International Journal of Current Sciences and Engineering Intelligence
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: *https://ijcsi.com*
Volume 01, Issue 01, September 2026
OPEN ACCESS
E-ISSN :

Stability Aware Association Rule Selection for Retail Transaction Analysis

Saurabh Pandey[1], Manish Bose[2], Kavya Agarwal[3], Radhika Das[4]
[1],[2],[3],[4] Students, Chimanbhai Patel Institute of Computer Applications, Sardar Vallabhbhai Global University, Ahmedabad

Article Info

Article History:
Published: 14 Nov 2026

Publication Issue:
Volume 01, Issue 01
September 2026

Page Number:
1-5

Corresponding Author:
Saurabh Pandey

Abstract:
Association rule mining can produce patterns with high confidence or lift that fail to remain useful when transaction conditions change. This paper develops an analytical screening framework that combines support, antecedent count, temporal stability, and a conservative confidence score. Two constructed periods of 1,000 transactions are used to compare four candidate rules. One rare rule has 90 percent confidence and a lift of 3.0 in the discovery period, but its validation confidence falls to 40 percent and its lift falls to 1.0. A more stable rule retains 68 percent confidence and a lift of 1.7 in validation. The proposed screening procedure retains that stable rule while rejecting candidates that are rare, unstable, or insufficiently informative relative to the current baseline. Wilson lower scores are used as an illustrative screening device under a binomial sampling interpretation, not as proof of statistical significance for synthetic data. The study demonstrates why pattern discovery and operational acceptance should be separate decisions and identifies the additional evidence required for real retail deployment.

Keywords: Association rules; market basket analysis; temporal stability; support; confidence; lift; pattern validation; data mining.

Introduction

Association rules summarize relationships between items occurring in the same transaction. A rule such as X implies Y can help an analyst examine purchasing combinations, product placement, or recommendation opportunities. The notation describes an observed association; it does not establish that purchasing X causes a customer to purchase Y. That distinction remains important when a discovered rule is later used to justify an intervention.

Several measures describe different aspects of a rule. Support indicates how often the item combination appears in the database. Confidence measures the proportion of transactions containing X that also contain Y. Lift compares that confidence with the overall prevalence of Y. A rule can perform well on one measure and poorly on another, especially when its antecedent is rare or the consequent is already common.

A further difficulty arises when customer behavior changes between discovery and deployment. A high confidence rule estimated from a small number of transactions can disappear in a later period. Changes in overall product popularity can also reduce lift even when conditional purchase probability remains stable. Selecting rules from one dataset without checking these effects can produce an unreliable recommendation set.

This paper asks how a transparent acceptance procedure can combine frequency, stability, and baseline relevance. The study uses deliberately constructed transaction periods to isolate these mechanisms. Its contribution is an interpretable comparison of rule selection policies, not a benchmark claim about a mining algorithm or evidence of improved retail sales. Every reported value can be reconstructed from the stated transaction counts.

Detailed Literature Review

Agrawal, Imieliński, and Swami [1] formalize association rule discovery in large transaction databases. Agrawal and Srikant [2] develop efficient algorithms for identifying frequent itemsets and rules. These foundational contributions

Copyright © 2026 The Author(s) : This is an open access article under the CC BY license (http://creativecommons.org/licenses/by-nc/4.0/)

address the discovery problem, including how to explore combinations efficiently. Finding candidate patterns, however, does not by itself determine whether those patterns should be retained for a later operational period.

Brin and colleagues [3] develop dynamic itemset counting and implication rule methods. Brin, Motwani, and Silverstein [4] examine correlations beyond the conventional market basket setting. Their work supports distinguishing frequent cooccurrence from informative dependence. This distinction motivates the present study's comparison against the contemporaneous prevalence of the consequent.

Hahsler, Grün, and Hornik [5] provide the arules computational environment for representing and mining transaction data. Tan, Kumar, and Srivastava [6] investigate the choice of objective measures for association analysis. Different measures can rank the same patterns differently because they respond to different properties of contingency tables. The present framework therefore reports several measures instead of treating one score as a universal definition of usefulness.

Webb [7] examines significant pattern discovery and the problem of spurious patterns in large search spaces. Hämäläinen and Webb [8] explain statistically sound pattern discovery, including the consequences of evaluating many candidate patterns. These works motivate separating exploratory selection from subsequent evaluation. The small example below does not implement a multiple testing procedure and does not claim familywise error control.

Wilson [9] develops a score based approach to inference for a proportion, while Agresti and Coull [10] compare interval estimation methods for binomial proportions. Their work informs the lower score used here to penalize uncertain conditional frequencies. However, applying that calculation to constructed counts is only an illustration of a possible screening rule. Statistical coverage would require an appropriate sampling process, and comparing a conditional bound with an estimated baseline is not itself a valid test of their difference.

Detailed Methodology & Framework

Two separate transaction periods each contain 1,000 records. The discovery period contains the consequent item Y in 300 transactions, while the validation period contains it in 400. Four antecedent items, A, B, C, and D, occur in 400, 100, 20, and 300 transactions respectively in each period. Their discovery joint counts with Y are 180, 70, 18, and 90. Their validation joint counts are 190, 68, 8, and 120.

These marginal and joint counts can be realized in consistent transaction databases. In each period, assign Y to the first required number of records. For each antecedent, assign its joint count among those Y records and its remaining occurrences among records without Y. Antecedents may overlap because a transaction can contain several items. The item identifiers have no hidden product meaning, and no real retailer supplied the data.

For each antecedent, support is its joint count with Y divided by 1,000. Confidence is the joint count divided by antecedent count. Lift is confidence divided by the period's prevalence of Y. Temporal drift is the absolute difference between discovery and validation confidence. These definitions ensure that the changing baseline enters the lift calculation separately in each period.

$$\text{Support} = \frac{\text{nXY}}{\text{N}} \; ; \; \text{Confidence} = \frac{\text{nXY}}{\text{nX}}$$

$$\text{Lift} = \frac{\text{nXY/nX}}{\text{nY/N}} \; ; \; \text{Drift} = |\text{pvalidation} - \text{pdiscovery}|$$

$$\text{L} = \frac{\hat{\text{p}} + \text{z}^2/(2\text{n}) - \text{z}\sqrt{\hat{\text{p}}(1 - \hat{\text{p}})/\text{n} + \text{z}^2/(4\text{n}^2)}}{1 + \text{z}^2/\text{n}}$$

N is transaction count, nX is antecedent count, nY is consequent count, and nXY is joint count. In the Wilson score, $\hat{p}$ = nXY/nX, n = nX and z = 1.96. The score is an illustrative screening device under the assumptions stated in the text.

Three simple selection policies use discovery data alone. The confidence policy retains rules with confidence of at least 0.60. The lift policy retains rules with lift of at least 2.0. The combined frequency policy requires support of at least 0.05 and confidence of at least 0.40. These thresholds are transparent design choices, not universal recommendations.

The proposed temporal screening procedure requires joint support of at least 0.05 in both periods, at least 50 antecedent occurrences in both periods, and confidence drift no greater than 0.10. It additionally requires the validation Wilson lower score to exceed the validation prevalence of Y by more than 0.10. The score uses a normal quantile of 1.96, corresponding to the conventional two sided 95 percent Wilson calculation under a binomial model.

The lower score is a conservative ranking aid in this constructed exercise. It does not establish that the selected rule has a statistically significant excess probability, because baseline uncertainty, rule dependence, and selection effects are not jointly modeled. The additional 0.10 margin represents a minimum practical difference chosen for the example.

The validation period is used for acceptance screening and is therefore part of model selection. A third, untouched period would be needed for an independent performance assessment of the selected policy. Sensitivity analysis changes the validation count for rule B and the minimum support threshold. The framework diagram shows rejected rules returning to monitoring rather than being treated as permanently false.

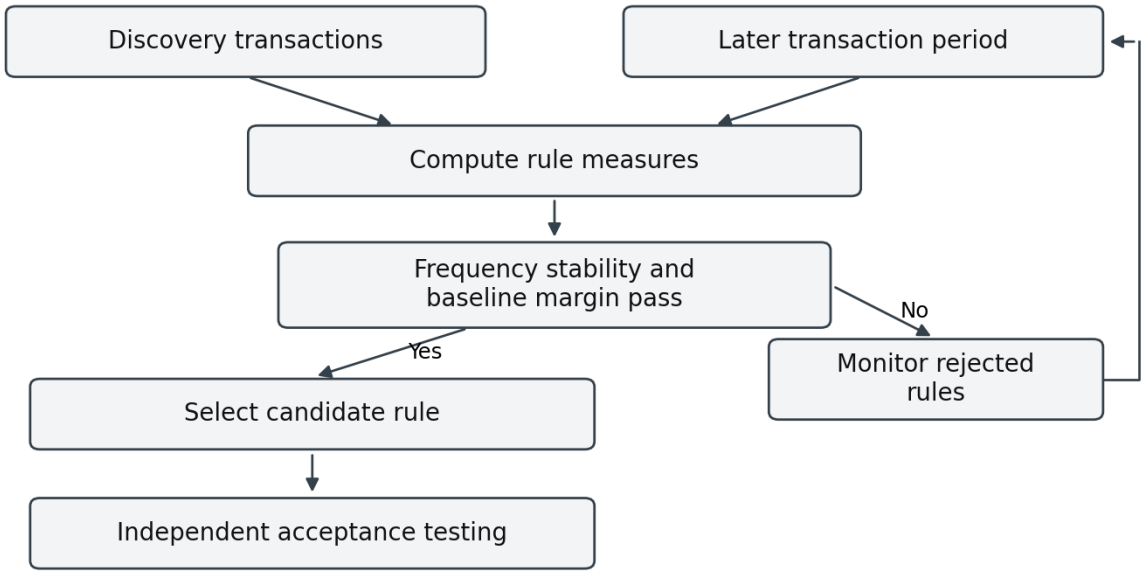


Figure 1. Temporal screening before independent rule testing

Detailed Results and Analysis

In the discovery period, rule C appears strongest under confidence and lift. Its confidence is 0.90 and its lift is 3.00. However, the antecedent occurs only twenty times. In validation, the joint count falls from eighteen to eight, confidence falls to 0.40, and lift becomes 1.00. The rule now provides no conditional increase over the overall frequency of Y.

Rule B has lower discovery confidence than C but substantially greater stability. Its confidence changes from 0.70 to 0.68, while lift changes from approximately 2.33 to 1.70 because Y becomes more common. Its validation lower score is approximately 0.583, exceeding the screening threshold of 0.50. It also satisfies both support requirements and the antecedent count requirement.

Table 1. Calculated rule comparison across periods

Rule	Discovery confidence	Validation confidence	Discovery lift	Validation lift	Temporal screen
A → Y	45.0%	47.5%	1.50	1.19	Reject
B → Y	70.0%	68.0%	2.33	1.70	Accept

Rule	Discovery confidence	Validation confidence	Discovery lift	Validation lift	Temporal screen
C → Y	90.0%	40.0%	3.00	1.00	Reject
D → Y	30.0%	40.0%	1.00	1.00	Reject

Note: Values are calculated from the stated analytical scenario.

Rule A demonstrates a different issue. Its confidence rises from 0.45 to 0.475, yet lift falls from 1.50 to approximately 1.19 because the background prevalence increases faster. Its lower score is approximately 0.427 and does not exceed the required threshold. Rule D has lift equal to one in both periods, so it reflects the baseline frequency without an additional association.

The confidence and lift policies both retain B and C from discovery data. The frequency policy retains A and B. The proposed temporal procedure retains only B. This result shows stricter screening in the constructed example; it is not a measured improvement in predictive accuracy, customer response, or profit. Rule coverage and business value would require additional data.

Table 2. Comparison of rule selection policies

Selection policy	Rules	Lowest validation lift
Discovery confidence ≥ 0.60	B, C	1.00
Discovery lift ≥ 2.00	B, C	1.00
Support ≥ 0.05 and confidence ≥ 0.40	A, B	1.19
Proposed temporal screening	B	1.70

Note: Values are calculated from the stated analytical scenario.

The sensitivity calculation reveals how close a candidate may be to acceptance. If B has sixty validation cooccurrences out of one hundred antecedent transactions, its lower score is approximately 0.502 and its confidence drift is exactly 0.10. It still passes. At fifty nine cooccurrences, the lower score falls to approximately 0.492 and drift rises to 0.11, so it fails. Raising minimum joint support to 0.10 removes B altogether and leaves no accepted rule. A stricter threshold therefore trades coverage for selectivity.

Future Enhancement

A practical extension should use chronological transaction partitions from several stores and seasons, with rules selected before the final test period is inspected. Evaluation should report retained rule coverage, stability across periods, recommendation response, and incremental business outcomes. A rule's statistical association should be distinguished from the causal effect of acting on that rule.

Further work can incorporate formal multiple testing control, uncertainty in the consequent baseline, customer clustering, and adjustments for repeated transactions from the same buyer. Drift detection could trigger reconsideration when a product's prevalence changes. Domain constraints can exclude unavailable items or unsuitable recommendations. The screening thresholds should be chosen using an explicit operational objective, and their selection process should be separated from the data used to report final performance.

Conclusion

The analytical example demonstrates that high discovery confidence or lift can coexist with weak temporal stability. A rare rule that initially appears strongest loses all lift in the validation period, while a more moderate rule retains useful association. Combining support, sample size, temporal drift, and a baseline aware lower score yields a more selective acceptance procedure in the constructed data. The procedure remains a screening heuristic, not a validated significance test or a causal recommendation method. Its main value is making the reasons for accepting or rejecting each rule explicit and reproducible.

References

[1] Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. ACM SIGMOD Record, 22(2), 207–216. Source

[2] Agrawal, R., and Srikant, R. (1994). Fast Algorithms for Mining Association Rules in Large Databases. Proceedings of VLDB, 487–499. Source

[3] Brin, S., Motwani, R., Ullman, J. D., and Tsur, S. (1997). Dynamic Itemset Counting and Implication Rules for Market Basket Data. ACM SIGMOD Record. Source

[4] Brin, S., Motwani, R., and Silverstein, C. (1997). Beyond Market Baskets: Generalizing Association Rules to Correlations. ACM SIGMOD Record. Source

[5] Hahsler, M., Grün, B., and Hornik, K. (2005). arules—A Computational Environment for Mining Association Rules and Frequent Item Sets. Journal of Statistical Software, 14(15), 1–25. Source

[6] Tan, P. N., Kumar, V., and Srivastava, J. (2004). Selecting the Right Objective Measure for Association Analysis. Information Systems, 29(4), 293–313. Source

[7] Webb, G. I. (2007). Discovering Significant Patterns. Machine Learning, 68(1), 1–33. Source

[8] Hämäläinen, W., and Webb, G. I. (2019). A Tutorial on Statistically Sound Pattern Discovery. Data Mining and Knowledge Discovery, 33, 325–377. Source

[9] Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. Journal of the American Statistical Association, 22, 209–212. Source

[10] Agresti, A., and Coull, B. A. (1998). Approximate Is Better than Exact for Interval Estimation of Binomial Proportions. The American Statistician, 52(2), 119–126. Source