

Abstract

Market Basket Analysis (MBA) is a core technique in retail analytics that uncovers purchasing patterns, such as customers who buy bread often also buying butter. Traditional MBA approaches, based on algorithms like Apriori, FP-Growth, and Eclat, use statistical measures such as support, confidence, and lift to generate association rules. While useful, these methods remain limited because they ignore the context of transactions, focus on frequency rather than profitability, and measure only correlation rather than causation. In practice, customer behavior is strongly shaped by contextual factors such as time of day, store location, seasonality, and promotional offers, and ignoring these aspects reduces the relevance of discovered rules. Similarly, rules that are frequent may not always be profitable, as they may involve low-margin items that do not increase overall revenue. Furthermore, traditional MBA cannot determine whether recommending an item genuinely causes additional sales, or if it merely reflects a statistical co-occurrence.

This project introduces a context-aware and profit-optimized MBA framework that integrates causal analysis to address these challenges. The system mines association rules within different contexts using FP-Growth and Eclat, ranks them through a multi-objective scoring function that combines lift with expected profit margins and diversity, and applies uplift modeling to estimate whether a recommendation truly changes customer behavior. To ensure practical utility, the framework is implemented on a streaming pipeline that continuously updates results and provides recommendations in near real-time through an API and interactive dashboards. Experiments on benchmark retail datasets such as Instacart and Dunnhumby, along with simulated promotion data, demonstrate that the proposed method improves profit contribution, click-through rates, basket size, and incremental revenue compared to classical MBA. The thesis contributes not only a reproducible end-to-end pipeline but also a novel rule-ranking strategy and an evaluation protocol that distinguishes correlation from genuine causation, making MBA more actionable for modern retail environments.

JASNA JABIN(24102595)

M.Tech(IT)