CARMA 2026

8th Int. Conf. on Advanced Research Methods and Analytics
1-3 July 2026, Valencia, Spain

Retail Site Selection: Augmenting Statistical Learning with Large Language Models

Dominik Heinisch, Benedikt Lugauer, Enrique Fita Sanmartin, Markus Herrmann
Daten & KI, Lidl Stiftung & Co. KG, Neckarsulm, Germany.

How to cite: Heinisch, D.; Lugauer, B.; Fita Sanmartin, E.; Herrmann, M. (2026). Retail Site Selection: Augmenting Statistical Learning with Large Language Models. In: 8th International Conference on Advanced Research Methods and Analytics (CARMA 2026). Valencia, 1-3 July 2026. https://doi.org/10.4995/CARMA2026.2026.21920

Abstract

Site selection is one of the most critical strategic decisions in the retail sector, as physical locations represent high-capital, long-term commitments. In addition to integrating vast amounts of location-specific data, inherent spatial autocorrelation must be addressed with robust statistical learning methods.

Therefore, we implement a Spatial Durbin Model using a Two-Stage Least Squares (2SLS) estimator to predict the turnover of potential sites. Moving beyond standard point predictions, we apply a wild bootstrap approach to estimate reliable prediction intervals. Although this methodology yields solid results in standard settings, it occasionally produces extreme outliers. We trace this issue to limitations in the training data rather than the model itself—specifically, a bias toward only realized locations, the high degrees of freedom required for generalization, and latent, site-specific heterogeneity that is not captured in our dataset.

Thus, we propose a shift from purely structured econometric modeling toward the integration of Large Language Models (LLMs) to expand the accessible information space. We systematically evaluate distinct architectural approaches to augment predictions, focusing on automated feature engineering through web search and context-aware refinement of overlooked regional drivers.

Keywords: *Retail Site Selection; Expansion Strategy; Large Language Models; Spatial Durbin Model.*

Article in Press

This work is licensed under a Creative Commons License CC BY-NC-SA 4.0
Editorial Universitat Politècnica de València