

Using machine learning to determine factors affecting product and product–service innovation

Oscar F. Bustinza and Luis M. Molina Fernandez
University of Granada, Granada, Spain, and

Marlene Mendoza Macías
Universidad Católica de Santiago de Guayaquil, Guayaquil, Ecuador

Machine
learning, product
innovation and
PSI

Received 30 June 2023
Revised 5 October 2023
11 January 2024
Accepted 4 February 2024

Abstract

Purpose – Machine learning (ML) analytical tools are increasingly being considered as an alternative quantitative methodology in management research. This paper proposes a new approach for uncovering the antecedents behind product and product–service innovation (PSI).

Design/methodology/approach – The ML approach is novel in the field of innovation antecedents at the country level. A sample of the Equatorian National Survey on Technology and Innovation, consisting of more than 6,000 firms, is used to rank the antecedents of innovation.

Findings – The analysis reveals that the antecedents of product and PSI are distinct, yet rooted in the principles of open innovation and competitive priorities.

Research limitations/implications – The analysis is based on a sample of Equatorian firms with the objective of showing how ML techniques are suitable for testing the antecedents of innovation in any other context.

Originality/value – The novel ML approach, in contrast to traditional quantitative analysis of the topic, can consider the full set of antecedent interactions to each of the innovations analyzed.

Keywords Machine learning, Product–service innovation, Servitization, Business analytics

Paper type Research paper

1. Introduction

In the domain of artificial intelligence (AI), technologies encompass the identification of precise information management challenges, the introduction of computational models to address these and the subsequent development of algorithms (Liebregts *et al.*, 2023). AI notably differs from traditional statistical and operational research methods due to its adaptability to new data streams and assimilation of existing knowledge (Rodríguez-Espíndola *et al.*, 2020). Key AI technologies span natural language processing, computer vision (e.g. object detection, image classification), robotics and expert systems that utilize software programs for decision-making and problem-solving. Among these, machine learning (ML) techniques are critical, focusing on algorithms enabling systems to learn and enhance innovation performance using data (Hastie *et al.*, 2009; Marr, 2019). Projections foresee the total economic impact of AI technology and ML techniques reaching \$15.7 trillion by 2030, with 45% of gains attributed to improvements in innovation (Rao and Verweij, 2017).

© Oscar F. Bustinza, Luis M. Molina Fernandez and Marlene Mendoza Macías. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Grant C-SEJ-020-UGR23 funded by Consejería de Universidad, Investigación e Innovación and by ERDF Andalusia Program 2021-2027.



Hence, integrating ML into innovation processes holds promise in fostering more effective strategies, enabling adaptation to competitive landscapes and enhancing overall innovation capabilities and performance (Mikalef *et al.*, 2023).

Incorporating ML techniques into firms' innovation processes offers significant potential for crafting more effective strategies, aiding firms in navigating intense market competition and handling the increasing volume of available data (Mariani *et al.*, 2023). The rise of innovation ecosystems has intensified competition, shifting away from traditional process and product evaluations to swiftly translating customer needs into innovative product-service bundles through product-service innovation (PSI). In this context, ML plays a pivotal analytical role in fostering processes, capabilities and improving the offerings of new and updated products and services (Sjödin *et al.*, 2020). These evolving ecosystems have identified PSI as a distinct area in innovation research (Baines *et al.*, 2017; Bustinza *et al.*, 2022; Kowalkowski *et al.*, 2017a, b; Rabetino *et al.*, 2018), encompassing diverse technology-driven business models focused on gaining a competitive edge by delivering knowledge-based customer services throughout the lifecycle of manufacturing products (Bustinza *et al.*, 2019). The evolving landscape of innovation ecosystems advocates for new theoretical and methodological approaches, calling for an enhanced understanding of the implications of incorporating these new innovative forms, such as PSI (Kolagar *et al.*, 2022) and ML techniques into the traditional spectrum of theoretical (Rabetino *et al.*, 2021) and quantitative methods (Lindner *et al.*, 2022).

In studies concerning product innovation and PSI, the choice of suitable methodologies for acquiring and analyzing empirical evidence has consistently sparked debate. These methodologies hold a critical role in evaluating the validity of foundational theories (Cornelissen, 2017; Wolstenholme, 1999). The discourse commonly revolves around the appropriateness of regression methods, widely employed in quantitative analyses within this domain (Cerniglia and Fabozzi, 2020). It is crucial to acknowledge that there's no universal method or standardized rules dictating the use of regression methods across the majority of empirical studies in innovation. Scientific inquiry hinges on both deductive and inductive reasoning. While deductive reasoning tests hypotheses by deriving logical consequences and comparing them to empirical data, a significant part of scientific exploration leans on inductive, probabilistic explanations, particularly when confronting complex phenomena (Prasad and Prasad, 2002). Theories endeavor to interpret intricate phenomena like PSI by proposing fundamental laws or principles that govern them. These theories utilize these principles to elucidate observed regularities and frequently predict new, analogous patterns (Hempel, 1966). Therefore, employing appropriate methodological approaches becomes crucial in comprehending and interpreting complex phenomena marked by diverse and intricate attributes, including various forms of innovation such as product innovation (Vendrell-Herrero *et al.*, 2023) or PSI (Kowalkowski *et al.*, 2017a, b). In the rapidly evolving landscape of innovation, the incorporation of ML techniques represents a novel frontier, revolutionizing traditional approaches to innovation processes by harnessing the potential of cutting-edge analytical tools and methodologies.

The traditional approach of incorporating all seemingly relevant variables into regression models encounters substantial challenges in contemporary managerial research (Kalmins, 2022) due to the need for extensive databases. Moreover, attempts to construct comprehensive regression models often exacerbate multicollinearity, diminishing result stability and interpretability (Kalmins, 2018). Furthermore, the intricate nature of variable relationships demands mediated and moderated models, further complicating result stability and hindering definitive interpretations (Johnston *et al.*, 2018). As scientific inquiry, crucial for unraveling complex phenomena like PSI, involves both deductive and inductive reasoning, the potential of inductive open-ended studies emerges as a promising avenue to comprehend such intricacies. Open-ended studies foster fluid and unrestricted data collection,

liberating researchers from rigid confines and facilitating a holistic and adaptable analysis of phenomena (Schmenner and Swink, 1998). While open-ended studies are focused on understanding complex phenomena without pre-defined hypotheses, the challenge arises when complex phenomena analysis models should incorporate all seemingly relevant and observable variables (Kalnins, 2018). Therefore, new methodological approaches are needed to overcome regression analysis limitations (Kalnins, 2022).

All of this has led to the advocacy for new methods capable of overcoming these obstacles and bolstering research outcomes (Tonidandel *et al.*, 2018). Several high-impact journals have released editorials or calls for papers aimed at introducing these ML techniques (Lindner *et al.*, 2022; Pagell *et al.*, 2019), signifying an increasing recognition of the necessity for sophisticated analytical approaches. These novel ML methodologies, proposed as efficient solutions, hold promise in enhancing research, particularly in areas such as breakthrough research identification (Li *et al.*, 2022) and tackling the limitations of classical regressions, notably concerning multicollinearity (Kalnins, 2022). These techniques offer an alternate path for advancing knowledge in innovation research. As Tonidandel *et al.* (2018, p. 534) stated: *“(this) new epistemological approach seeks to develop knowledge from data as opposed to using the data to test existing theory”*. In alignment with this perspective, our study does not commence with preconceived theories indicating the most relevant variables for predicting innovative outcomes. Instead, we leverage ML tools to uncover the attributes and their relationships that best elucidate those primary outcomes, specifically product innovation and PSI innovation (Bustinza *et al.*, 2019; Gault, 2018; Slater *et al.*, 2014; Snyder *et al.*, 2016). This leads us to propose the following research questions:

- RQ1. What factors exert the most positive and negative influences on companies' product and PSI outcomes from a pool of over 500 initial variables?
- RQ2. How are the patterns that emerge from the interplay of those factors related to existing theories on innovation?

To address the research questions, we utilize the National Survey on Science, Technology and Innovation Activities in Ecuador (ACTI), which offers data on key indicators related to scientific research, technological development and innovation in the country, as well as information on human resources and assets dedicated to these activities. This survey adheres to the Oslo Manual (OCDE, 2005), a widely recognized conceptual and methodological framework for gathering firm-level data on innovation activities within a specific national system.

2. Methodology

2.1 Machine learning

Data analytics employs statistical methods to uncover patterns and extract meaning from data (Liebregts *et al.*, 2023). This process extends to predictive analytics, which leverages advanced techniques to analyze historical data, utilizing data mining models and ML algorithms to reveal potential scenarios. These predictive analytics techniques combine new algorithms with traditional models like regression to assign probabilities to instances for outcome classification. Finally, prescriptive analysis is built upon the patterns uncovered during the predictive analysis, providing actionable decision options and future opportunities within the analyzed context. In this context, ML is “the subset of artificial intelligence devoted to defining computer algorithms that automatically improve through experience” (Liebregts *et al.*, 2023).

The core objective of ML is to craft a model capable of accurately predicting outcomes based on a given set of input or predictor variables. It is important to note that the data used in ML is typically divided into three subsets: (1) The training set: This subset is used to train the model by feeding it with input variables and their corresponding outcomes. The model learns from this data to understand the underlying patterns and relationships. (2) The validation set:

After training the model, it is necessary to fine-tune its performance. The validation set is employed for this purpose, allowing the classifier to undergo feature selection and data balancing techniques. These steps help optimize the model's predictive capabilities. (3) The test set: Once the model has been trained and validated, it needs to be evaluated on unseen data (known as out-of-sample data) to assess its performance. Notably, supervised ML entails out-of-sample validation, a distinguishing feature when contrasted with the traditional econometric regression, which primarily focuses on in-sample fitting (Chou *et al.*, 2022). The test set serves as a final benchmark, providing an estimation of how well the model can predict outcomes when faced with new input variables.

Moreover, preceding model training, ML entails several preparatory steps to ensure a robust model foundation, including:

- (1) *Data cleaning and missing value handling*: The training data may contain errors, outliers, or missing values. Therefore, it is necessary to perform data cleaning, which involves removing or correcting inaccurate data points and handling missing values appropriately, either by imputing them or excluding the corresponding inputs.
- (2) *Input scaling*: In order to standardize the range and distribution of the inputs, input scaling techniques are applied. This typically involves two common methods: min-max normalization, which scales the values to a specific range (e.g. 0–1) and standardization, which transforms the inputs to have zero mean and unit variance (Han *et al.*, 2012). This step ensures that all inputs are on a comparable scale and prevents certain features from dominating the learning process due to their larger magnitude.
- (3) *Input selection (if necessary)*: Depending on the nature and complexity of the problem, it may be beneficial to select a subset of relevant inputs. Input selection techniques help identify the most informative and discriminative inputs, reducing the dimensionality of the input space and potentially improving model performance.
- (4) *Data balancing*: In situations where the training data is imbalanced, meaning there is a significant difference in the number of inputs belonging to different classes (e.g. positive and negative outcomes), it is crucial to address this issue. Data balancing techniques, such as resampling, are employed to adjust the class distribution and equalize the number of inputs in each class (He and Garcia, 2009). This ensures that the model is not biased toward the majority class and can effectively learn from both positive and negative outcomes.

As to the ML approaches and depending on the data, various techniques can be selected, namely “*supervised learning*,” “*unsupervised learning*,” “*semi-supervised learning*,” and “*reinforcement learning*” (Dasgupta and Nath, 2016). In this paper, we will utilize supervised learning algorithms since we have knowledge of the output to be predicted (product innovation and PSI). Within the set of supervised learning algorithms, the main groups are as follows: *Rule Induction* algorithms such as ZeroR (often considered the baseline algorithm), PART, OneR, Decision table, or Jrip; *Decision Tree and CART* (Classification and Regression tree) algorithms such as SimpleCart, REPTree, LMT, J48, HoefidingTree, or RandomForest; *Nearest Neighbors* algorithms such as Lazy.Ibk; *Bagging/Boosting* ensemble algorithms such as Bagging, or AdaBoostM1; *Neural Networks* algorithms such as RBFNetwork; *Support Vector Machines* (SVM) algorithms such as LibLinear; *Statistical Classifiers* algorithms such as logistic, NaiveBayes, or BayesNet (Stockdale and Standing, 2006).

2.2 Data selection

For our analysis, we used data from the National Survey on Science, Technology and Innovation Activities conducted by the Government of Ecuador (<https://www.ecuadorencifras.gob.ec/>)

[encuesta-nacional-de-actividades-de-ciencia-tecnologia-e-innovacion-acti/](#)). This database includes a total of 566 fields of information about 6,275 companies, the majority of which are related to innovative activities, as the survey adheres to the standards of the Oslo Manual (OCDE, 2005). The outcome objectives of this study are product innovation and PSI, which are measured by the introduction/commercialization of new products and services. Product and PSI are concepts critical in the field of innovation (Mendoza-Silva, 2020; Vendrell-Herrero *et al.*, 2023), as well as in the analysis of competitive position in the market (Baines *et al.*, 2009; Bustinza *et al.*, 2015; Gunday *et al.*, 2011).

Subsequently, leveraging Python along with libraries such as Pandas, Scikit-learn, Weka and Statsmodels, we conducted an in-depth analysis of the dataset. The missing data, which was limited in quantity, was handled as follows: first, we identified the columns with null values using the “isnull()” function. Then, we applied an imputation technique using the “KNNimputer()” function, where the missing values were predicted based on the mean of neighboring data points. This approach allowed us to retain as much data as possible, and subsequently, we applied appropriate techniques to select the attributes (inputs) that best predict the classes (outcomes). Regarding the number of manufacturers and servitized manufacturers, there were 1,412 instances (observations) of pure manufacturers and 177 instances of servitized manufacturers. One of the advantages of using ML techniques is that researchers do not need to preselect predictors and their interactions, as is required in traditional regression techniques. ML tools incorporate “meta” classifiers such as *AttributeSelectedClassifier*, which provides an automated feature selection tool that is particularly useful when working with databases containing a large number of inputs, such as the one used in this study. Meta-classifiers are algorithms designed to enhance the performance and address specific challenges in classification tasks by either combining or modifying predictions generated by other base algorithms. Several prominent meta-classifiers (Fernández-Delgado *et al.*, 2014) include:

- (1) *Bagging (bootstrap aggregating)*: Bagging creates an ensemble of base algorithms by training them on various bootstrap samples derived from the original training data. The predictions of these classifiers are subsequently combined through methods such as majority voting or averaging to formulate the final prediction.
- (2) *AdaBoostM1 (adaptive boosting)*: AdaBoost iteratively adjusts the weights of training inputs based on the performance of the base algorithms. AdaBoostM1 amalgamates predictions from multiple weak algorithms, often decision trees, by assigning greater weights to misclassified inputs, thereby constructing a robust classifier.
- (3) *RandomSubSpace*: RandomSubSpace introduces random feature subsetting during the training of base algorithms. This entails the random selection of a subset of inputs for each base algorithm, enabling them to specialize in distinct input subsets. The predictions of these specialized base algorithms are then merged to generate the ultimate prediction.
- (4) *AttributeSelectedClassifier*: This approach integrates an input selection algorithm with a base algorithm. Initially, it employs an input selection method to identify a subset of relevant inputs, subsequently training the base algorithm exclusively on these selected inputs.

3. Results

In the current study, we employed 10-fold cross-validation, which involves randomly dividing the data into 10 parts, ensuring that the output is almost equally represented in each

part. The algorithm is then executed 10 times using different training datasets, and the 10 error estimates are averaged (Witten and Frank, 2002). The results, in terms of accuracy achieved by the different algorithms, are presented in Figures 1 and 2.

As shown in both Figures 1 and 2, the highest accuracy is achieved by Trees and Rules algorithms. In the case of Trees, LMT (Logistic Model Trees) is a logarithm for building classification trees with logistic regression functions at the leaves. J48 is the traditional C4.5 algorithm (Quinlan, 1993) used for generating pruned or unpruned decision trees. SimpleCart implements minimal cost-complexity pruning to the tree, while RandomForest uses a greedy algorithm that splits the data at the best point for each step of the tree building process. Regarding Rules algorithms, PART reports a decision list through separate-and-conquer searches. JRIP (Repeated Incremental Pruning to Produce Error Reduction) is a variant of the

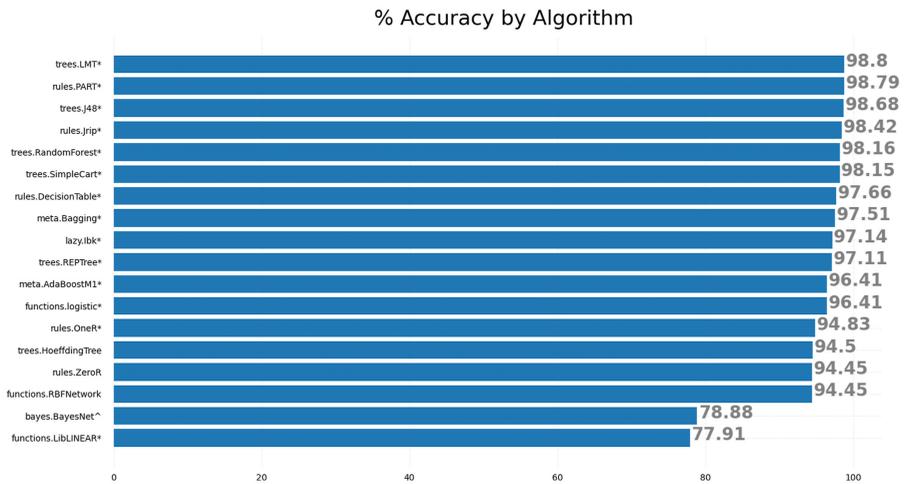


Figure 1. Accuracy achieved by the different algorithms when the outcome is product innovation

Source(s): Authors work

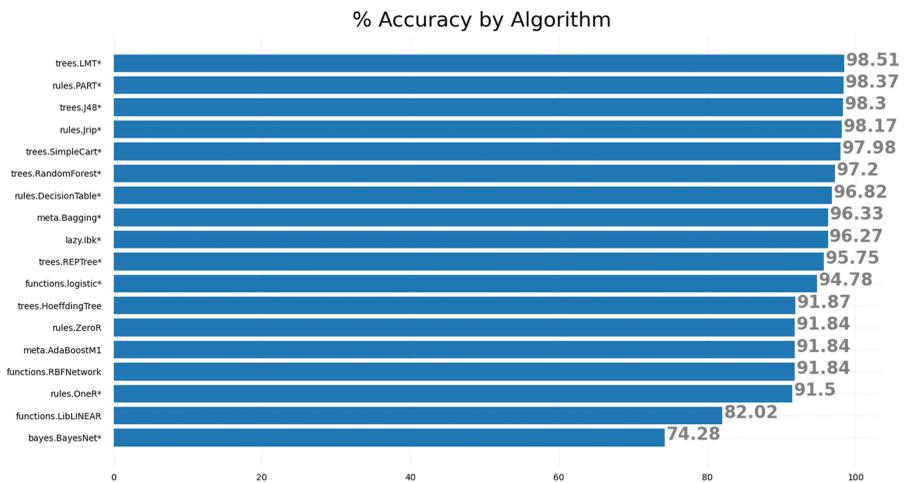


Figure 2. Accuracy achieved by the different algorithms when the outcome is PSI

Source(s): Authors work

RIPPER algorithm (Cohen, 1995). Finally, DecisionTable is a simple decision table majority classifier (Witten and Frank, 2002).

When examining the accuracy achieved by different algorithms, it is crucial to consider all predictors, even those potentially insignificant in the analysis. To identify the set of predictors that truly contribute to predicting the outcome, especially in cases with a large number of inputs, the *AttributeSelectedClassifier* is used. This classifier offers various evaluators, such as *CfsSubsetEvaluation* which evaluates the worth of a subset of inputs by considering both the individual predictive ability of each predictor and the degree of redundancy between them. Other evaluators are *GainRatioAttributeEval* or *WrapperSubsetEval*, which use a learning algorithm as the baseline (e.g. ZeroR), among others. Additionally, three different search techniques are available: *BestFirst*, *GreedyStepwise* and *Ranker*.

Among all the evaluators and search techniques, *GainRatioAttributeEval*, which evaluates the worth of a predictor by measuring the gain ratio with respect to the outcome, combined with the *Ranker* search technique, which ranks predictors based on their individual evaluations, consistently offered the highest accuracy for most of the algorithms. This search technique provided the ranked contribution of the most important inputs for predicting product and PSI (see Table 1). All predictors are grounded in the literature, as the survey follows the OECD standards for analyzing innovation. In our case, the inputs or predictors unveiled by the ML algorithms are related to two different innovation determinants:

- (1) *Sources of information* (Amara and Landry, 2005): They comprises the internal sources of information and knowledge such as headquarter (HQ) cooperation and in-house R&D departments, the external or Open Innovation sources (Chesbrough, 2006; Laursen and Salter, 2006; Vendrell-Herrero *et al.*, 2023) and the specialized and generally available sources as trade fairs and exhibitions, technical standards licensed through Intellectual Property Patents (IPO), or safety and environmental standards such as ISO 14001.
- (2) and competitive priorities determinants: that is, the set of internal and external competitive variables that exert influence to generate innovation: (1) internal

Product innovation		Product-service innovation	
Input	Information gain	Input	Information gain
Headquarter cooperation	0.494	IPO cooperation	0.526
Other firms cooperation	0.494	Competitors cooperation	0.525
Government cooperation	0.494	University cooperation	0.525
Suppliers cooperation	0.494	Total percentage of funding	0.525
Total percentage of funding	0.494	Government cooperation	0.525
IPO cooperation	0.473	Clients cooperation	0.525
Ind. manufacturing sector	0.418	Other firms cooperation	0.525
University cooperation	0.398	Labs cooperation	0.449
Labs cooperation	0.357	Headquarter cooperation	0.367
Customers cooperation	0.320	Customers cooperation	0.315
Competitors cooperation	0.312	Suppliers cooperation	0.311
Clients cooperation	0.299	Obj.: Increase quality	0.277
Obj.: Incr. mark. share	0.248	Objective: Incr. flexibility	0.203
Obj.: reduce environ. impact	0.239	Obj.: Incr. mark. share	0.172
R&D Department	0.144	Obj.: Entry into new markets	0.169
Obj.: entry into new markets	0.122	Obj.: Reduce environ. impact	0.151

Source(s): Authors' own work

Table 1.
Attribute gain ratio
with respect for
product innovation
and PSI

competitive priorities of the firm (Alegre-Vidal *et al.*, 2004; Hayes, 1984) such as cost efficiency, quality, delivery and flexibility, and (2) external market orientation strategies such as increasing market share, entering new markets, or reducing environmental impact (Appiah-Adu and Ranchhod, 1998; Tajeddini *et al.*, 2006).

In exploring the determinants of information sources encompass a multi-dimensional perspective related to the origins of information and knowledge required for innovation activities: (1) Internal Sources: These sources emanate from within an organization, frequently stemming from in-house departments that contribute valuable insights for product development and innovation. Collaborative efforts with headquarters, involving various functions such as marketing, production, or management staff, can also be of instrumental importance (Amara and Landry, 2005). These internal sources originate from the organization's own activities, operations and resources, assuming a pivotal role in decision-making, problem-solving and process enhancement; (2) Market sources that encompass knowledge-driven innovations derived from interactions with suppliers, clients, competitors, consultants and commercial laboratories. The well-established literature underscores the significance of involving lead users or customers in the innovation process (von Hippel, 2007). Their participation is highly esteemed for the complementary skills and knowledge they contribute, along with their role in mitigating the risks associated with innovation development and market adoption. Additionally, suppliers enjoy widespread recognition as fundamental sources of innovation. Previous empirical studies investigating the impact of market-based cooperation on the level of innovation novelty within the manufacturing sector have consistently demonstrated a positive correlation (Mention, 2011); (3) General and Specialized Sources: These encompass a range of sources including regulations, environmental and safety standards such as ISO 14001, professional conferences, trade associations, initial public offerings (IPOs), scientific publications and professional associations. These sources have previously been associated with specific types of innovation, such as green innovation (Thao and Xie, 2023) or biotechnological innovation (Gertler and Levitte, 2005).

Furthermore, within the realm of competitive priorities shaping innovation, ever since Skinner's seminal work in (1969), the literature on operations strategy has consistently emphasized the delineation of competitive priorities through the lens of four fundamental components: low cost, quality, delivery time and flexibility (Wheelwright and Hayes, 1985). These components are defined as follows: (1) Cost Importance: The concept of cost importance revolves around the effective management of manufacturing costs, encompassing direct production expenses, productivity, capacity utilization and inventory reduction, all aimed at minimizing the monetary valuation of production (Ward *et al.*, 1998). This encompasses various facets, such as overhead costs and inventory management, all with the overarching goal of effectively controlling production costs and adding value. (2) Quality Importance: Quality importance is associated with notions of excellence, value, adherence to specifications and the ability to meet or exceed customer expectations (Reeves and Bednar, 1994). (3) Delivery Time Importance: The significance of delivery time lies in the capacity to promptly provide goods and services, adhering to promised schedules. It also encompasses considerations related to the time-to-market for new products (Leong *et al.*, 1990). (4) Flexibility Importance: Flexibility importance pertains to the capacity to deploy and/or reallocate resources in response to changes in contractual agreements, often instigated by customer demands (Phusavat and Kanchana, 2007). This encompasses various facets, including adjustments in design and planning, changes in production volume and product variety.

The evaluation of attribute gain ratio, an essential measure assessing input relevance concerning product innovation and PSI prediction, delves into the intrinsic information of

predictors while considering their information gain. It helps to identify inputs that have a strong relationship with the outcome variable. In the case of Product Innovation, the attribute gain ratio measures the importance of each input in predicting whether a company will introduce new products. Similarly, for PSI, the attribute gain ratio assesses the significance of inputs in predicting the introduction of new bundles of product-services by a company. Predictors with higher gain ratios indicate a stronger association with the respective innovation type, making them more influential in predicting the occurrence of product or PSI respectively.

Transitioning to logistic regression, this method helps identify the key predictors and their impact on the likelihood of achieving product or PSI. It provides insights into the statistical significance of the selected predictors and their contribution to the innovation outcomes. In Tables 2 and 3, the logistic regression results for both product innovation and PSI can be observed. In both cases, the relevance of variables in explaining the propensity to innovate in product and PSI is demonstrated, considering the increase in the probability of occurrence. In this analysis, the marginal effect is considered, taking into account the effect of the other variables on the propensity to innovate. Thus, considering the marginal effects, in the case of product innovation, all previously analyzed factors are significant except supplier cooperation, IPO cooperation and the competitive priority of increasing market share. In the case of PSI, all the presented factors have significant marginal effects, except competitor and other firm cooperation and the pursuit of competitive priorities such as increasing market share or flexibility. It is particularly noteworthy that in the case of PSI, supplier and lab cooperation have a negative marginal effect, making it more challenging to successfully innovate in PSI. This result contradicts the findings presented in the case of product

Logistic regression results

Dep. Variable: product innovation	Observations: 1,412 pure manufacturers
Model: Logit (Method: MLE)	Pseudo R-squ.: 0.664
AIC: 2956.710	BIC: 3064.617
converged: True (Interactions = 9)	Log-Likelihood: -1462.4
LL-Null: -4348.8	LLR <i>p</i> -value: 0.00

	Coef	Std. err	z	P> z	[0.025	0.975]	OR
Intercept	-7.483	0.246	-30.381	0.000	-7.966	-7.001	0.001
Headquarter cooperation	0.352	0.170	2.073	0.038	0.019	0.686	1.422
Other firms cooperation	0.438	0.164	2.666	0.008	0.116	0.761	1.550
Government cooperation	0.532	0.240	2.213	0.027	0.061	1.003	1.702
Suppliers cooperation	-0.144	0.114	-1.265	0.206	-0.368	0.079	0.866
Total perc. of funding	0.004	0.001	2.693	0.007	0.001	0.006	1.004
IPO cooperation	0.093	0.292	0.318	0.750	-0.480	0.666	1.097
Ind. Manuf. sector	4.106	0.122	33.542	0.000	3.866	4.346	60.703
University cooperation	0.674	0.239	2.821	0.005	0.206	1.143	1.962
Labs cooperation	0.659	0.189	3.480	0.001	0.288	1.030	1.933
Customers cooperation	1.221	0.093	13.126	0.018	1.038	1.404	3.391
Competitors cooperation	0.457	0.137	3.352	0.001	0.190	0.725	1.579
Clients cooperation	0.567	0.111	5.102	0.000	0.350	0.786	1.763
Obj.: Incr. mark. share	-0.039	0.060	-0.648	0.517	-0.157	0.079	0.962
Obj.: Reduce envi. impact	1.003	0.091	11.019	0.002	1.001	1.005	2.726
R&D Department	1.484	0.106	13.981	0.000	1.276	1.692	4.411
Obj.: Entry new mark	0.439	0.056	7.890	0.000	0.330	0.549	1.551

Source(s): Authors' own work

Table 2. Logistic regression results for product innovation

Logistic regression results

Dep. Variable: Product-service Innovation
 Model: Logit (Method: MLE)
 AIC: 3669.284
 converged: True (Interactions = 8)
 LL-Null: -3548.9

Observations: 177 servitized manufacturers
 Pseudo R-squ.: 0.488
 BIC.: 3773.938
 Log-Likelihood: -1818.6
 LLR *p*-value: 0.00

	Coef	Std. err	z	P> z	[0.025	0.975]	OR
Intercept	-4.466	0.212	-21.074	0.000	-4.882	-4.051	0.011
IPO cooperation	2.988	0.458	6.520	0.000	2.090	3.886	19.849
Competitors cooperation	0.051	0.114	-0.450	0.653	-0.275	0.172	0.950
University cooperation	0.536	0.216	2.488	0.013	0.114	0.959	1.709
Total percentage of funding	0.050	0.004	12.497	0.000	0.041	0.058	1.051
Government cooperation	0.909	0.274	3.313	0.001	0.371	1.447	2.482
Clients cooperation	0.373	0.093	4.017	0.000	0.191	0.555	1.452
Other firms cooperation	0.161	0.131	1.233	0.217	-0.095	0.418	1.175
Labs cooperation	-0.795	0.203	-3.909	0.000	-1.194	-0.396	0.451
Headquarter cooperation	0.983	0.148	6.649	0.000	0.694	1.274	2.674
Customers cooperation	0.598	0.101	5.924	0.000	0.400	0.796	1.819
Suppliers cooperation	-0.659	0.094	-7.008	0.000	-0.843	-0.475	0.517
Obj.: Increase quality	1.090	0.124	8.794	0.000	0.946	1.234	2.974
Objective: Incr. flex	0.064	0.047	1.381	0.167	-0.027	0.156	1.066
Obj.: Incr. mark. share	0.073	0.047	1.546	0.122	-0.020	0.166	1.076
Obj.: Entry new markets	0.228	0.044	5.168	0.000	0.141	0.314	1.256
Obj.: Red. env. impact	0.406	0.037	10.854	0.000	0.333	0.479	1.501

Table 3.
 Logistic regression
 results for product-
 service innovation

Source(s): Authors' own work

innovation, where the marginal effect was not significant for supplier cooperation and was positive for Lab cooperation.

Post preselection of inputs or predictors, two decision trees are developed to delineate the pathways and rules governing product innovation and PSI, thereby providing a comprehensive understanding of the process. These decision trees provide a visual representation of the factors and their relationships that contribute to each type of innovation. Decision trees has been generated by the J48 algorithm. J48 provides the ability to visualize the generated decision tree, aiding in better comprehension and interpretation. J48 is a versatile and widely utilized decision tree algorithm that offers a robust and efficient approach to classification tasks. The J48 algorithm is an implementation of the C4.5 decision tree algorithm (Quinlan, 1993). It is a popular decision tree algorithm known for its top-down, greedy approach to building decision trees from training data. The name "J48" signifies that it is a Java implementation of the C4.5 algorithm. Some noteworthy features and functionalities of the J48 algorithm includes (1) Input Selection: J48 employs information gain or gain ratio (used in the current study) as criteria for selecting inputs. These measures assess the quality of an input in terms of its ability to reduce uncertainty or impurity in the data; (2) Missing Values Handling: J48 can effectively manage missing values in the dataset by utilizing surrogate splits. Surrogate splits allow the algorithm to make predictions using alternative inputs when the primary input is missing; (3) Pruning: J48 incorporates pruning techniques, such as reduced error pruning or cost-complexity pruning, to prevent overfitting. Pruning involves the removal of nodes or branches from the decision tree to enhance its generalization capability; (4) Confidence Factor: J48 allows for the adjustment of the confidence factor parameter, which governs the extent of pruning. A higher confidence factor leads to more

aggressive pruning, resulting in smaller trees; (5) Binary and Multi-class Classification: J48 supports both binary classification (two outcome classes) and multi-class classification (more than two outcome classes) tasks.

By integrating insights from both logistic regression analysis and decision trees, a holistic comprehension of the drivers behind product and PSI emerges. These findings hold immense potential to guide strategic decision-making, offering invaluable insights for organizations seeking to bolster their innovation prowess in these domains. Figures 3 and 4, coupled with Table 4, intricately detail the outcomes derived from the decision tree analysis elucidating the triggers for innovation activation in both product innovation (Figure 3) and PSI (Figure 4). These visual representations significantly contribute to understanding the nuanced pathways to innovation. These decision trees depict the path of necessary conditions leading to an outcome, in our case, product innovation and PSI [1]. These illustrate the combination of company characteristics that result in a high probability of success for each type of innovation. For example, focusing on Figure 4, the first characteristic that appears is customer cooperation. Thus, if the company does not cooperate with its customers, the decision tree informs us that the company has low chances of success. Once it cooperates with customers, there are different ways to improve the probability of success in PSI, depending on whether the company is manufacturing and transitioning to services or vice versa, entering into goods production from services. We will delve deeper into analyzing these relationships in the results discussion.

In essence, this section encapsulates an extensive exploration comprising one regression analysis and 3 ML technique analyses conducted on a database housing over 500 explanatory variables pertinent to product innovation and PSI. Findings derived from attribute gain ratio, ML logistic regression and decision tree algorithms persistently underscore the criticality of fostering collaboration with ecosystem stakeholders alongside aligning organizational competitive priorities and market orientation. These analyses collectively unravel a

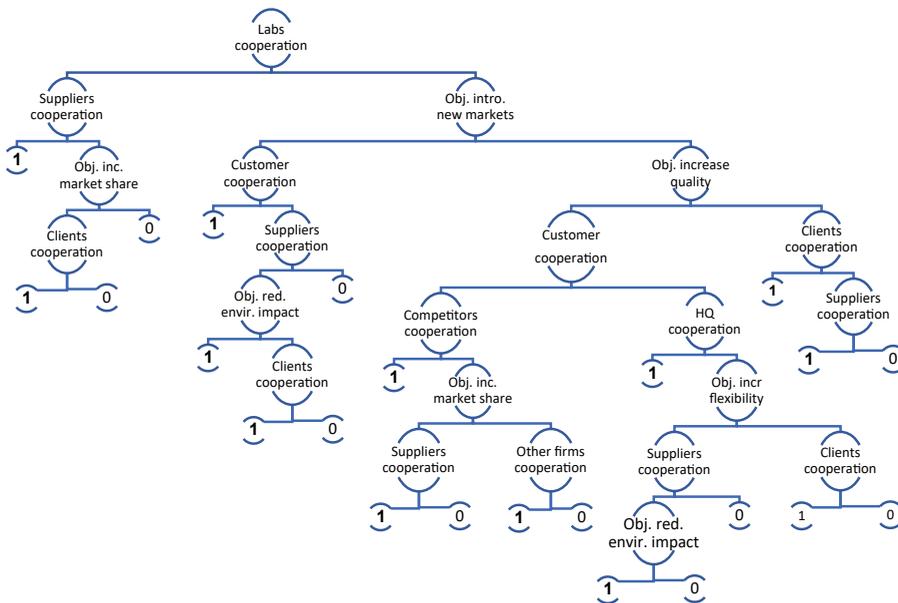


Figure 3.
Decision tree for
product innovation

Source(s): Authors work

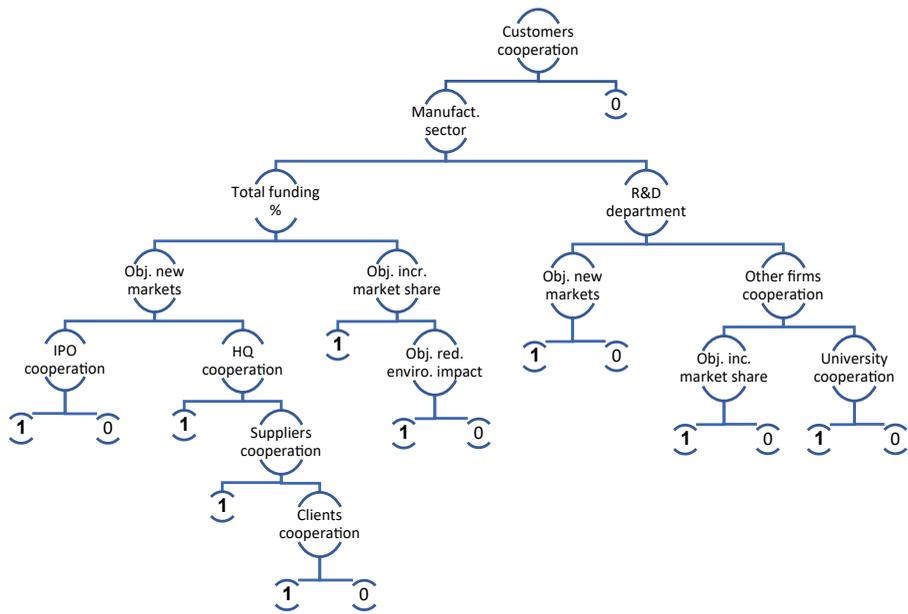


Figure 4.
Decision tree for PSI

Source(s): Authors work

multifaceted framework previously inaccessible through conventional data analysis methodologies.

4. Discussion

The discussion of these results enables us to gain a deeper understanding of the factors that exert the greatest influence on innovation success, both within product and PSI domains. By collectively examining the results, we can identify the most pertinent factors and determine if they exhibit consistency across both types of innovation. Furthermore, the techniques employed in this study allow us to explore the interactions among these factors and investigate the configurations that lead to success in both product and PSI. This analysis offers valuable insights into the intricate relationships and interdependencies between various factors. Lastly, these findings can be scrutinized within the context of innovation and manufacturing strategy theories to assess their alignment with the propositions presented by theoretical models. Through comparing the results with existing theories of innovation and the literature on competitive priorities, we can evaluate the extent to which our findings either support or diverge from these theoretical frameworks.

In previous investigations concerning ML techniques in innovation, the primary focus has been on the broader field of Information and Communication Technology (ICT) business innovation. This pertains to the systematic utilization of advancements in information and communication technologies to generate novel or refined products, services, processes, or business models within an organizational or industrial context (Yunis *et al.*, 2018). For example, building upon this foundation, Eom *et al.* (2022) conducted a comprehensive analysis to ascertain the most suitable ML techniques in terms of their predictive accuracy concerning specific innovation outcomes, such as innovation performance. In a similar vein, Lim *et al.* (2020) utilized ML tools and analytical methodologies to investigate the synergistic

relationships between product lifecycle management and business innovation. Lastly, [Nafizah et al. \(2023\)](#) conducted an evaluation of the innovation benefits resulting from diverse strategies adopted by micro-businesses when implementing AI and machine learning techniques. The results obtained in our study align with those of previous studies in terms of revealing the relationships between determinant factors and innovation outcomes.

Delving deeper into the most influential factors considered individually, it becomes evident that cooperation with other members of the innovation ecosystem emerges as pivotal in the context of the analyzed companies ([Kolagar et al., 2022](#); [Vaccaro et al., 2010](#)). Regarding product innovation, combining the results presented in [Tables 1 and 2](#), it is evident that the most relevant factors are “headquarter cooperation” and cooperation with the ecosystem members (government, universities, labs, customers, competitors, clients and other firms). On the other hand, when considering other types of cooperation, although “supplier cooperation” and “IPO cooperation” can predict the degree of product innovation according to *Attribute Gain Ratio*, they do not have a direct effect on the degree of product innovation in the logistic analysis results. This seemingly contradictory result is clarified when considering the interactions between variables, as both types of cooperation may not be individually significant but play a part in configurations that lead the company to success in product innovation. Regarding other variables with predictive capacity, they revolve around the competitive priority pursued and the percentage of funding. In the case of product innovation, if the company’s objective is reducing environmental impact ([Dangelico, 2016](#); [Szász and Seer, 2018](#)) or entering a new market ([Zhou and Li, 2008](#)), the likelihood of success in product innovation increases. However, when the objective is to increase market share ([Cooper and Kleinschmidt, 1987](#)), despite its explanatory power, it does not have individual capacity but rather operates within configurations of factors that influence product innovation success when considered collectively.

Regarding the factors that individually affect PSI, “headquarter cooperation” remains relevant ([Cenamor et al., 2017](#)). In terms of collaboration with ecosystem companies, “IPO cooperation,” “universities,” “government,” “client,” and “customer cooperation” are significant. However, when considered individually, the results for “competitor,” “lab,” “supplier,” and “other firms cooperation” are not entirely conclusive. While they are deemed relevant according to *Attribute Gain Ratio*, logistic regression results indicate a lack of significance. As to the innovation objective, similar to product innovation, if the objective is to reducing environmental impact or entering new markets, the probability of successful PSI increases. However, when considered individually, if the objective is to enhance flexibility ([Yeniaras et al., 2021](#)) or increase market share, despite their relevance according to *Attribute Gain Ratio*, they are not significant once the other explanatory factors are considered in logistic regression.

Considering the individual factors that affect both product innovation and PSI, we can observe that in both cases, “headquarter cooperation” and the objectives of reducing environmental impact or entering new markets are relevant and have individual explanatory power. The remaining factors are related to cooperation with ecosystem agents ([Kolagar et al., 2022](#)), but their relevance varies depending on the type of innovation. There are similarities and differences between the two analyses regarding the significance of “supplier cooperation” and the relevance of “government,” “universities,” “customers,” and “clients cooperation.” However, the results vary for other agents. In the case of product innovation, “IPO collaboration” is not significant, but it is in the case of PSI. On the other hand, the results for “competitor,” “lab,” and “other firm cooperation” are inconclusive for PSI, but they are significant for product innovation. These findings indicate that cooperation is highly relevant, but the company should focus on cooperating with different ecosystem agents depending on the type of innovation sought.

The analysis using decision trees allows us to delve deeper by examining the interactions that exist between these factors, thereby increasing the probability of

innovation in both products and PSI. In the case of product innovation, the most direct approach is to collaborate with suppliers and laboratories. When collaboration with labs (De Faria *et al.*, 2010) is not feasible, the network of relationships emphasizes that cooperation with companies in the business ecosystem (customers, suppliers and clients) is crucial for forming different combinations (Li, 2009). In certain instances, it is possible to foster product innovation through collaboration with competitors, other firms, or the company headquarters, but this can only be achieved under the assumption that specific competitive priorities are pursued.

Furthermore, in the realm of cooperation types leading to PSI, customer cooperation emerges as the pivotal factor (Bustinza *et al.*, 2013; Reim *et al.*, 2019). Without it, PSI cannot be achieved under any circumstances. Therefore, co-innovation with customers is a prerequisite for the successful development of this type of innovation (Sjödin *et al.*, 2020). Once this cooperation has been established, it becomes increasingly important to collaborate with external entities outside the company's business ecosystem, such as IPOs, universities, other firms and the headquarters. Cooperation with suppliers and clients is comparatively less significant, particularly when compared to the case of product innovation. In this context, collaboration with laboratories, which was pivotal for product innovation, is not relevant. These findings demonstrate that the nature of relevant collaboration in both types of innovation differs significantly.

When it comes to competitive priorities, it is evident that focusing on improving quality and increasing market share is highly relevant for product innovation, with either one of the two being crucial in almost all combinations. However, in the case of PSI, competitive priorities play a much less decisive role. Traditional manufacturing priorities such as quality and flexibility are not significant in any combination, and the remaining priorities only hold importance in very specific cases. The most noteworthy priority is entering new markets (Aquilante and Vendrell-Herrero, 2021; Lafuente *et al.*, 2023), particularly when combined with collaboration with IPOs (Liu *et al.*, 2023). This suggests that when manufacturing companies aim to innovate in product-service with the goal of entering new markets, collaborating with IPOs becomes essential, especially if they lack the necessary resources to capture from the market.

Concerning internal variables, they are only relevant in the case of PSI. The most significant variable is the distinction between manufacturing firms that adopt servitization and service firms that expand their offerings to include physical goods, that is, productization process (Leoni, 2019). The combinations of collaboration types and competitive priorities differ between these two innovation processes. For servitization or manufacturing companies seeking to introduce new services, successful PSI can be achieved through collaborations with companies in the business ecosystem (customers, clients, suppliers, or the headquarters), depending on the competitive priority. On the other hand, for productization or service companies aiming to innovate in products, there are two possibilities. They can either establish their own R&D department (Ziaee Bigdeli *et al.*, 2017), particularly if their competitive priority is entering new markets, or collaborate with universities or other companies outside the ecosystem (in addition to mandatory collaboration with customers in all cases). This finding demonstrates that not only do the combinations of collaboration and competitive priorities differ between product innovation and PSI, but they also depend on the path (servitization/productization) chosen by the company.

These results present a highly complex map of collaboration possibilities depending on the type of innovation (product or product-service) and the competitive priorities and market orientation (new market, market share, environmental performance, quality improvement, etc.). Such interactions cannot be adequately studied using traditional research techniques in the field of innovation and management. These findings hold great relevance on their own, and importantly, they highlight the intricate nature of interactions between types of internal

and external variables (Huizingh, 2011). With over 500 potential factors or explanatory inputs of innovation success introduced, the results strongly reinforce the propositions of open innovation theory (Dahlander and Gann, 2010; Enkel *et al.*, 2020). Not only do different types of cooperation have individual significance, but when analyzed collectively through decision trees, cooperation with various stakeholders based on the type and objective of innovation becomes crucial. It is essential to note that we did not start with preconceived research hypotheses regarding the most relevant factors, and the research method is not bound by requirements of normality or linearity in relationships. Therefore, the fact that most explanatory factors are associated with open innovation further bolsters the central idea of this theory, and our results provide substantial clarification in this regard.

Within this theoretical framework, a multitude of studies address the complexity of collaboration and emphasize that open innovation is not a one-size-fits-all approach. Instead, its implementation must be carefully analyzed (Xie and Wang, 2020). Our results strongly support the idea that open innovation is contingent on two main types of factors. Firstly, the way it is implemented differs between product innovation and PSI (Vendrell-Herrero *et al.*, 2023). Secondly, a commonly overlooked factor is the competitive priorities and market orientation of the firms (Tajeddini *et al.*, 2006). Our findings conclude that depending on the objective, our collaborative efforts should be tailored accordingly.

5. Conclusion

The need for improved research methods to obtain more robust results and models that better reflect complex phenomena within organizations has been widely recognized (Chou *et al.*, 2022; Lindner *et al.*, 2022; Prasad and Prasad, 2002). Our research presents an open-ended study that provides alternative options to extend beyond current methodological techniques, offering insights and conclusions that are challenging to obtain using traditional regression-based research methods. This work represents a novel attempt to utilize ML techniques to develop theory from data. The rationale for complementing open-ended studies with ML techniques is grounded in their capacity to approximate complex functions and uncover non-obvious patterns within data. Therefore, it represents a methodological advancement and a way to validate theories on innovation and management by relying on induction rather than being confined to incomplete theoretical models that may lack relevant variables that cannot be incorporated due to the limitations of the models, or the analytical method used. We contribute to previous studies where ML techniques have proven valuable in addressing various aspects of supply chain management (Chuang *et al.*, 2021; Mandl and Minner, 2023), environmental innovation (Chang *et al.*, 2021) and the enhancement of end-customer experiences (Ilk and Fan, 2022; Jagabathula and Rusmevichientong, 2019).

Furthermore, our results are highly relevant to innovation theories as they strongly support research in open innovation (Chesbrough, 2006; Vendrell-Herrero *et al.*, 2023). The factors proposed by this theory are crucial in determining success in both product and PSI. Moreover, the collaboration landscape is complex and contingent upon the type of innovation, competitive priorities and market orientation pursued. Therefore, it is essential to incorporate this complexity into future studies, as mere collaboration is often insufficient to achieve product and PSI. It is necessary to deploy specific combinations of collaborations based on the type and ultimate competitive priority established. Similarly, the presented results are relevant to studies on servitization and productization (Baines *et al.*, 2017; Leoni, 2019). The conducted analyses conclude that the factors and, above all, the way they are combined and deployed, do not fully coincide in the case of product innovation and PSI from manufacturers or service firms. This finding reinforces the body of knowledge on PSI as a specific type of innovation with its own characteristics that must be considered in research and management (Opazo-Basález *et al.*, 2022).

The implications of these results hold particular relevance for practitioners. They suggest that managers should consider the intricate relationships between collaboration types and competitive variables when implementing innovation initiatives. The decision regarding whom to collaborate with and their significance for innovation success depends on both the ultimate competitive priority and the established market orientation. Moreover, the findings delineate essential pathways for firms to achieve product innovation and PSI in alignment with their external and internal operational contexts. ML techniques prove instrumental in understanding and pursuing specific objectives, such as identifying potential breakthroughs in solar cell technology (Li *et al.*, 2022), optimizing production processes in the chemical industry (Arboretti *et al.*, 2022) and implementing improvements in the urban administrative public sector (Luo *et al.*, 2023). Hence, ML techniques serve as valuable decision-making tools for establishing and attaining strategic as well as operational objectives.

In terms of potential future research directions, additional logarithmic approaches may prove advantageous for analyzing innovation outcomes. For instance, Logistic Model Trees (LMT) integrate decision trees with logistic regression models. This method follows a two-step procedure where decision tree nodes are established using specific splitting criteria, and logistic regression models are subsequently applied at the terminal nodes (Landwehr *et al.*, 2005). The logistic regression models at the terminal nodes offer valuable insights into the relationship between attribute values and class probabilities, making them suitable for the analysis of both categorical and numerical attributes. Additionally, novel ML approaches for hypothesis testing (Chou *et al.*, 2022) involve fitting training samples with the Random Forest method. They identify variables of interest, evaluate their predictive significance, employ H-statistics for interaction quantification and utilize Jackknifed-based confidence intervals to determine significance. These recent ML advancements significantly broaden the potential application of these methodologies for testing and building theory.

Notes

1. Another methodological approach for identifying the necessary conditions to achieve an innovation outcome is fsQCA (Xie and Wang, 2020). A key distinction between a decision tree and an fsQCA analysis is that, in the former, achieving a specific outcome requires following the sequence of decisions made at each internal node of the tree. Conversely, the latter identifies the set of necessary and sufficient conditions for reaching the outcome without proposing a specific path to attain them.

References

- Alegre-Vidal, J., Lapedra-Alcamí, R. and Chiva-Gómez, R. (2004), "Linking operations strategy and product innovation: an empirical study of Spanish ceramic tile producers", *Research Policy*, Vol. 33 No. 5, pp. 829-839, doi: [10.1016/j.respol.2004.01.003](https://doi.org/10.1016/j.respol.2004.01.003).
- Amara, N. and Landry, R. (2005), "Sources of information as determinants of novelty of innovation in manufacturing firms: evidence from the 1999 statistics Canada innovation survey", *Technovation*, Vol. 25 No. 3, pp. 245-259, doi: [10.1016/S0166-4972\(03\)00113-5](https://doi.org/10.1016/S0166-4972(03)00113-5).
- Appiah-Adu, K. and Ranchhod, A. (1998), "Market orientation and performance in the biotechnology industry: an exploratory empirical analysis", *Technology Analysis and Strategic Management*, Vol. 10 No. 2, pp. 197-210, doi: [10.1080/09537329808524311](https://doi.org/10.1080/09537329808524311).
- Aquilante, T. and Vendrell-Herrero, F. (2021), "Bundling and exporting: evidence from German SMEs", *Journal of Business Research*, Vol. 132, pp. 32-44, doi: [10.1016/j.jbusres.2021.03.059](https://doi.org/10.1016/j.jbusres.2021.03.059).
- Arboretti, R., Ceccato, R., Pegoraro, L., Salmaso, L., Housmekerides, C., Spadoni, L., Pierangelo, E., Quaggia, S., Tveit, C. and Vianello, S. (2022), "Machine learning and design of experiments with an application to product innovation in the chemical industry", *Journal of Applied Statistics*, Vol. 49 No. 10, pp. 2674-2699, doi: [10.1080/02664763.2021.1907840](https://doi.org/10.1080/02664763.2021.1907840).

- Baines, T., Lightfoot, H., Peppard, J., Johnson, M., Tiwari, A., Shehab, E. and Swink, M. (2009), "Towards an operations strategy for product-centric servitization", *International Journal of Operations and Production Management*, Vol. 29 No. 5, pp. 494-519, doi: [10.1108/01443570910953603](https://doi.org/10.1108/01443570910953603).
- Baines, T., Ziaee Bigdeli, A., Bustinza, O.F., Shi, V.G., Baldwin, J. and Ridgway, K. (2017), "Servitization: revisiting the state-of-the-art and research priorities", *International Journal of Operations and Production Management*, Vol. 37 No. 2, pp. 256-278, doi: [10.1108/ijopm-06-2015-0312](https://doi.org/10.1108/ijopm-06-2015-0312).
- Bustinza, O.F., Parry, G.C. and Vendrell-Herrero, F. (2013), "Supply and demand chain management: the effect of adding services to product offerings", *Supply Chain Management*, Vol. 18, p. 6, doi: [10.1108/SCM-05-2013-0149](https://doi.org/10.1108/SCM-05-2013-0149).
- Bustinza, O.F., Bigdeli, A.Z., Baines, T. and Elliot, C. (2015), "Servitization and competitive advantage: the importance of organizational structure and value chain position", *Research Technology Management*, Vol. 58 No. 5, pp. 53-60, doi: [10.5437/08956308X5805354](https://doi.org/10.5437/08956308X5805354).
- Bustinza, O.F., Gomes, E., Vendrell-Herrero, F. and Baines, T. (2019), "Product-service innovation and performance: the role of collaborative partnerships and R&D intensity", *R&D Management*, Vol. 49 No. 1, pp. 33-45, doi: [10.1111/radm.12269](https://doi.org/10.1111/radm.12269).
- Bustinza, O.F., Opazo-Basáez, M. and Tarba, S. (2022), "Exploring the interplay between Smart Manufacturing and KIBS firms in configuring product-service innovation performance", *Technovation*, Vol. 118, 102258, doi: [10.1016/j.technovation.2021.102258](https://doi.org/10.1016/j.technovation.2021.102258).
- Cenamor, J., Sjödin, D.R. and Parida, V. (2017), "Adopting a platform approach in servitization: leveraging the value of digitalization", *International Journal of Production Economics*, Vol. 192, pp. 54-65, doi: [10.1016/j.ijpe.2016.12.033](https://doi.org/10.1016/j.ijpe.2016.12.033).
- Cerniglia, J.A. and Fabozzi, F.J. (2020), "Selecting computational models for asset management: financial econometrics versus machine learning—is there a conflict?", *The Journal of Portfolio Management*, Vol. 47 No. 1, pp. 107-118, doi: [10.3905/jpm.2020.1.184](https://doi.org/10.3905/jpm.2020.1.184).
- Chang, X., Huang, Y., Li, M., Bo, X. and Kumar, S. (2021), "Efficient detection of environmental violators: a big data approach", *Production and Operations Management*, Vol. 30 No. 5, pp. 1246-1270, doi: [10.1111/poms.13272](https://doi.org/10.1111/poms.13272).
- Chesbrough, H.W. (2006), *Open Innovation: the New Imperative for Creating and Profiting from Technology*, Harvard Business Press, New Delhi.
- Chou, Y.-C., Chuang, H.H.C., Chou, P. and Oliva, R. (2022), "Supervised machine learning for theory building and testing: opportunities in operations management", *Journal of Operations Management*, Vol. 69 No. 4, pp. 643-675, doi: [10.1002/joom.1228](https://doi.org/10.1002/joom.1228).
- Chuang, H.H.C., Chou, Y.-C. and Oliva, R. (2021), "Cross-item learning for volatile demand forecasting: an intervention with predictive analytics", *Journal of Operations Management*, Vol. 67 No. 7, pp. 828-852, doi: [10.1002/joom.1152](https://doi.org/10.1002/joom.1152).
- Cohen, W.W. (1995), "Fast effective rule induction", *Presented at the Machine Learning: Proceedings of the Twelfth International Conference*, 1995.
- Cooper, R.G. and Kleinschmidt, E.J. (1987), "Success factors in product innovation", *Industrial Marketing Management*, Vol. 16 No. 3, pp. 215-223, doi: [10.1016/0019-8501\(87\)90029-0](https://doi.org/10.1016/0019-8501(87)90029-0).
- Cornelissen, J.P. (2017), "Preserving theoretical divergence in management research: why the explanatory potential of qualitative research should be harnessed rather than suppressed", *Journal of Management Studies*, Vol. 54 No. 3, pp. 368-383, doi: [10.1111/joms.12210](https://doi.org/10.1111/joms.12210).
- Dahlander, L. and Gann, D.M. (2010), "How open is innovation?", *Research Policy*, Vol. 39 No. 6, pp. 699-709, doi: [10.1016/j.respol.2010.01.013](https://doi.org/10.1016/j.respol.2010.01.013).
- Dangelico, R.M. (2016), "Green product innovation: where we are and where we are going", *Business Strategy and the Environment*, Vol. 25 No. 8, pp. 560-576, doi: [10.1002/bse.1886](https://doi.org/10.1002/bse.1886).
- Dasgupta, A. and Nath, A. (2016), "Classification of machine learning algorithms", *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, Vol. 3 No. 3, pp. 6-11.

-
- De Faria, P., Lima, F. and Santos, R. (2010), "Cooperation in innovation activities: the importance of partners", *Research Policy*, Vol. 39 No. 8, pp. 1082-1092, doi: [10.1016/j.respol.2010.05.003](https://doi.org/10.1016/j.respol.2010.05.003).
- Enkel, E., Bogers, M. and Chesbrough, H. (2020), "Exploring open innovation in the digital age: a maturity model and future research directions", *R&D Management*, Vol. 50 No. 1, pp. 161-168, doi: [10.1111/radm.12397](https://doi.org/10.1111/radm.12397).
- Eom, T., Woo, C. and Chun, D. (2022), "Predicting an ICT business process innovation as a digital transformation with machine learning techniques", *Technology Analysis and Strategic Management*, pp. 1-13.
- Fernández-Delgado, M., Cernadas, E., Barro, S. and Amorim, D. (2014), "Do we need hundreds of classifiers to solve real world classification problems?", *The Journal of Machine Learning Research*, Vol. 15 No. 1, pp. 3133-3181.
- Gault, F. (2018), "Defining and measuring innovation in all sectors of the economy", *Research Policy*, Vol. 47 No. 3, pp. 617-622, doi: [10.1016/j.respol.2018.01.007](https://doi.org/10.1016/j.respol.2018.01.007).
- Gertler, M.S. and Levitte, Y.M. (2005), "Local nodes in global networks: the geography of knowledge flows in biotechnology innovation", *Industry and Innovation*, Vol. 12 No. 4, pp. 487-507, doi: [10.1080/13662710500361981](https://doi.org/10.1080/13662710500361981).
- Gunday, G., Ulusoy, G., Kilic, K. and Alpkan, L. (2011), "Effects of innovation types on firm performance", *International Journal of Production Economics*, Vol. 133 No. 2, pp. 662-676, doi: [10.1016/j.ijpe.2011.05.014](https://doi.org/10.1016/j.ijpe.2011.05.014).
- Han, J., Kamber, M. and Pei, J. (2012), "Outlier detection", *Data Mining: Concepts and Techniques*, pp. 543-584.
- Hastie, T., Tibshirani, R., Friedman, J.H. and Friedman, J.H. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Vol. 2, Springer, New York.
- Hayes, R.H. (1984), *Restoring Our Competitive Edge: Competing through Manufacturing*/Robert H. Hayes, Steven C. Wheelwright, Wiley, New York, Chichester.
- He, H. and Garcia, E.A. (2009), "Learning from imbalanced data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21 No. 9, pp. 1263-1284, doi: [10.1109/tkde.2008.239](https://doi.org/10.1109/tkde.2008.239).
- Hempel, C.G. (1966), *Philosophy of Natural Science*, Printice Hall, Englewood Cliffs.
- Huizingh, E.K.R.E. (2011), "Open innovation: state of the art and future perspectives", *Technovation*, Vol. 31 No. 1, pp. 2-9, doi: [10.1016/j.technovation.2010.10.002](https://doi.org/10.1016/j.technovation.2010.10.002).
- Ilk, N. and Fan, S. (2022), "Combining textual cues with social clues: utilizing social features to improve sentiment analysis in social media", *Decision Sciences*, Vol. 53 No. 2, pp. 320-347, doi: [10.1111/deci.12490](https://doi.org/10.1111/deci.12490).
- Jagabathula, S. and Rusmevichientong, P. (2019), "The limit of rationality in choice modeling: formulation, computation, and implications", *Management Science*, Vol. 65 No. 5, pp. 2196-2215.
- Johnston, R., Jones, K. and Manley, D. (2018), "Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour", *Quality and Quantity*, Vol. 52 No. 4, pp. 1957-1976, doi: [10.1007/s11135-017-0584-6](https://doi.org/10.1007/s11135-017-0584-6).
- Kalnins, A. (2018), "Multicollinearity: how common factors cause Type 1 errors in multivariate regression", *Strategic Management Journal*, Vol. 39 No. 8, pp. 2362-2385, doi: [10.1002/smj.2783](https://doi.org/10.1002/smj.2783).
- Kalnins, A. (2022), "When does multicollinearity bias coefficients and cause type 1 errors? A reconciliation of Lindner, Puck, and Verbeke (2020) with Kalnins (2018)", *Journal of International Business Studies*, Vol. 53 No. 7, pp. 1-13, doi: [10.1057/s41267-022-00531-9](https://doi.org/10.1057/s41267-022-00531-9).
- Kolagar, M., Parida, V. and Sjödin, D. (2022), "Ecosystem transformation for digital servitization: a systematic review, integrative framework, and future research agenda", *Journal of Business Research*, Vol. 146, pp. 176-200, doi: [10.1016/j.jbusres.2022.03.067](https://doi.org/10.1016/j.jbusres.2022.03.067).
- Kowalkowski, C., Gebauer, H., Kamp, B. and Parry, G. (2017a), "Servitization and deservitization: overview, concepts, and definitions", *Industrial Marketing Management*, Vol. 60, pp. 4-10, doi: [10.1016/j.indmarman.2016.12.007](https://doi.org/10.1016/j.indmarman.2016.12.007).

-
- Kowalkowski, C., Gebauer, H. and Oliva, R. (2017b), "Service growth in product firms: past, present, and future", *Industrial Marketing Management*, Vol. 60, pp. 82-88, doi: [10.1016/j.indmarman.2016.10.015](https://doi.org/10.1016/j.indmarman.2016.10.015).
- Lafuente, E., Vaillant, Y. and Vendrell-Herrero, F. (2023), "Editorial: product-service innovation Systems—opening-up servitization-based innovation to manufacturing industry", *Technovation*, Vol. 120, 102665, doi: [10.1016/j.technovation.2022.102665](https://doi.org/10.1016/j.technovation.2022.102665).
- Landwehr, N., Hall, M. and Frank, E. (2005), "Logistic model trees", *Machine Learning*, Vol. 59 No. 1, pp. 161-205, doi: [10.1007/s10994-005-0466-3](https://doi.org/10.1007/s10994-005-0466-3).
- Laursen, K. and Salter, A. (2006), "Open for innovation: the role of openness in explaining innovation performance among UK manufacturing firms", *Strategic Management Journal*, Vol. 27 No. 2, pp. 131-150, doi: [10.1002/smj.507](https://doi.org/10.1002/smj.507).
- Leong, G.K., Snyder, D.L. and Ward, P.T. (1990), "Research in the process and content of manufacturing strategy", *Omega*, Vol. 18 No. 2, pp. 109-122, doi: [10.1016/0305-0483\(90\)90058-h](https://doi.org/10.1016/0305-0483(90)90058-h).
- Leoni, L. (2019), "Productisation as the reverse side of the servitisation strategy", *International Journal of Business Environment*, Vol. 10 No. 3, pp. 247-269, doi: [10.1504/ijbe.2019.097981](https://doi.org/10.1504/ijbe.2019.097981).
- Li, Y.-R. (2009), "The technological roadmap of Cisco's business ecosystem", *Technovation*, Vol. 29 No. 5, pp. 379-386, doi: [10.1016/j.technovation.2009.01.007](https://doi.org/10.1016/j.technovation.2009.01.007).
- Li, X., Wen, Y., Jiang, J., Daim, T. and Huang, L. (2022), "Identifying potential breakthrough research: a machine learning method using scientific papers and Twitter data", *Technological Forecasting and Social Change*, Vol. 184, 122042, doi: [10.1016/j.techfore.2022.122042](https://doi.org/10.1016/j.techfore.2022.122042).
- Liebrechts, W.J., van den Heuvel, W.-J. and van den Born, A. (2023), *Data Science for Entrepreneurship: Principles and Methods for Data Engineering, Analytics, Entrepreneurship, and the Society*, Springer, New York.
- Lim, K.Y.H., Zheng, P. and Chen, C.-H. (2020), "A state-of-the-art survey of Digital Twin: techniques, engineering product lifecycle management and business innovation perspectives", *Journal of Intelligent Manufacturing*, Vol. 31 No. 6, pp. 1313-1337, doi: [10.1007/s10845-019-01512-w](https://doi.org/10.1007/s10845-019-01512-w).
- Lindner, T., Puck, J. and Verbeke, A. (2022), "Beyond addressing multicollinearity: robust quantitative analysis and machine learning in international business research", *Journal of International Business Studies*, Vol. 53 No. 7, pp. 1-8, doi: [10.1057/s41267-022-00549-z](https://doi.org/10.1057/s41267-022-00549-z).
- Liu, Y., Xing, Y., Vendrell-Herrero, F. and Bustinza, O.F. (2023), "Setting contextual conditions to resolve grand challenges through responsible innovation: a comparative patent analysis in the circular economy", *Journal of Product Innovation Management*, Vol. n/a No. n/a, doi: [10.1111/jpim.12659](https://doi.org/10.1111/jpim.12659).
- Luo, J., Wang, Y. and Li, G. (2023), "The innovation effect of administrative hierarchy on intercity connection: the machine learning of twin cities", *Journal of Innovation and Knowledge*, Vol. 8 No. 1, 100293, doi: [10.1016/j.jik.2022.100293](https://doi.org/10.1016/j.jik.2022.100293).
- Mandl, C. and Minner, S. (2023), "Data-driven optimization for commodity procurement under price uncertainty", *Manufacturing and Service Operations Management*, Vol. 25 No. 2, pp. 371-390, doi: [10.1287/msom.2020.0890](https://doi.org/10.1287/msom.2020.0890).
- Mariani, M.M., Machado, I., Magrelli, V. and Dwivedi, Y.K. (2023), "Artificial intelligence in innovation research: a systematic review, conceptual framework, and future research directions", *Technovation*, Vol. 122, 102623, doi: [10.1016/j.technovation.2022.102623](https://doi.org/10.1016/j.technovation.2022.102623).
- Marr, B. (2019), *Artificial Intelligence in Practice: How 50 Successful Companies Used AI and Machine Learning to Solve Problems*, John Wiley & Sons, New Jersey.
- Mendoza-Silva, A. (2020), "Innovation capability: a systematic literature review", *European Journal of Innovation Management*, Vol. 24 No. 3, pp. 707-734, doi: [10.1108/EJIM-09-2019-0263](https://doi.org/10.1108/EJIM-09-2019-0263).
- Mention, A.L. (2011), "Co-operation and co-opetition as open innovation practices in the service sector: which influence on innovation novelty?", *Technovation*, Vol. 31 No. 1, pp. 44-53, doi: [10.1016/j.technovation.2010.08.002](https://doi.org/10.1016/j.technovation.2010.08.002).

-
- Mikalef, P., Islam, N., Parida, V., Singh, H. and Altwaijry, N. (2023), "Artificial intelligence (AI) competencies for organizational performance: a B2B marketing capabilities perspective", *Journal of Business Research*, Vol. 164, 113998, doi: [10.1016/j.jbusres.2023.113998](https://doi.org/10.1016/j.jbusres.2023.113998).
- Nafizah, U.Y., Roper, S. and Mole, K. (2023), "Estimating the innovation benefits of first-mover and second-mover strategies when micro-businesses adopt artificial intelligence and machine learning", *Small Business Economics*, Vol. 62, pp. 1-24, doi: [10.1007/s11187-023-00779-x](https://doi.org/10.1007/s11187-023-00779-x).
- OCDE (2005), *Proposed Guidelines for Collecting and Interpreting Technological Innovation Data*, OCDE: Statistical Office of the European Communities, Paris.
- Opazo-Basáez, M., Vendrell-Herrero, F. and Bustinza, O.F. (2022), "Digital service innovation: a paradigm shift in technological innovation", *Journal of Service Management*, Vol. 33 No. 1, pp. 97-120, doi: [10.1108/JOSM-11-2020-0427](https://doi.org/10.1108/JOSM-11-2020-0427).
- Pagell, M., Flynn, B. and Fugate, B. (2019), "Call for papers for the 2020 emerging discourse incubator emerging approaches for developing SCM theory", *Journal of Supply Chain Management*, Vol. 55 No. 4, pp. 129-131, doi: [10.1111/jscm.12209](https://doi.org/10.1111/jscm.12209).
- Phusavat, K. and Kanchana, R. (2007), "Competitive priorities of manufacturing firms in Thailand", *Industrial Management and Data Systems*, Vol. 107 No. 7, pp. 979-996, doi: [10.1108/02635570710816702](https://doi.org/10.1108/02635570710816702).
- Prasad, A. and Prasad, P. (2002), "The coming of age of interpretive organizational research", *Organizational Research Methods*, Vol. 5 No. 1, pp. 4-11, doi: [10.1177/1094428102051002](https://doi.org/10.1177/1094428102051002).
- Quinlan, J.R. (1993), *Program for Machine Learning, CA. 5*, Morgan Kaufmann, Burlington.
- Rabetino, R., Harmsen, W., Kohtamäki, M. and Sihvonen, J. (2018), "Structuring servitization-related research", *International Journal of Operations and Production Management*, Vol. 38 No. 2, pp. 350-371, doi: [10.1108/IJOPM-03-2017-0175](https://doi.org/10.1108/IJOPM-03-2017-0175).
- Rabetino, R., Kohtamäki, M., Kowalkowski, C., Baines, T.S. and Sousa, R. (2021), "Guest editorial: servitization 2.0: evaluating and advancing servitization-related research through novel conceptual and methodological perspectives", *International Journal of Operations and Production Management*, Vol. 41 No. 5, pp. 437-464, doi: [10.1108/IJOPM-05-2021-840](https://doi.org/10.1108/IJOPM-05-2021-840).
- Rao, A.S. and Verweij, G. (2017), "PwC's global artificial intelligence study: exploiting the AI revolution", available at: [Ttps://Www.Pwc.Com/Gx/En/Issues/Analytics/Assets/Pwc-Ai-Analysis-Sizing-the-Prize-Report.Pdf](https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf)
- Reeves, C.A. and Bednar, D.A. (1994), "Defining quality: alternatives and implications", *The Academy of Management Review*, Vol. 19 No. 3, 419, doi: [10.2307/258934](https://doi.org/10.2307/258934).
- Reim, W., Sjödin, D.R. and Parida, V. (2019), "Servitization of global service network actors – a contingency framework for matching challenges and strategies in service transition", *Journal of Business Research*, Vol. 104, pp. 461-471, doi: [10.1016/j.jbusres.2019.01.032](https://doi.org/10.1016/j.jbusres.2019.01.032).
- Rodríguez-Espíndola, O., Chowdhury, S., Beltagui, A. and Albores, P. (2020), "The potential of emergent disruptive technologies for humanitarian supply chains: the integration of blockchain, Artificial Intelligence and 3D printing", *International Journal of Production Research*, Vol. 58 No. 15, pp. 4610-4630, doi: [10.1080/00207543.2020.1761565](https://doi.org/10.1080/00207543.2020.1761565).
- Schmenner, R.W. and Swink, M.L. (1998), "On theory in operations management", *Journal of Operations Management*, Vol. 17 No. 1, pp. 97-113, doi: [10.1016/s0272-6963\(98\)00028-x](https://doi.org/10.1016/s0272-6963(98)00028-x).
- Sjödin, D., Parida, V., Kohtamäki, M. and Wincent, J. (2020), "An agile co-creation process for digital servitization: a micro-service innovation approach", *Journal of Business Research*, Vol. 112, pp. 478-491, doi: [10.1016/j.jbusres.2020.01.009](https://doi.org/10.1016/j.jbusres.2020.01.009).
- Skinner, W. (1969), "Manufacturing: missing link in corporate strategy", *Harvard Business Review*, Vol. 47 No. 3, pp. 136-145.
- Slater, S.F., Mohr, J.J. and Sengupta, S. (2014), "Radical product innovation capability: literature review, synthesis, and illustrative research propositions", *Journal of Product Innovation Management*, Vol. 31 No. 3, pp. 552-566, doi: [10.1111/jpim.12113](https://doi.org/10.1111/jpim.12113).

-
- Snyder, H., Witell, L., Gustafsson, A., Fombelle, P. and Kristensson, P. (2016), "Identifying categories of service innovation: a review and synthesis of the literature", *Journal of Business Research*, Vol. 69 No. 7, pp. 2401-2408, doi: [10.1016/j.jbusres.2016.01.009](https://doi.org/10.1016/j.jbusres.2016.01.009).
- Stockdale, R. and Standing, C. (2006), "An interpretive approach to evaluating information systems: a content, context, process framework", *European Journal of Operational Research*, Vol. 173 No. 3, pp. 1090-1102, doi: [10.1016/j.ejor.2005.07.006](https://doi.org/10.1016/j.ejor.2005.07.006).
- Szász, L. and Seer, L. (2018), "Towards an operations strategy model of servitization: the role of sustainability pressure", *Operations Management Research*, Vol. 11 No. 1, pp. 51-66, doi: [10.1007/s12063-018-0132-0](https://doi.org/10.1007/s12063-018-0132-0).
- Tajeddini, K., Trueman, M. and Larsen, G. (2006), "Examining the effect of market orientation on innovativeness", *Journal of Marketing Management*, Vol. 22 Nos 5-6, pp. 529-551, doi: [10.1362/026725706777978640](https://doi.org/10.1362/026725706777978640).
- Thao, H.T. and Xie, X. (2023), "Fostering green innovation performance through open innovation strategies: do green subsidies work?", *Environment, Development and Sustainability*, doi: [10.1007/s10668-023-03409-4](https://doi.org/10.1007/s10668-023-03409-4), available at: <https://link.springer.com/article/10.1007/s10668-023-03409-4#citeas>
- Tonidandel, S., King, E.B. and Cortina, J.M. (2018), "Big data methods: leveraging modern data analytic techniques to build organizational science", *Organizational Research Methods*, Vol. 21 No. 3, pp. 525-547, doi: [10.1177/1094428116677299](https://doi.org/10.1177/1094428116677299).
- Vaccaro, A., Parente, R. and Veloso, F.M. (2010), "Knowledge management tools, inter-organizational relationships, innovation and firm performance", *Technological Forecasting and Social Change*, Vol. 77 No. 7, pp. 1076-1089, doi: [10.1016/j.techfore.2010.02.006](https://doi.org/10.1016/j.techfore.2010.02.006).
- Vendrell-Herrero, F., Bustinza, O.F., Opazo-Basaez, M. and Gomes, E. (2023), "Treble innovation firms: antecedents, outcomes, and enhancing factors", *International Journal of Production Economics*, Vol. 255, 108682, doi: [10.1016/j.ijpe.2022.108682](https://doi.org/10.1016/j.ijpe.2022.108682).
- von Hippel, E. (2007), "The sources of innovation", in Boersch, C. and Elschen, R. (Eds), *Das Summa Summarum des Management: Die 25 wichtigsten Werke für Strategie, Führung und Veränderung*, Gabler, Wiesbaden, pp. 111-120, doi: [10.1007/978-3-8349-9320-5_10](https://doi.org/10.1007/978-3-8349-9320-5_10).
- Ward, P.T., McCreery, J.K., Ritzman, L.P. and Sharma, D. (1998), "Competitive priorities in operations management", *Decision Sciences*, Vol. 29 No. 4, pp. 1035-1046, doi: [10.1111/j.1540-5915.1998.tb00886.x](https://doi.org/10.1111/j.1540-5915.1998.tb00886.x).
- Wheelwright, S.C. and Hayes, R.H. (1985), "Competing through manufacturing", *Harvard Business Review*, Vol. 63 No. 1, pp. 99-109.
- Witten, I.H. and Frank, E. (2002), *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Elsevier, New York.
- Wolstenholme, E.F. (1999), "Qualitative vs quantitative modelling: the evolving balance", *Journal of the Operational Research Society*, Vol. 50 No. 4, pp. 422-428, doi: [10.2307/3010462](https://doi.org/10.2307/3010462).
- Xie, X. and Wang, H. (2020), "How can open innovation ecosystem modes push product innovation forward? An fsQCA analysis", *Journal of Business Research*, Vol. 108, pp. 29-41, doi: [10.1016/j.jbusres.2019.10.011](https://doi.org/10.1016/j.jbusres.2019.10.011).
- Yeniaras, V., Di Benedetto, A. and Dayan, M. (2021), "Effects of relational ties paradox on financial and non-financial consequences of servitization: roles of organizational flexibility and improvisation", *Industrial Marketing Management*, Vol. 99, pp. 54-68, doi: [10.1016/j.indmarman.2021.09.006](https://doi.org/10.1016/j.indmarman.2021.09.006).
- Yunis, M., Tarhini, A. and Kassar, A. (2018), "The role of ICT and innovation in enhancing organizational performance: the catalysing effect of corporate entrepreneurship", *Journal of Business Research*, Vol. 88, pp. 344-356, doi: [10.1016/j.jbusres.2017.12.030](https://doi.org/10.1016/j.jbusres.2017.12.030).
- Zhou, C. and Li, J. (2008), "Product innovation in emerging market-based international joint ventures: an organizational ecology perspective", *Journal of International Business Studies*, Vol. 39 No. 7, pp. 1114-1132, doi: [10.1057/jibs.2008.51](https://doi.org/10.1057/jibs.2008.51).

Ziaee Bigdeli, A., Bustinza, O.F., Vendrell-Herrero, F. and Baines, T. (2017), "Network positioning and risk perception in servitization: evidence from the UK road transport industry", *International Journal of Production Research*, Vol. 56 No. 6, pp. 2169-2183, doi: [10.1080/00207543.2017.1341063](https://doi.org/10.1080/00207543.2017.1341063).

Machine
learning, product
innovation and
PSI

Corresponding author

Oscar F. Bustinza can be contacted at: oscarfb@ugr.es

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com