong, non-random associations. An average of 77.23 rules (SD = 74.32) reflected expected variability in household water perceptions. As shown in Table 1, the itemsets capture interpretable links between water quality concerns, supply reliability, service dissatisfaction, and household treatment behaviors, suggesting adaptive responses to perceived risks.

Table 1: Frequent Itemsets Derived from Household Water Perception Survey Data

No.	Itemsets (Interpreted)	Support
1	Irregular water supply	0.141176
2	Concern about drinking water quality	0.141176
3	Irregular water supply AND quality concern	0.141176
4	Use of household water treatment methods (boiling/filtering)	0.129412
.....
8	Household treatment practice AND quality concern	0.129412
9	Irregular supply AND dissatisfaction with water service	0.129412
10	Quality concern AND use of alternative water sources	0.129412

Nodes represent perception variables and edges denote strong Apriori-identified associations (Figure 1), with the network visualization intuitively revealing interconnected perception structures, dominant variables shaping household attitudes, and complementing quantitative association rule metrics as demonstrated in prior survey-based perception studies.

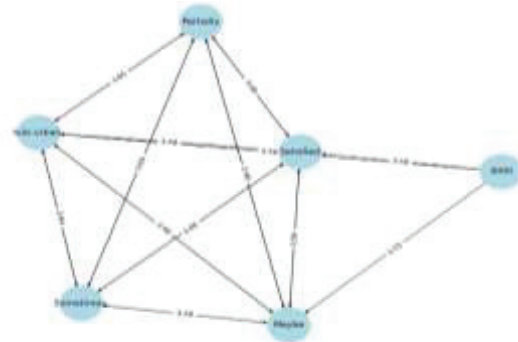


Figure 1: Network Graph of Association Rules for Household Water Perceptions

Conclusions:

This study confirms that Apriori-based ARM is effective for identifying latent perception structures in household water survey data. Household water safety perceptions are influenced by interconnected factors, particularly perceived quality, supply reliability, and precautionary behaviours, with observed variability reflecting genuine differences in household experiences.

References:

- Al-Maolegi, M. & Arkok, B. 2020 An improved Apriori algorithm for association rules mining. *Int. Journal of Computer Science and Information Technology*, **11**(1), pp. 39–46.
- Sarker, I.H. 2021 Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, **2**(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>
- Wu, X. & Zhang, C. 2021 The application of improved Apriori algorithm in the analysis of university students' mental health survey. *Journal of Physics: Conference Series*, **1881**(4), 042024.