Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: https://www.science-gate.com/IJAAS.html

Ontology-driven machine learning for marketing activation forecasting

Enkhtuul Bukhsuren, Nandin-Erdene Enkhmyagmar, Munkhtsetseg Namsraidorj *

Department of Information and Computer Sciences, School of Information Technology and Electronics, National University of Mongolia, Ulaanbaatar, Mongolia

ARTICLE INFO

Article history:
Received 15 January 2026
Received in revised form
4 July 2026
Accepted 10 July 2026

Keywords:
Marketing activation forecasting
Knowledge graphs
Ontology
Regression models
Semantic features

ABSTRACT

Marketing activation generates large volumes of data, but these data are often organized in flat tables that do not explicitly capture the relationships among brands, channels, partners, and promotions. As a result, forecasting models may identify statistical associations without fully representing how marketing elements interact in practice. This study addresses this limitation by organizing marketing knowledge through an ontology and implementing it as a knowledge graph. SPARQL queries were used to generate relational indicators that represent interaction patterns among products, promotions, channels, and partners. These semantic features were integrated into regression-based forecasting models. The empirical analysis is based on 2.9 million transactional records from a large FMCG distribution company covering the period 2020–2024. Multiple linear regression, random forest regression, and gradient boosting regression models were developed and compared with baseline models that relied only on conventional tabular features. Model performance was evaluated using RMSE, MAE, and adjusted R^2. Gradient boosting achieved the best performance (RMSE = 116.93; MAE = 10.15; adjusted R^2 = 0.27). Although the performance improvements were moderate, models enriched with ontology-derived features consistently outperformed the baseline models. The findings suggest that incorporating structured relational knowledge improves model interpretability and provides incremental predictive gains. The study demonstrates how semantic modeling can be systematically integrated into regression-based marketing activation forecasting in an enterprise context. Future research may apply the proposed framework to independent enterprise datasets to further assess its external validity.

1. Introduction

Marketing remains central to competitive strategy, though its impact now extends well beyond communication and brand visibility. Pricing, promotional intensity, and channel configuration all influence revenue dynamics and customer retention. Firms operate across an increasingly complex mix of digital platforms. Social media, online marketplaces, and partner-managed systems generate substantial volumes of data. However, greater data availability does not automatically yield clearer explanations. In practice, it can become difficult to trace how specific marketing actions are associated with subsequent performance outcomes. Prior work in marketing analytics has noted that data abundance may complicate managerial interpretation rather than simplify it (Wedel and Kannan, 2016).

The underlying challenge is often structural. Marketing data is generated within systems that evolved independently. Records related to products, brands, partners, and channels are rarely integrated in a relationally coherent form. Reporting tools summarize performance indicators, but they typically present simplified views of inherently interconnected activities. When campaigns overlap, or pricing adjustments coincide with distribution changes, isolating their effects becomes problematic. Managers may observe shifts in sales without being able to clearly attribute them to specific actions. Similar concerns appear in enterprise data research, where relational meaning diminishes when interconnected information is reduced to isolated tables (Hogan et al., 2021).

A growing stream of research has therefore turned toward more explicit representations of

* Corresponding Author.
Email Address: munkhtsetseg_n@num.edu.mn (M. Namsraidorj)
https://doi.org/10.21833/ijaas.2026.07.005
Corresponding author's ORCID profile:
https://orcid.org/0000-0001-7767-5311

relationships. Ontologies formalize domain concepts and clarify how they relate to one another. Knowledge graphs operationalize these conceptual structures in data form (Ji et al., 2022). Refinement techniques can enhance graph consistency and improve semantic coherence (Paulheim, 2016). Such representations are particularly relevant in marketing activation settings, where interactions among brands, channels, and partners evolve continuously rather than remaining fixed.

Advances in representation learning enable the incorporation of these relational structures into predictive analysis. Knowledge graph embedding methods encode entities and their relationships into vector spaces compatible with predictive models (Wang et al., 2017). Relational machine learning approaches likewise account for structured dependencies during model estimation (Nickel et al., 2016). Taken together, these developments indicate that structured domain knowledge can move beyond purely descriptive use and contribute directly to predictive tasks.

Despite these methodological advances, applied studies linking marketing ontologies to forecasting practice remain limited. While data-driven CRM initiatives have shown measurable organizational benefits (Chatterjee et al., 2021) and broader reviews document the growing integration of advanced analytics into marketing research (Mariani et al., 2022), empirical applications of ontology-based approaches in activation forecasting remain relatively scarce.

This study examines whether structuring marketing knowledge through an ontology and implementing it as a knowledge graph improves activation-related sales forecasting within a regression-based framework. An ontology is specified to represent the relationships among products, brands, channels, partners, and promotional mechanisms. From this structure, a knowledge graph is constructed, and semantic indicators are derived using SPARQL queries. These indicators are then incorporated into predictive models to evaluate their contribution to forecasting accuracy and interpretability relative to more conventional feature-engineering approaches.

2. Literature review

2.1. Machine learning in marketing analytics

Machine learning has assumed a more prominent role in marketing over the past decade, mainly because firms now collect far more detailed information about customers, channels, and transactions than before. A wide range of techniques, including linear regression, random forests, gradient boosting, and neural networks, are applied to tasks such as demand forecasting, churn analysis, and promotion effectiveness assessment. Large-scale forecasting competitions further demonstrate that structured feature design and ensemble-based methods often outperform purely statistical baselines in complex prediction environments (Makridakis et al., 2022). Prior research suggests that such data-driven systems can enhance organizational performance, particularly when integrated into customer relationship management processes (Chatterjee et al., 2021). Broader reviews similarly document the growing use of advanced analytical methods in marketing research and practice (Mariani et al., 2022).

Despite these advances, many practical implementations rely on large, flattened datasets created through extensive merging and aggregation. While this approach supports predictive accuracy, it often obscures the relational structure underlying marketing activities. Once data are combined in this way, connections among brands, partners, channels, and promotional mechanisms may no longer be explicitly represented. As a result, models may detect statistical patterns without clearly reflecting how marketing elements interact in practice. This limitation has encouraged researchers to consider approaches that preserve structural and semantic relationships within marketing data, thereby improving both interpretability and analytical coherence.

2.2. Ontology-based enterprise data integration

Ontologies are widely used to maintain consistent meaning across enterprise information systems and to support semantic interoperability. By defining shared concepts and relationships, organizations can reduce ambiguity and make distributed data easier to interpret and connect (Staab and Studer, 2016). When knowledge is structured in this way, it becomes easier to reuse and analyze across systems.

From a strategic perspective, systematically organized information assets can strengthen analytical capability and contribute to competitive advantage (Barney, 1991). In marketing practice, explicitly defining channel roles, incentive structures, and partner responsibilities can reduce ambiguity and support more coherent activation decisions (Kotler et al., 2019). Even so, examples showing how ontology-based models are converted into variables for predictive modeling remain relatively limited in applied business settings. Clear demonstrations of how formal semantic structures are operationalized into forecasting features are still developing.

2.3. Knowledge graphs in business and marketing analytics

Knowledge graphs have been adopted in enterprise environments to integrate heterogeneous business data and explicitly model relationships across organizational entities (Ruan et al., 2016).

In business analytics contexts, graph-based representations are applied to connect customers, products, and organizational entities within integrated enterprise data environments (Gomez-

Perez et al., 2017). Moreover, enterprise knowledge graph initiatives highlight their role in enabling cross-system integration and structured reasoning in evolving organizational settings (Chis et al., 2024).

Despite these developments, much of the literature remains focused on general enterprise intelligence or data integration. Empirical studies that embed knowledge graphs directly into end-to-end predictive modeling workflows for marketing activation performance and channel-level decision support remain comparatively limited.

2.4. Semantic reasoning and SPARQL-driven feature construction

Semantic reasoning extends analytical capabilities beyond traditional SQL-based querying, particularly when data are represented as graphs rather than flat tables. Using SPARQL, analysts can trace multi-step relationships and apply logical conditions across interconnected entities. This approach enables the detection of dependencies that are often difficult to identify in conventional transactional datasets. In predictive contexts, the results of semantic queries can be transformed into structured variables that reflect interactions across entities, time periods, and business settings. However, applied studies that systematically combine semantic querying, structured feature construction, and predictive modeling, particularly in marketing activation contexts, remain limited. Recent surveys on knowledge graph reasoning emphasize the increasing importance of structured relational inference and reasoning mechanisms for extracting richer knowledge from complex relational data (Liang et al., 2024).

2.5. Machine learning, explainability, and KG-augmented forecasting

In marketing analytics, machine learning models, particularly ensemble methods such as random forests and gradient boosting, are widely used because they can capture nonlinear relationships and interaction effects that linear models often overlook (Hastie et al., 2009). These methods are well-suited to enterprise datasets, where multiple interdependent factors shape patterns. There is also growing interest in incorporating structured domain knowledge into predictive workflows. Knowledge graphs provide a way to organize relationships among business entities and to derive features that reflect those connections (Gomez-Perez et al., 2017). By making relationships explicit, graph-based representations help bridge the gap between raw transactional data and domain-aware modeling.

At the same time, the practical use of predictive models in business settings requires some transparency. Decision-makers often need to understand why a model produces a particular output, especially when predictions inform resource allocation or strategic planning. Techniques such as SHAP can be used to estimate how individual variables contribute to model predictions, offering a clearer view of model behavior (Lundberg and Lee, 2017). Despite these advances, relatively few applied studies combine knowledge-graph-based feature construction with interpretable machine learning models within a single, end-to-end framework tailored to marketing activation analysis. Much of the existing work treats semantic modeling, prediction, and explanation as separate tasks instead of integrating them within a unified analytical pipeline. The present study brings these components together in an applied setting using real marketing activation data.

3. Background theory and hypothesis

3.1. Research gap

Marketing activation is now conducted in environments where channels, partners, and promotional mechanisms are closely integrated. Firms are required to manage multiple sales channels simultaneously, run overlapping campaigns, and respond to rapidly changing customer behavior. Because many marketing actions take place simultaneously, their effects often interact, making it difficult to isolate the contribution of a single campaign or decision to overall performance. In practice, shifts in sales or customer responses rarely appear as clear and unified signals. Instead, they are scattered across multiple information systems and reported in different formats. While business-intelligence dashboards and conventional econometric tools are useful for summarizing historical outcomes, they typically provide limited insight into how concurrent marketing actions interact. As a consequence, managers often struggle to identify which combinations of actions are responsible for observed activation results. These challenges have encouraged a move away from purely descriptive reporting toward analytical approaches that incorporate business context, structured representations, and predictive methods (Kotler et al., 2019).

However, despite this shift, the existing literature does not yet provide a clearly articulated analytical framework that systematically integrates semantic structuring, knowledge graph reasoning, and predictive modeling within marketing activation contexts.

3.2. Theory-related gaps

Although data-driven marketing has advanced considerably, several theory-related gaps remain. From a Resource-Based View (RBV) perspective, marketing knowledge and internal data are treated as strategic resources capable of generating sustained competitive advantage (Wernerfelt, 1984; Barney, 1991).

However, empirical research has rarely shown how such knowledge resources are formally structured into operational semantic models that can be embedded in decision-support systems. While ontologies provide conceptual tools for structuring knowledge (Staab and Studer, 2016), their transformation into applied knowledge graph representations for managerial use remains underexplored.

Dynamic Capability Theory (DCT) emphasizes the importance of sensing environmental changes and reconfiguring resources accordingly (Teece, 2016). Yet, limited empirical evidence demonstrates how semantic modeling infrastructures can support continuous updating of relational knowledge in response to evolving channel performance, customer demand, and partner configurations.

Within the Technology–Organization–Environment (TOE) framework, technological innovations generate value only when aligned with organizational routines and external pressures (Tornatzky et al., 1990). However, integrated analytical systems that combine knowledge graph technologies with organizational decision-making processes in marketing activation settings remain comparatively scarce.

3.3. Methodological limitations and gap statement

Existing research in enterprise and marketing analytics reveals methodological fragmentation. While semantic modeling, predictive analytics, and explainability have each been examined in prior studies, they are often addressed separately. Practical implementations that integrate ontology design, knowledge graph construction, structured feature generation, machine learning prediction, and interpretability within a single empirical workflow remain limited.

This limitation becomes more visible when viewed through established theoretical perspectives. From a Resource-Based View, structured knowledge assets can enhance analytical capability and support competitive advantage (Wernerfelt, 1984). Dynamic capabilities emphasize adaptation to evolving environments (Teece, 2016), while the Technology–Organization–Environment framework highlights alignment between technology and organizational context (Tornatzky et al., 1990). However, empirical research rarely demonstrates how these theoretical perspectives can be operationalized through an integrated semantic–predictive infrastructure.

Accordingly, this study develops and evaluates a unified analytical framework that combines ontology-based modeling, knowledge graph feature construction, machine learning-based prediction, and interpretable outputs in the context of marketing activation.

3.4. Research hypothesis and theoretical background

Drawing on RBV, TOE, and DCT perspectives, this study examines whether structured semantic modeling strengthens predictive analytics in marketing activation contexts. From an RBV perspective, marketing information can be viewed as an organized knowledge resource that enhances analytical capability when systematically structured (Barney, 1991). Knowledge graphs provide a practical mechanism for operationalizing such structured knowledge.

The TOE framework suggests that technological infrastructures such as semantic repositories and machine learning workflows create value when aligned with organizational processes (Tornatzky et al., 1990).

Dynamic Capability Theory emphasizes the importance of organizational capabilities that adapt to evolving environments (Teece, 2016). Based on these theoretical foundations, four hypotheses are proposed:

H1: Incorporating ontology-based structuring improves the explanatory power of activation-related sales models compared to unstructured tabular data representations.
H2: Knowledge-graph-derived features improve the model's ability to capture cross-entity interaction effects in activation-related sales forecasting.
H3: Machine learning models incorporating knowledge-graph-derived semantic features achieve higher predictive performance than models relying solely on conventional feature engineering.
H4: Integrating knowledge-graph-based features with interpretable machine-learning techniques improves the interpretability of variable contributions in activation-related sales forecasting.

4. Data and methodology

This research is informed by three complementary theoretical perspectives that shape how marketing data are structured and analyzed. The Resource-Based View (RBV) suggests that information assets and analytical capabilities contribute to sustained competitive advantage when systematically organized and embedded in organizational practice (Barney, 1991).

Dynamic Capability Theory (DCT) emphasizes the importance of continuously reconfiguring organizational capabilities to respond to changing market conditions (Teece, 2016).

The Technology–Organization–Environment (TOE) framework further indicates that technological solutions create value only when aligned with internal processes and external competitive pressures (Tornatzky et al., 1990).

Building on these foundations, the analytical design of this study reflects their core assumptions. The ontology was developed as a structured representation of relational organizational knowledge, consistent with the RBV perspective. The knowledge graph was employed to model evolving interactions among marketing actors, reflecting dynamic capability reasoning. Embedding the semantic layer within the enterprise analytics workflow corresponds to the alignment principle emphasized in the TOE framework. In this way, theoretical perspectives informed both the conceptual structure and the empirical implementation of the framework.

4.1. Proposed methodology

The analytical framework combines data from multiple enterprise systems and organizes them through an ontology-based structure. A knowledge graph is then created to represent semantic relationships among brands, channels, partners, and promotional activities. Information derived from this graph is translated into structured variables for use in machine learning models. This process aims to improve forecasting performance while maintaining interpretability within marketing activation contexts.

To implement this approach, a sequential workflow was designed, as illustrated in Fig. 1. The workflow begins with data preparation and concludes with model interpretation and managerial application.

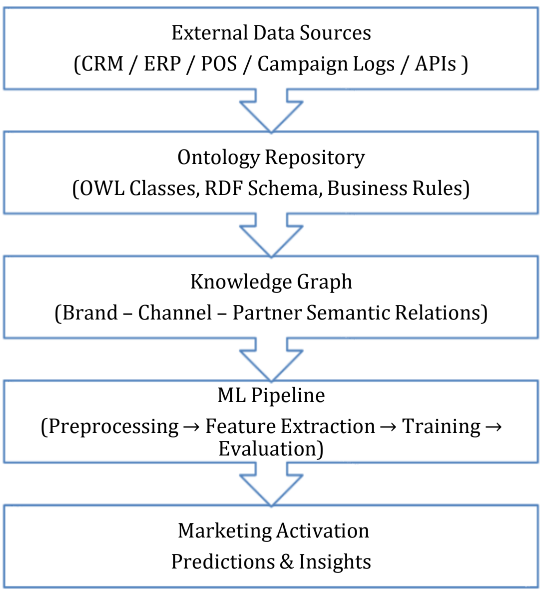

Fig. 1: Proposed methodology

Data preparation: Marketing and sales data covering the period from 2020 to 2024 were merged into a unified dataset. During preprocessing in Python, missing values were addressed, duplicate records were removed, and inconsistent or extreme observations were reviewed. Date fields and numerical variables were standardized to ensure consistency across sources.

Ontology design: The ontology was constructed gradually while working with the marketing activation data and reviewing relevant business documentation. Rather than beginning with a purely abstract schema, the structure evolved from practical analytical needs. During the early stages, we identified recurring business questions, for example, how specific promotions relate to particular products, how partner involvement interacts with channel activity, and how combinations of these elements correspond to sales outcomes. These guiding questions are consistent with ontology engineering practices described in prior studies (Staab and Studer, 2016). To address these analytical needs, we examined enterprise data schemas and internal process descriptions to determine which entities and relationships were essential. Core concepts, including Product, Brand, Channel, Promotion, Partner, and Sales, were then defined in Protégé, together with their object and data properties. The modeling decisions were informed by established marketing logic and activation theory (Kotler et al., 2019), particularly regarding how promotional mechanisms and channel strategies operate in practice. The structure was refined several times during development. Relationships were adjusted when inconsistencies appeared, and SPARQL queries were used to verify that the model returned meaningful patterns aligned with real marketing operations. Through this iterative process, the ontology gradually reached a form that was both logically coherent and suitable for downstream feature extraction. The ontology was validated internally through structural consistency checks, verification of class hierarchies and property constraints, and SPARQL-based query testing to ensure logical coherence and practical usability. While formal external domain-expert validation was not conducted, the model was developed and iteratively refined using enterprise documentation and established marketing theory to maintain conceptual alignment.

Knowledge graph construct: The validated ontology was exported in RDF format and deployed in GraphDB to create the operational knowledge graph. This graph represents interactions among brands, channels, partners, and promotional mechanisms within the enterprise context. Graph-based representations provide a structured way to organize relational business data in enterprise environments.

Semantic variable extraction: SPARQL queries were designed to extract relational indicators from the knowledge graph, including promotional intensity, channel activity, partner participation, and seasonal interaction patterns. The resulting query outputs were converted into structured features for subsequent predictive modeling.

The transformation of semantic relationships into quantitative indicators draws on established principles in knowledge graph embedding and structured representation learning (Wang et al., 2017; Ji et al., 2022). Rather than treating entities as independent attributes, relational information was encoded in a form suitable for integration into predictive models.

Machine learning stage: To evaluate whether semantic features improve forecasting performance, three supervised models were implemented: Multiple Linear Regression as a transparent linear baseline, Random Forest to capture nonlinear interactions, and Gradient Boosting to model more complex feature dependencies. These models reflect increasing modeling complexity, from a linear baseline to nonlinear ensemble approaches. All models were trained using grid search for hyperparameter optimization within a five-fold cross-validation framework. Performance was assessed using RMSE, MAE, and adjusted R^2 to compare predictive accuracy across model specifications. This evaluation design enabled a systematic comparison of models estimated with and without semantic features.

Interpretation and managerial interface: To support practical decision-making, model outputs were visualized in Power BI dashboards. SHAP (Shapley Additive Explanations) values were computed to estimate the contribution of individual variables to model predictions. Explainability techniques such as SHAP help clarify how different features influence model outputs, making predictive results easier to interpret in managerial settings.

4.2. Dataset

The dataset comprises approximately 2.9 million transactional records collected from a large distribution company in the fast-moving consumer goods (FMCG) sector. It includes product and brand identifiers, sales channel configurations, incentive types and durations, sales outcomes, and indicators of partner activity. Data from 2020 to 2024 were analyzed to capture seasonal effects and recurring activation patterns. All variables were cleaned and standardized to ensure temporal consistency. Six core concepts, Product, Brand, Channel, Promotion, Partner, and Sales, were defined to reflect the company's marketing structure. These concepts were modeled in Protégé and exported to GraphDB to construct an enterprise knowledge graph of marketing interactions. Representing the data in graph form enabled direct examination of cross-entity relationships and supported SPARQL-based feature extraction. The dependent variable in the regression models was activation-related sales revenue.

4.3. Reliability measures

To check the robustness of the results, several practical steps were taken during the analysis. The dataset was reviewed for missing values, duplicate entries, and unusual observations, and corrections were made where needed. Model performance was evaluated using five-fold cross-validation, along with additional tests based on alternative data splits. Training and validation results were compared to detect possible overfitting. No substantial discrepancies were observed between training and validation metrics. Hyperparameters were tuned via grid search during cross-validation to avoid overly optimistic model fitting. Feature importance was also examined across validation folds. The main semantic variables repeatedly appeared among the leading predictors, indicating that their contribution is not tied to a particular partition of the data. Together, these procedures enhance the robustness and reliability of the reported results.

5. Results and discussion

5.1. Ontology and knowledge graph overview

The marketing activation records were first aligned with the ontology's semantic structure. Key business entities, brands, products, channels, partners, promotions, and sales were represented together with the connections that link them in practice. Through this representation, it became possible to follow how products relate to their brands, how promotional activities are distributed across partners and channels, and how sales outcomes correspond to particular time frames and transaction characteristics. When these interactions were formalized, several structural patterns that were difficult to observe in ordinary tables, such as channel overlap or timing effects in promotions, became more transparent and easier to evaluate.

Fig. 2 presents the conceptual arrangement of the main ontology entities. Object properties such as ownsProduct, promotedBy, appliedToPartner, usesChannel, generatesSales, and hasDate were formalized as RDF triples. This representation enables systematic tracing of brand–channel alignment, partner participation, and promotional cycles and temporal sales outcomes.

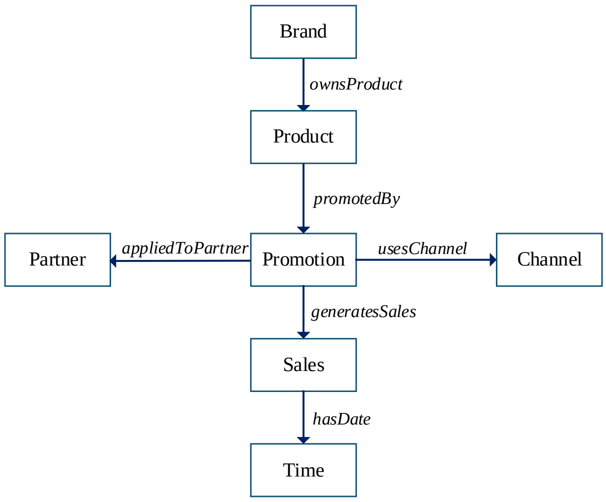

Fig. 2: Semantic relationships in the marketing ontology

5.2. Semantic queries and feature extraction

SPARQL enables the retrieval of relational signals from graph-structured data by following connections across multiple entities. This approach allows for the detection of dependencies that are often difficult to identify in conventional relational tables. To operationalize the ontology, we queried the knowledge graph using SPARQL to identify multi-entity patterns that cannot be easily captured via flat-table joins. In marketing activation settings, standard relational joins frequently overlook interactions involving promotional timing, partner roles, and cross-channel dynamics.

Through structured queries, information on channel activity, promotion periods, seasonal sales variation, and partner participation was integrated into a unified semantic framework. The outputs were converted into structured variables derived from the relationships among the Product, Promotion, Channel, and Partner entities. Instead of being manually engineered as isolated variables, these features were systematically derived from formally defined entity relationships and aggregated into measures of activation intensity, cross-entity interaction strength, and temporal patterns. Because these indicators were based on explicit relationships, they retained clear business meaning while providing additional structural context for the predictive modeling stage.

Representative SPARQL query logic used to generate semantic features included channel-level promotion ROI extraction, seasonal sales extraction by channel, and brand–partner–channel interaction extraction, as illustrated below.

```
Query 1. Channel-level promotion ROI extraction
SELECT channel, promotion type,
    SUM(sales amount),
    SUM(promotion cost),
    ROI = (sales − promotion cost) / promotion cost
GROUP BY channel and promotion type

Query 2. Seasonal sales extraction by channel
SELECT channel, month,
    SUM(sales amount)
GROUP BY channel and month

Query 3. Brand–partner–channel interaction extraction
SELECT brand, channel,
    COUNT(DISTINCT partner)
GROUP BY brand and channel
```

In total, 52 semantic features were generated from the knowledge graph using SPARQL queries. These features captured channel activity, promotion intensity, partner involvement, seasonal variation, and cross-entity interaction patterns. Before modeling, the feature set was reviewed to remove variables with near-zero variance and reduce redundancy. Highly correlated features and those exceeding acceptable multicollinearity thresholds were excluded to improve model stability. After this screening step, 38 semantic variables were retained and combined with the conventional transactional features for prediction.

5.3. Model training and evaluation

The experiments were conducted using a 70:30 train–test split. Categorical variables were encoded using one-hot encoding, while numerical variables were standardized with StandardScaler. Model optimization was performed using a Grid Search combined with 5-fold cross-validation. Performance was assessed using RMSE, MAE, and adjusted R^2. Before model training, a structured data pre-processing pipeline was applied, including handling missing values, outlier removal, feature encoding, feature scaling, multicollinearity screening, and cross-validation to ensure model robustness. The detailed pre-processing configuration is summarized in Table 1.

Table 1: Data preprocessing steps and parameter settings

Process	Method
Missing values	Median imputation
Outliers	IQR filtering
Categorical variables	One-Hot Encoding
Numerical variables	StandardScaler
Multicollinearity	VIF < 5
Train/test split	70% / 30%
Validation	5-fold cross-validation
Reproducibility	Random seed = 42

To ensure optimal model performance, key hyperparameters were tuned for each algorithm using a systematic search strategy. The primary hyperparameter configurations used in this study are shown in Table 2.

Table 2: Hyperparameter configurations used for machine learning models

Model	Key parameters
Gradient Boosting	n_estimators = 500, learning_ rate = 0.05, max_depth = 5, subsample = 0.8
Random Forest	n_estimators = 1000, max_depth = 7, min_samples_leaf = 2, max_features = sqrt
Multiple Linear Regression	No hyperparameters tuned (ordinary least squares baseline)

As shown in Table 3, Gradient Boosting achieved the lowest RMSE (116.93) and the highest adjusted R^2 (0.27), indicating superior predictive accuracy and explanatory performance. Random Forest produced the lowest MAE (7.74), suggesting stable average prediction errors, although its adjusted R^2 remained lower than that of Gradient Boosting. Multiple Linear Regression demonstrated the weakest performance across RMSE, MAE, and adjusted R^2.

Table 3: Performance comparison of machine learning models based on RMSE, MAE, and adjusted R^2

Model	RMSE ↓	MAE ↓	Adj R^2 ↑
Gradient Boosting	116.93	10.15	0.27
Random Forest	123.89	7.74	0.18
Multiple Linear Regression	143.55	16.14	−0.096

Although the adjusted R^2 values are moderate, such levels of explanatory power are common in enterprise forecasting contexts, where activation outcomes are influenced by multiple partially observable and context-specific factors. The semantic variables were derived from explicitly defined relationships in the ontology and consistently appeared among leading predictors across validation folds. Their contribution is further supported by consistent performance patterns observed across validation folds, suggesting that formalizing these relationships adds structural depth rather than merely increasing the number of predictors. Marketing activities rarely operate independently. Channels, partners, and promotional mechanisms interact, and their influence evolves. By explicitly encoding these connections, the knowledge graph provides a structured relational context that ensemble models appear better equipped to utilize than linear approaches. The observed performance patterns are consistent with the hypothesis that structured relational representations improve the modeling of interaction dynamics. Each hypothesis was examined using performance indicators aligned with the study's regression framework. H1 was evaluated by comparing RMSE, MAE, and adjusted R^2 across conventional tabular feature configurations and semantically enriched specifications. H2 was assessed through reductions in RMSE and increases in adjusted R^2, reflecting improved modeling of cross-entity interaction effects. H3 was examined through cross-validation stability and the consistency of SHAP-based feature importance patterns. The results are summarized in Table 4.

Table 4: Summary of hypothesis testing results

Hypothesis	Justification	Result
H1. Ontology improves model performance	Lower RMSE and higher adjusted R^2 were observed.	Supported
H2. KG features improve interaction modeling	Cross-entity variables improved RMSE and adjusted R^2.	Supported
H3. KG semantic features improve prediction accuracy	Gradient Boosting achieved the best performance (RMSE = 116.93; adjusted R^2 = 0.27).	Supported
H4. KG + interpretable ML improves transparency	SHAP clarified variable contributions and improved interpretability.	Supported

5.4. Discussion

This study examined changes that occur when marketing activation data are reorganized using an ontology and expressed as a knowledge graph before modeling. Once relationships among brands, channels, partners, incentives, and sales were explicitly defined, specific interaction patterns became easier to trace. In conventional tables, these links are often dispersed across separate fields and systems. Introducing relational signals into the feature set allowed the models to account for cross-entity dependencies more directly.

The improvement in predictive performance was moderate. Given the complexity of activation environments, this outcome is reasonable. Sales responses depend on many contextual and organizational factors that structured enterprise data cannot fully represent. When baseline models already include detailed transactional variables, substantial increases in explained variance are unlikely. Even so, the semantic variables appeared consistently across validation folds. Their recurring importance suggests that they reflect stable interaction patterns rather than random variation.

Gradient Boosting achieved the strongest performance (RMSE = 116.93; adjusted R^2 = 0.27). The relative advantage of ensemble methods indicates that activation-related sales behavior involves nonlinear and interaction effects. In addition, a comparison between predicted and observed sales shows that the models generally followed the overall sales trend. Deviations became more noticeable during peak activation periods, when promotional intensity and channel overlap were high. These larger errors likely reflect contextual and behavioral factors that are not fully captured in structured enterprise data. This pattern is consistent with the moderate adjusted R^2 values reported and suggests that the models capture the core relational structure while some external influences remain unobserved. Linear models can approximate these dynamics, but ensemble methods appear more adaptable when relationships become layered or context-dependent.

The semantic layer's contribution is primarily structural. Transactional features describe individual events; ontology-derived variables describe how those events relate to one another. By formalizing promotion–channel–partner relationships, the models receive contextual information that would otherwise require extensive manual engineering. Although a separate ablation analysis isolating relational-only variables was not conducted, the repeated improvements observed in semantically enriched models suggest that these variables contributed additional explanatory structure relative to conventional tabular representations.

The results also relate to the theoretical perspectives underlying the study. Structuring knowledge aligns with the RBV view that organized information resources strengthen analytical capability. Modeling evolving relationships resonates with dynamic capability reasoning. The TOE perspective further emphasizes that technological value depends on integration with organizational processes. In this sense, the ontology and knowledge graph operate as components of a broader analytical configuration rather than isolated tools.

From a managerial perspective, the framework provides a clearer view of how promotional strategies, partner participation, and channel configurations interact. The knowledge graph enhances visibility of relational patterns, while SHAP-based analysis clarifies the contribution of individual variables to predicted sales outcomes.

Rather than focusing on large metric improvements, the study highlights the value of making relational dependencies explicit within the analytical process. Strengthening structural clarity supports a more coherent interpretation of activation dynamics.

Several limitations remain. The effectiveness of semantic features depends on the completeness and accuracy of the ontology. The study was conducted within a single enterprise setting, limiting generalizability beyond that setting. In addition, the predictive improvements, while consistent, remain moderate, indicating that behavioral and environmental factors beyond the available enterprise data continue to influence activation outcomes.

6. Conclusions

This study examined whether reorganizing marketing activation data using an ontology and a knowledge graph adds analytical value before machine learning modeling. The process began by identifying and formally structuring the core elements of the activation environment: brands, channels, promotions, partners, and sales relationships. Encoding these elements within a semantic framework enabled exploration of relational patterns that are less visible in conventional tabular representations.

When semantic variables derived from this structure were incorporated into predictive models, performance improved modestly, and model behavior became easier to interpret. Gradient Boosting produced the strongest and most stable results (RMSE = 116.93; adjusted R^2 = 0.27), indicating the presence of nonlinear and interaction-based dynamics in activation forecasting.

Beyond numerical gains, the results suggest that expressing domain knowledge in a structured semantic form can support a more coherent analytical workflow. The study also illustrates how capability-oriented and alignment-based theoretical perspectives can inform practical analytical design choices. Explicitly modeling business concepts makes relational information more analytically accessible within regression-based forecasting settings.

Overall, the findings indicate a gradual transition from analytics that rely solely on transactional tables to approaches that integrate formal domain structure with transparent machine learning methods. Although the performance improvements are incremental, the combined framework enhances interpretability, structural clarity, and decision support within marketing activation contexts. Future research could extend this framework by applying it to independent enterprise datasets to assess broader applicability and external validity.

6.1. Practical implications

This framework provides managers with clearer visibility into how brands, channels, promotions, and partners interact within activation settings. By structuring these relationships, firms can plan campaigns more systematically and evaluate partner roles within a structured relational context. Because the models are interpretable, managers can examine how key variables influence predicted outcomes rather than relying solely on opaque model outputs.

6.2. Limitations

The analysis is based on data from a single company, so the results reflect the structure and routines of that organization. Marketing systems vary widely across firms, and the role of semantic features may differ in other contexts. Applying the framework in additional companies would help clarify how broadly it can be used. The ontology covers the main entities involved in activation decisions, but it does not capture every dimension of marketing complexity. Detailed customer behavior, competitor dynamics, and complete digital journey data were not included. The knowledge graph, therefore, represents only part of the activation environment rather than the entirety. Model stability was assessed through cross-validation. Although the results were generally consistent, ensemble models can still be sensitive to the structure of a particular dataset. Testing the framework on independent datasets would provide more substantial evidence of robustness. In addition, broader contextual influences such as economic conditions or seasonal shifts were not explicitly modeled and may affect activation performance. While SHAP improved transparency, further methodological work could help clarify how semantic features contribute to predictive changes.

6.3. Future research

Further work could broaden the ontology to include additional dimensions such as customer journey stages, competitor activity, and more detailed online-offline interactions. Applying the framework in different industries and geographic settings would clarify how it performs under varied market structures and organizational practices. Methodologically, there is room to experiment with graph-based learning approaches and more dynamic forms of reasoning that better reflect rapidly changing marketing conditions. Closer integration with existing BI systems is another practical direction. Linking the semantic layer with operational dashboards could allow managers to view real-time data alongside structured relational insights and interpretable model outputs. This would help connect prediction results more directly with day-to-day decision processes.

List of abbreviations

BI	Business intelligence
CRM	Customer relationship management
DCT	Dynamic Capability Theory
ERP	Enterprise resource planning
FMCG	Fast-moving consumer goods
IQR	Interquartile range
KG	Knowledge graph
MAE	Mean absolute error
ML	Machine learning
OLS	Ordinary least squares
OWL	Web Ontology Language
POS	Point of sale
R^2	Coefficient of determination
RBV	Resource-Based View
RDF	Resource Description Framework
RMSE	Root mean squared error

ROI	Return on investment
SHAP	Shapley Additive Explanations
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Structured Query Language
TOE	Technology–Organization–Environment
VIF	Variance inflation factor

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Barney J (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1): 99-120. https://doi.org/10.1177/014920639101700108

Chatterjee S, Rana NP, Tamilmani K, and Sharma A (2021). The effect of AI-based CRM on organization performance and competitive advantage: An empirical analysis in the B2B context. Industrial Marketing Management, 97: 205-219. https://doi.org/10.1016/j.indmarman.2021.07.013

Chis A, Ghiran AM, and Alter S (2024). Informing enterprise knowledge graphs with a work system perspective. International Journal of Conceptual Modeling, 19(7): 1-21.

Gomez-Perez JM, Pan JZ, Vetere G, and Wu H (2017). Enterprise knowledge graph: An introduction. In: Pan J, Vetere G, Gomez-Perez J, and Wu H (Eds.), Exploiting linked data and knowledge graphs in large organisations: 1–14. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-45654-6_1

Hastie T, Tibshirani R, and Friedman J (2009). The elements of statistical learning: Data mining, inference, and prediction. 2nd Edition, Springer, New York, USA. https://doi.org/10.1007/978-0-387-84858-7

Hogan A, Blomqvist E, Cochez M et al. (2021). Knowledge graphs. ACM Computing Surveys (CSUR), 54(4): 71. https://doi.org/10.1145/3447772

Ji S, Pan S, Cambria E, Marttinen P, and Yu PS (2022). A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems, 33(2): 494-514. https://doi.org/10.1109/TNNLS.2021.3070843 **PMid:33900922 PMCid:PMC10643713**

Kotler P, Brady M, Goodman M, and Hansen T (2019). Marketing management: European edition. 4th Edition, Pearson Education Limited, Harlow, UK.

Liang K, Meng L, Liu M, Liu Y, Tu W, Wang S, Zhou S, Liu X, Sun F, and He K (2024). A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12): 9456–9478. https://doi.org/10.1109/TPAMI.2024.3417451 **PMid:38941209**

Lundberg SM and Lee SI (2017). A unified approach to interpreting model predictions. In the Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran Associates Inc., California, USA: 4765-4774.

Makridakis S, Spiliotis E, and Assimakopoulos V (2022). M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting, 38(4): 1346-1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

Mariani MM, Perez-Vega R, and Wirtz J (2022). AI in marketing, consumer research and psychology: A systematic literature review and research agenda. Psychology & Marketing, 39(4): 755-776. https://doi.org/10.1002/mar.21619

Nickel M, Murphy K, Tresp V, and Gabrilovich E (2016). A review of relational machine learning for knowledge graphs. Proceedings of the IEEE, 104(1): 11-33. https://doi.org/10.1109/JPROC.2015.2483592

Paulheim H (2016). Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web, 8(3): 489-508. https://doi.org/10.3233/SW-160218

Ruan T, Xue L, Wang H, Hu F, Zhao L, and Ding J (2016). Building and exploring an enterprise knowledge graph for investment analysis. In: Groth P, Simperl E, Gray A et al. (Eds.), The semantic web – ISWC 2016: 418–436. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-46547-0_35

Staab S and Studer R (2016). Handbook on ontologies. 2nd Edition, Springer, Berlin, Germany.

Teece DJ (2016). Dynamic capabilities and entrepreneurial management in large organizations: Toward a theory of the (entrepreneurial) firm. European Economic Review, 86: 202-216. https://doi.org/10.1016/j.euroecorev.2015.11.006

Tornatzky LG, Fleischer M, and Chakrabarti AK (1990). The processes of technological innovation. Lexington Books, Lexington, USA.

Wang Q, Mao Z, Wang B, and Guo L (2017). Knowledge graph embedding: A survey of approaches and applications. IEEE Transactions on Knowledge and Data Engineering, 29(12): 2724-2743. https://doi.org/10.1109/TKDE.2017.2754499

Wedel M and Kannan PK (2016). Marketing analytics for data-rich environments. Journal of Marketing, 80(6): 97-121. https://doi.org/10.1509/jm.15.0413

Wernerfelt B (1984). A resource-based view of the firm. Strategic Management Journal, 5(2): 171-180. https://doi.org/10.1002/smj.4250050207