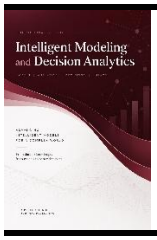
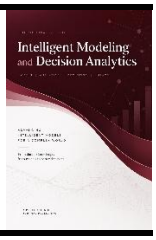


Intelligent Modeling and Decision Analytics

Journal homepage: www.imda.journal-publishing.org
ISSN: 2838-1807



Profiling and Prioritizing Export Success in a Technopark with LLM-Based Feature Extraction, Explainable Machine Learning, and Fuzzy MCDM

Abdullah Kürşat Merter[1,*], Murat Çemberci[2], Fahim Jaman[3]

[1] Department of Business Administration, Faculty of Business Administration, Gebze Technical University, Kocaeli, Türkiye
[2] Department of Business Administration, Faculty of Economics and Administrative Sciences, Yıldız Technical University, İstanbul, Türkiye
[3] Department of Civil, Environmental, and Transportation Engineering, Morgan State University, 1700 E. Cold Spring Lane, Baltimore, MD 21251, United States of America

ARTICLE INFO

Article history:
Received 13 June 2026
Received in revised form 3 August 2026
Accepted 15 September 2026
Available online 20 September 2026

Keywords:
Technology parks; Export prediction; Natural language processing; Explainable artificial intelligence; Resource allocation

ABSTRACT

Science and technology parks function as primary instruments of regional innovation policy designed to accelerate economic development. However, the evaluation literature predominantly estimates average historical park effects rather than identifying specific tenant firms requiring targeted public support, leaving park administrators without objective tools for allocating scarce resources. The purpose of this research is to address this resource allocation problem by developing a data-driven decision support framework that systematically profiles and prioritizes tenant firms based on their empirical export potential. The methodology applies a large language model to extract technological complexity metrics from 557 unstructured administrative project narratives. It then utilizes explainable machine learning classifiers to predict firm-level export success and translates the resulting algorithmic feature attributions into objective criterion weights for a fuzzy multi-criteria prioritization model. The principal results indicate that park-specific institutional tenure and realized research revenue are the dominant predictors of commercial export success. The regularized predictive model achieves strong out-of-sample classification performance, while the extracted textual innovation scores provide additional predictive information beyond traditional headcount metrics. The results demonstrate that integrating natural language processing with explainable artificial intelligence provides a data-driven alternative to subjective expert judgments in resource allocation. The resulting prioritization framework provides policymakers with an auditable and reproducible tool for effectively targeting internationalization grants and optimizing the distribution of public innovation incentives.

1. Introduction

Science and technology parks function as primary instruments of public innovation policy, financed to concentrate research and development activity and raise the technological intensity of regional economies. Despite two decades of evaluation research, the academic literature lacks a consensus regarding whether these zones consistently achieve their intended economic outcomes.

* *Corresponding author.*
E-mail address: akmerter@gtu.edu.tr

Lindelöf and Löfsten [1], comparing new technology-based firms located on and off Swedish science parks, found measurable differences in R&D intensity and financing conditions but observed only a slight overall advantage for on-park firms. Squicciarini [2, 3], using duration models of patenting activity, reported comparatively better innovative performance among park tenants while documenting non-trivial matching effects between specific firm profiles and park characteristics. Albahari *et al.,* [4] concluded that the empirical evidence remains contrasting and that the likelihood of detecting positive park effects rises considerably with larger sample sizes.

This inconclusive record reflects structural methodological limitations rather than mere measurement noise. The evaluation literature is built almost entirely on average-effect comparisons between on-park and off-park firms. Albahari *et al.,* [4] argued that average-effect designs structurally ignore heterogeneous outcomes, and that some firms benefit significantly more from an on-park location than others. Consequently, a critical resource allocation challenge remains unaddressed. Park managers and funding agencies operate under binding budget constraints and must distribute scarce financial support across highly heterogeneous firms inside a single park. While existing macro-level research evaluates whether parks function effectively on average, it provides limited operational guidance regarding which specific tenant firms possess the organizational capabilities to best utilize the next unit of public support.

Türkiye offers a particularly relevant institutional setting for examining this intra-park resource allocation problem. The Technology Development Zones Law No. 4691, enacted in 2001, grants tenant firms a comprehensive incentive regime including corporate income tax exemptions on R&D earnings, payroll income tax relief, and value added tax exemptions on software deliveries. By the end of February 2026, the country hosted 95 operational technology development zones accommodating 12,739 firms and 128,274 personnel, with cumulative exports reaching USD 16.2 billion. Computer programming accounts for 56.54% of all zone firms, making the national system predominantly a software-driven economy [5]. The substantial scale of public financial commitment necessitates precise policy targeting. Empirical evidence from emerging economies confirms that support is not always allocated to areas where it generates absolute additionality. Taş and Erdil [6] estimated that Turkish R&D tax incentives raise beneficiary R&D intensity by approximately 7% with a sub-unity multiplier, indicating a partial crowding out of the firms' own internal funds. Similarly, Teng *et al.,* [7] highlighted that government R&D funding in the Zhangjiang zone can crowd out the innovation performance of firms that already invest heavily in their own R&D capabilities. These observations suggest that the potential misallocation of public incentives constitutes a persistent structural challenge.

A parallel structural problem concerns the measurement of innovation performance in software-heavy environments. In software-dominated parks, intellectual property is typically protected through copyright and trade secrets rather than formal patents. Consequently, traditional patent counts exhibit little or no variance in these settings and are structurally uninformative as indicators of firm-level capabilities. Text-based innovation research offers a robust methodological alternative to overcome this limitation. Bellstam *et al.,* [8] showed that the textual analysis of analyst reports accurately reveals innovation activity for firms that traditional patent-based measures miss entirely. Kelly *et al.,* [9] established that unstructured textual descriptions effectively track technological progress over the long run. Building on the extraction of textual features to capture intangible technological knowledge, the present study concurrently adopts export success as a complementary and sector-consistent performance metric. While text mining captures latent innovation potential from project documents, exporting provides an observable, administratively recorded, and commercially meaningful indicator of realized performance for software firms whose primary output remains largely intangible.

To integrate these disparate data sources into actionable management insights, this paper proposes a hybrid decision support framework tailored for technopark administration, built upon the operational records of a major Turkish technopark. The methodology operates through three integrated phases. In Phase I, a large language model extracts structured technological complexity scores and domain labels from 557 Turkish-language R&D project narratives. In Phase II, explainable machine learning models classify 114 unique firms by their exporter status spanning the years 2021 to 2025, with SHAP values decomposing each algorithmic prediction into distinct feature-level contributions. In Phase III, the mean absolute SHAP values serve as objective criterion weights for a Fuzzy TOPSIS firm prioritization algorithm. This sequential framework systematically moves from the extraction of unstructured text to an interpretable classification, and ultimately yields a data-driven priority ordering. The resulting predictor variables are contextualized through the theoretical lenses of the resource-based view, organizational learning models, and absorptive capacity theory. The framework operates on the premise that firms gradually develop unique capabilities through sustained engagement with the park environment, and that these accumulated resources govern their ability to translate external public incentives into measurable commercial success in international markets.

The study delivers five distinct contributions to technology management literature. First, it provides the initial application of LLM-based feature extraction to technopark firm evaluation utilizing non-English project narratives. Second, it formulates the first firm-level prediction of export success inside a single technopark with complete explainability. This localized approach strongly complements macro-level exporter prediction studies, prominently including the work of Micocci and Rungi [10], who modeled exporter classification across a national sample of more than 57,000 French manufacturing firms. Third, it introduces a purely data-driven weighting mechanism for multi-criteria decision making in firm prioritization, replacing the subjective analytical hierarchy models proposed by Durak *et al.,* [11]. Fourth, it yields a prescriptive output capable of identifying latent high-potential firms. These entities exhibit strong predicted export capabilities that current metric-based resource allocation policies typically overlook. Fifth, it supplies concrete tenant-level evidence supporting a capability-accumulation perspective on science park effects. By demonstrating that localized tenure dominates as a predictor of commercial success, the findings directly link intra-park performance differences to the theoretical premises of the resource-based view and organizational learning models.

The remainder of the paper is organized as follows. Section 2 reviews the relevant theoretical literature and establishes the conceptual background. Section 3 details the administrative data sources and outlines the three-phase methodology in depth. Section 4 presents the empirical results of the extraction, prediction, and ranking procedures. Section 5 discusses the theoretical implications alongside policy recommendations. Finally, Section 6 provides concluding remarks.

Existing average-effect evaluations leave the problem of allocating scarce support across heterogeneous tenants within a single park unresolved. Addressing this gap is important for providing park administrators with an auditable, firm-level basis for targeted support. The empirical investigation is structured around three interconnected research objectives. The first objective investigates whether large language models can reliably convert unstructured R&D project narratives into quantitative technological complexity and domain features that are statistically usable in machine learning algorithms. The second objective examines the specific factors, spanning administrative incentive types, human resources, and the derived innovation scores, that best predict empirical export success among technopark firms. This stage concurrently identifies the exact directions of their marginal effects utilizing explainable artificial intelligence techniques. The third objective assesses the viability of replacing subjective expert-survey weights in a Fuzzy TOPSIS firm

prioritization framework with objective, data-driven importance values extracted directly from machine learning classifications.

2. Literature Review and Theoretical Framework

2.1 Theoretical Foundations of Capability Accumulation

The framework developed in this study relies on a predictive design, yet its predictor structure fundamentally rests on established organizational theory regarding how firms build the internal capabilities that successful exporting requires. The integration of machine learning features with established management theory ensures that the predictive algorithms capture meaningful economic mechanisms rather than spurious statistical correlations. The primary theoretical lens applied here is the resource-based view [12, 13]. This perspective holds that sustained performance differences originate in valuable, scarce, and hard-to-imitate resources accumulated over time rather than in current market positioning [14]. Within the specific context of a technology park, physical office space and basic infrastructure are uniformly available to all tenants and therefore cannot constitute a unique competitive advantage. Instead, the true resources driving competitive differentiation are localized and intangible assets. These assets include embedded network integration, regulatory familiarity, and specialized technological knowledge that require sustained engagement to develop. Because these resources are intangible and socially complex, they cannot be rapidly acquired on the open market, meaning that competitive advantage accrues primarily to organizations with prolonged exposure to the local innovation ecosystem.

Organizational learning research complements this theoretical view by demonstrating that such intangible resources are heavily path-dependent and built gradually through operational experience inside a specific institutional context [15, 16]. Firms do not instantly acquire innovation capabilities upon entering a technology development zone. Instead, they undergo a sequential learning process [17] where they adapt to the regulatory environment, integrate into local knowledge networks, and optimize their utilization of available support structures. This path dependency implies that time spent within the park environment serves as a fundamental proxy for the accumulation of location-specific capabilities. Therefore, general firm age and park-specific tenure represent two conceptually distinct constructs. General age captures standard organizational maturation, whereas park tenure explicitly captures localized institutional adaptation.

Absorptive capacity theory adds a critical additional dimension by proposing that firms convert external knowledge and institutional support into measurable commercial performance only to the extent of their prior related investment [18, 19]. Consequently, the exact same incentive or support service produces highly heterogeneous returns across different firms. External public funding or tax exemptions do not generate commercial innovation independently. A firm must possess sufficient internal technical expertise to recognize the value of the incentive, assimilate it into existing research processes, and apply it toward commercial ends [20]. A firm lacking a baseline of technical and administrative competency will struggle to assimilate the knowledge spillovers or commercialize the financial subsidies available within the park ecosystem.

These three theoretical lenses organize the empirical predictor set into distinct operational families. Park tenure and firm age measure accumulated, park-specific and general organizational capability in the sense of the resource-based view and organizational learning. R&D revenue and the intensity of incentive use proxy realized absorptive capacity, because both scale with the volume of prior R&D activity that allows firms to convert public support into commercial output [18]. The LLM-derived innovation score captures the technological distinctiveness of a firm project portfolio, an intangible knowledge resource that traditional administrative records fail to register [12]. The VAT exemption, by contrast, enters as a market-orientation marker rather than a capability measure,

since under Law No. 4691 it accrues on domestic software deliveries. This theoretical structure predicts that capability-accumulation variables should dominate the explanation of export success, that the text-derived distinctiveness measure should carry independent rather than redundant information, and that the VAT exemption should behave differently from the remaining incentive instruments.

2.2 Empirical Evaluations of Science and Technology Parks

The empirical evaluation of science and technology park performance has long relied on a single dominant methodological design, specifically the statistical comparison of tenant firms with matched off-park counterparts. The fundamental premise driving this literature suggests that physical proximity to academic institutions and shared specialized infrastructure should manifest as superior commercial and innovative outcomes through localized knowledge spillovers. However, the empirical realization of this theoretical premise remains persistently inconsistent across different geographical contexts. In a foundational comparative study of 273 Swedish new technology-based firms, Lindelöf and Löfsten [1] found higher R&D intensity and distinct financing patterns among on-park firms but observed no clear statistical advantage in overall sales or profitability. A companion survey conducted by the same authors reported that park tenants emphasized innovation ability, market orientation, and expected growth more strongly, yet these strategic orientations amounted to only a slight actual performance edge [21]. Moving beyond cross-sectional comparisons, duration models of Finnish firms patenting between 1970 and 2002 showed that park tenants sustained stronger innovative output over the firm life cycle, while concurrently documenting first-mover disadvantages and non-negligible matching effects between specific firm profiles and park characteristics [2, 3]. These early longitudinal studies established a fundamental methodological challenge. Observational comparisons struggle to separate the treatment effect of the park infrastructure from the selection effect of highly capable firms choosing to locate within these specialized zones.

Systematic reviews help explain this mixed empirical record by pointing to inherent methodological limitations in how park effects are conceptualized and estimated. Albahari *et al.,* [4] comprehensively reviewed 221 discrete evaluation studies and concluded that the likelihood of detecting positive park effects rises mechanically with sample size, and that analytical designs estimating only average treatment effects structurally ignore the underlying heterogeneity that drives the contradictory findings. Lecluyse *et al.,* [22] argue similarly that park contributions depend jointly on the specific configuration of park services and the individual absorptive characteristics of the tenant firms. Furthermore, Theeranattapong *et al.,* [23] catalogued more than twenty distinct performance measures utilized in on-park off-park studies and identified a persistent definitional problem regarding what statistically counts as park effectiveness across different regulatory regimes. Heterogeneity also operates profoundly at the macro institutional level. Dynamic panel models of 53 Chinese parks show that foreign direct investment and localized human capital explain persistent performance differences among parks, with convergence clubs forming along the urban city-tier system [24]. These macro-level variations suggest that treating science parks as uniform policy interventions fundamentally mischaracterizes the nature of geographic innovation clusters.

Evidence from emerging economies reinforces the conditional nature of spatial park effects and highlights the specific risks of resource misallocation under public subsidy regimes. Turkish incubator tenants have been shown to outperform off-incubator firms on economic outcomes but not on innovative output, with direct financial support mechanisms explaining much of the observed differential [25]. Focusing on specific incentive mechanisms, Taş and Erdil [6] demonstrated that Turkish R&D tax incentives raise recipient innovative intensity by about seven percent, yet the implied multiplier lies between zero and one, indicating a measurable partial crowding out of the

firms own internal funds. In the Zhangjiang demonstration zone, Teng *et al.,* [7] found that government R&D funding stimulated innovation on average but was shown to crowd out the innovation performance of firms with strong internal investment capabilities, revealing a critical mismatch between standardized park services and heterogeneous tenant needs. The Indian software technology park scheme similarly demonstrates that macro park policies can successfully build export-oriented firm populations without guaranteeing uniform firm-level efficiency gains [26]. This broader literature systematically evaluates whether parks work on average. It does not address how a park authority should strategically allocate scarce support across heterogeneous tenants inside a single park, which constitutes the primary operational challenge for technology park administrators.

2.3 Text as Data and Automated Feature Extraction

Converting unstructured text into quantitative variables for statistical analysis is now an established empirical strategy that circumvents the limitations of traditional administrative datasets. Gentzkow *et al.,* [27] codified the text-as-data paradigm, detailing how vast corpora of qualitative documents become structured arrays of computable features suitable for rigorous econometric modelling. Dell [28] demonstrated that advanced deep learning architectures can impute structured information from large unstructured text and image corpora, effectively extending the analytical paradigm from mere frequency representation to deep contextual interpretation. Korinek [29] catalogued the broader research uses of generative AI and documented substantial productivity gains from automating classification and extraction micro-tasks in economic and managerial research workflows.

Within innovation studies, text-derived measures carry strong external validity because they capture latent technological knowledge that patents and financial statements often obscure or miss entirely. This is particularly relevant in software industries where algorithms and codes are protected by trade secrets rather than formal patents, rendering traditional patent counts statistically useless. Kelly *et al.,* [9] built patent importance and breakthrough indicators from textual similarity metrics and successfully validated them against forward citations and subsequent market value. Bellstam *et al.,* [8] derived a quantitative firm-level innovation measure from textual analyst reports that remains highly informative for firms without patents or reported R&D expenditures. Arts *et al.,* [30] used natural language processing techniques to detect the emergence of new technologies directly within patent text and to trace their subsequent commercial and scientific impact across industrial sectors. Bowen *et al.,* [31] measured whether firm innovation sits in rapidly evolving technological areas based on textual descriptions and linked this measure to favorable startup exit modes. These diverse applications demonstrate that the narrative descriptions of technological projects contain rich and quantifiable signals about organizational market potential and innovative trajectory.

A more recent methodological stream investigates whether large language models can entirely replace human coding in the extraction of these complex qualitative signals. Benchmark studies show that ChatGPT exceeded the zero-shot annotation accuracy of human crowd workers by about 25 percentage points while achieving notably higher intercoder agreement than trained expert annotators [32]. Furthermore, LLM-generated perceptual ratings of brands and products agreed with human survey data at rates above 75 percent, supporting the reliable use of language model outputs as quantitative psychometric scores in social science research [33, 34]. Carlson and Burbano [35] set out comprehensive validation and reporting standards for LLM annotation in management research, emphasizing the necessity of reproducible prompts and fixed evaluation rubrics. Feuerriegel *et al.,* [36] document the extensive zero-shot and few-shot inference capabilities that make such advanced annotation feasible without requiring task-specific weight training or extensive model fine-tuning. For Turkish natural language processing tasks specifically, benchmark evidence shows that modern

language models handle zero-shot classification competently, with the best Turkish-specific model outperforming a multilingual baseline by nearly 20 percentage points [37]. Despite this documented progress, LLM-based feature extraction has not been systematically applied to technopark firm evaluation, nor has it been adapted to the short non-English project narratives that firms routinely produce for national innovation support systems. Addressing this methodological gap allows for the extraction of standardized innovation metrics in environments where traditional intellectual property indicators fail.

2.4 Explainable Machine Learning in Firm-Level Prediction

Supervised classification has a mature benchmarking tradition in firm outcome prediction, primarily focused on financial risk assessment and binary corporate events. Firm outcome datasets typically suffer from moderate to severe class imbalance, where the event of interest represents a minority of the sample. Lessmann *et al.,* [38] compared 41 distinct classifiers across eight corporate credit datasets and found that ensemble methods systematically outperform individual linear learners in predictive accuracy due to their capacity to model complex non-linear interaction surfaces. Dumitrescu *et al.,* [39] showed that augmenting standard logistic regression with nonlinear tree effects successfully recovers much of the ensemble advantage while preserving essential econometric interpretability, and Bayesian hyperparameter optimization further improves boosted decision trees in this specific predictive domain [40]. Applied directly to international trade economics, Micocci and Rungi [10] predicted exporter status for a massive sample of more than 57,000 French manufacturing firms with up to 90 percent accuracy. Based on these results, they proposed converting the algorithmically predicted probabilities into continuous scores designed specifically for targeted trade promotion. Ben Jabeur *et al.,* [41] combined XGBoost algorithms with variable-importance feature engineering for corporate bankruptcy prediction, achieving one-year-ahead AUC values well above traditional logistic and linear discriminant baselines.

Because these advanced algorithmic ensemble models are inherently opaque, explanation methods have become integral to applied prediction in public policy and resource allocation contexts where decisions must be explicitly justified. SHAP assigns each feature a Shapley-based contribution to every individual prediction, unifying earlier model-agnostic explanation approaches in a single mathematically consistent additive framework [42]. By grounding variable importance in cooperative game theory, this framework provides a mathematically sound method for isolating the marginal impact of specific firm characteristics on their probability of export success. This ensures that the distributed importance values sum exactly to the final prediction, satisfying the critical mathematical property of local accuracy. Two validation properties determine whether such algorithmic explanations are methodologically trustworthy in small administrative datasets. Vabalas *et al.,* [43] showed mathematically that naive cross-validation yields optimistically biased performance estimates in small samples, whereas nested resampling with tuning confined strictly to training folds produces unbiased estimates that generalize correctly to unseen observations. Chen *et al.,* [44] showed that SHAP and LIME explanations lose statistical stability as class imbalance grows, with material deterioration confined to extreme ratios rather than the moderate imbalances typical of firm-level administrative applications. What remains absent from this extensive predictive literature is firm-level predictive modelling executed inside the boundaries of science and technology parks. Existing applications address national exporter populations or highly regulated credit outcomes. No prior study utilizes machine learning to predict export success among the distinct tenants of a single park, nor does any study make the underlying drivers of such predictions transparent for local park management using theoretically grounded feature spaces.

2.5 Multi-Criteria Decision Making and Objective Weighting

Multi-criteria decision making supplies the analytical ranking machinery that pure predictive models inherently lack. While machine learning classifiers excel at establishing the statistical probability of a binary outcome based on historical data, they do not automatically generate a prescriptive ordering of alternatives across multiple conflicting dimensions. This conversion requires formal operational research methods. The fuzzy extension of TOPSIS by Chen [45] represents criterion ratings as triangular fuzzy numbers and computes geometric distances to ideal solutions using a specialized vertex method. This mathematical transformation is particularly valuable because it explicitly accommodates the inherent linguistic and observational imprecision prevalent in subjective administrative evaluation judgments. By converting crisp operational metrics into fuzzy sets, this methodology prevents marginal measurement errors in administrative data from disproportionately altering the final priority rankings. In the Turkish public policy context, AHP-TOPSIS has been used to rank technoparks as spatial location alternatives for technology companies [11], and network data envelopment analysis with common weights has been applied to benchmark science and technology parks against each other based on aggregated efficiency metrics [46]. Both established frameworks operate strictly at the aggregate macro institutional level, evaluating entire parks rather than individual firms. Furthermore, the application based on the analytical hierarchy process derives its internal criterion weights entirely from subjective expert pairwise comparisons.

The systemic reliance on expert elicitation represents the recognized weak point of this analytical tradition. When public authorities allocate tangible financial support, relying on subjective expert intuition limits the transparency of the decision-making process, inadvertently embeds human cognitive biases into the ranking algorithm, and makes the system notoriously difficult to audit or reproduce systematically over time. Ahmed *et al.,* [47] address this problem directly by developing a methodology where the calculated feature importances of a calibrated machine learning classifier iteratively replace subjective expert weights in a complex site selection model. Their methodology successfully derives multi-criteria weights from verifiable empirical patterns in historical data rather than relying on the stated preferences of human evaluators. Their approach suggests a much broader epistemological principle for policy analytics. Criterion weights can and theoretically should reflect each criterion measured empirical contribution to predicting the decision-relevant outcome, rather than reflecting subjective expert judgment or purely statistical properties such as variance or data dispersion. This integration of predictive machine learning with operational research provides a mathematically rigorous alternative to subjective administrative evaluations. Despite the theoretical strength of this integrated principle, no existing study applies this data-driven weighting technique to prioritize individual tenant firms within a single operational technopark, leaving a substantial methodological gap between predictive analytics and actionable resource allocation.

2.6 Synthesis and Identification of Literature Gaps

The comprehensive review of the theoretical frameworks and empirical methodologies discussed above reveals a significant disconnect between the analytical tools available for firm evaluation and the practical resource allocation challenges faced by technopark administrators. The traditional evaluation literature documents the heterogeneous effects of science parks but structurally lacks the prescriptive tools to manage this heterogeneity at the individual firm level [4]. Text-as-data methodologies successfully capture latent innovation signals from corporate narratives [8, 9], yet they remain largely unapplied to the non-English administrative texts generated within public innovation zones. Concurrently, the integration of explainable machine learning with multi-criteria decision making offers a mechanism to eliminate subjective bias in resource allocation, but this synthesis has not been explicitly adapted for intra-park management.

The proposed framework is positioned precisely at the intersection of four distinct analytical streams that have not previously been combined to address the resource allocation problem in technology parks. First, while exporter prediction models demonstrate high predictive accuracy at the national macroeconomic scale [10], they historically lack a localized technopark focus and operate entirely without a prescriptive allocation layer. Second, conventional technopark applications of multi-criteria decision making predominantly rank the parks themselves rather than the individual tenant firms, and they rely heavily on subjective expert opinions or purely mathematical data envelopment weights that fail to capture out-of-sample predictive validity [11]. Third, although the validity of large language models for complex text annotation is firmly established in general management research, this extraction technology has not been systematically validated for technopark evaluation frameworks using highly technical administrative texts [32, 33]. Finally, the broader empirical science park evaluation literature continues to focus almost entirely on retrospective between-park comparative designs, leaving park administrations without forward-looking, data-driven methodologies to differentiate among their own highly heterogeneous tenants [1–3].

This study systematically addresses these identified methodological divergences by integrating natural language processing, predictive machine learning, and objective multi-criteria decision making into a cohesive decision support tool. By extracting latent innovation features directly from native language project texts using a large language model, the methodology bypasses the structural lack of patent data in software-dominated zones. By predicting firm-level export success through explainable machine learning, the framework identifies the exact micro-level drivers of commercialization that average-effect studies typically obscure. Ultimately, by feeding the resulting objective SHAP importance metrics into a fuzzy prioritization algorithm, the study translates theoretical and predictive insights into a reproducible and criteria-driven ranking system for technopark managers. This synthesis ensures that subsequent resource allocation decisions are firmly anchored in empirical performance probabilities and theoretical capability accumulation models, rather than relying on subjective qualitative assessments or generic administrative heuristics.

3. Methodology

3.1 Research Design and Framework Architecture

This study employs a sequential mixed-methods research design to evaluate and prioritize tenant firms within a technology development zone. The analytical framework systematically integrates natural language processing, predictive machine learning, and operational research to overcome the limitations of traditional observational evaluations. Figure 1 illustrates the proposed hybrid decision support framework, which converts raw administrative and textual inputs into a reproducible and empirically grounded prioritized ranking of technopark firms. The inclusion of this visual architecture is critical for comprehending the sequential flow of data, as the outputs of each distinct computational module directly dictate the inputs of the subsequent stage.

The framework follows a structurally integrated three-phase logic. Phase I applies a large language model to the R&D project narratives of each firm and extracts a quantitative technological complexity score together with a domain label. Phase II trains a set of supervised classifiers on the combined feature set, predicts export success, and explains the winning model with Shapley-based attributions. Phase III converts these algorithmic attributions into objective criterion weights and ranks all firms with a fuzzy TOPSIS algorithm. Each phase consumes only the outputs of the preceding phase, so the full computational pipeline can be re-run, audited, and updated independently at every stage.

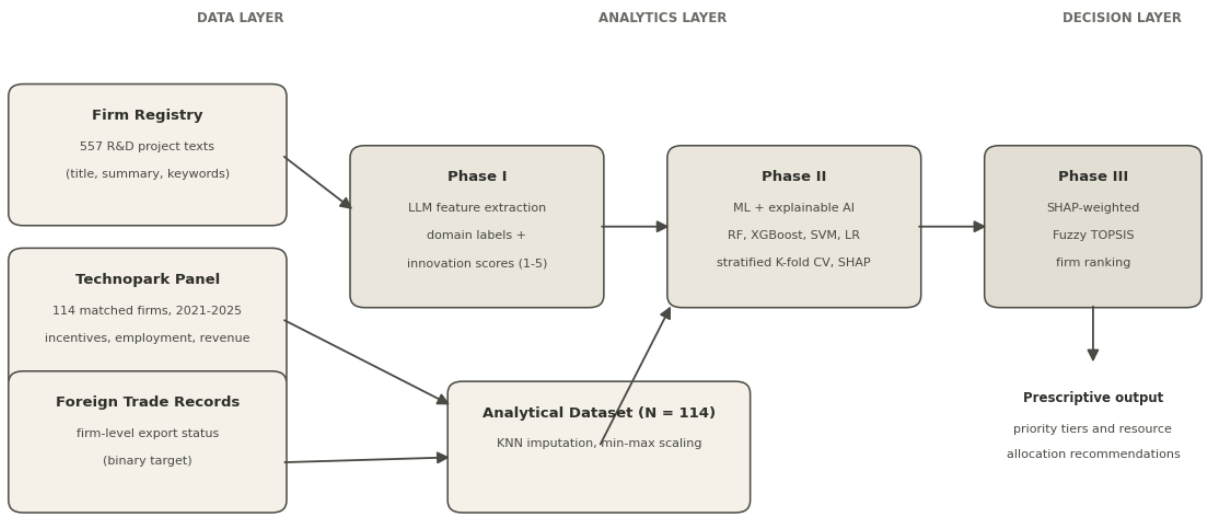


Fig. 1. The proposed hybrid decision support framework

A hybrid design is methodologically required because no single analytical component answers the decision problem on its own. The natural language processing algorithm computationally recovers latent technological information from project texts that standard administrative records do not contain. The predictive machine learning model establishes which specific firm characteristics actually discriminate exporters from non-exporters based on historical data rather than assuming their relevance a priori based on subjective institutional guidelines. The multi-criteria decision-making stage then translates these validated empirical relationships into an actionable priority ordering for park management. Prediction and prescription are therefore linked by mathematical construction, since the exact same model that demonstrates out-of-sample predictive validity also supplies the criterion weights used for the final resource ranking.

3.2 Research Setting and Data Collection

The empirical setting for this analysis is a major Turkish technopark located on the highly industrialized Istanbul and Kocaeli axis. Three distinct administrative sources covering the period 2021 to 2025 are combined to construct the analytical dataset. The firm registry contributes 557 R&D project texts. The technopark annual survey panel contributes 737 firm-year records across 252 park firms. Foreign trade records contribute deterministic firm-level export activity over the same five-year observation period.

The registry initially covers 171 firms, of which 167 have a counterpart in the panel after corporate name normalization. Because some registry entries are merely spelling variants or subsidiary branches of the same legal entity, these matches consolidate into 114 unique panel firms, which form the final cross-sectional analytical sample. Four sole proprietorships hold no comprehensive panel record and are explicitly excluded from the analysis to maintain data uniformity. Patent and utility-model variables show absolute zero variance in this software-dominated park, where 56.54 percent of firms in Turkish technology development zones nationally operate in computer programming [5]. Copyright and trade secrets substitute for formal intellectual property protection in this specific institutional setting. Innovation success is therefore modeled as a binary export outcome rather than a patenting outcome, in line with established text-based innovation measurement protocols developed specifically for non-patenting firms [8].

The target variable TARGET_EXP equals 1 if the firm recorded positive verified exports in at least one year between 2021 and 2025, and 0 otherwise. Thirty-six of the 114 firms qualify as exporters,

giving a positive class prevalence rate of 31.6 percent. The predictor set combines firm-level averages of the panel variables with the computationally derived innovation score produced in Phase I. Features and target are measured over the same five-year window, so the primary learning task is the contemporaneous discrimination of exporting firms rather than the longitudinal forecasting of future export market entry.

The analytical sample contains no missing values on any of the twelve independent features, so the reported performance metrics are completely unaffected by imputation artifacts. A k-nearest-neighbor imputation protocol with k equal to 5 is nevertheless retained in the computational pipeline as a structural safeguard for future annual updates of the panel. To rigorously prevent data leakage, this imputer is fitted exclusively within the training folds so that no statistical information from held-out validation observations can enter the model. All continuous features are rescaled utilizing min-max normalization, likewise fitted strictly within the training folds. Min-max scaling is specifically selected over standard score normalization because the downstream fuzzy TOPSIS algorithm requires all criterion inputs to strictly reside within a bounded positive interval. Right-skewed monetary and count variables enter the machine learning models in log-transformed form to stabilize extreme variance, while the descriptive statistics reported in Section 4 use the original administrative units to ensure practical readability. Table 1 defines every variable and specifies its operational role in the analysis.

Table 1
Variable Dictionary and Model Roles

Variable Code	Definition	Type	Role
TARGET_EXP	Positive exports in at least one year, 2021-2025	Binary	Target
F_EMP_TOT	Mean total employment over the panel period	Continuous	Predictor
F_INC_CORP	Mean corporate income tax exemption (TRY)	Continuous	Predictor
F_INC_PAY	Mean payroll income tax relief (TRY)	Continuous	Predictor
F_INC_VAT	Mean VAT exemption on software deliveries (TRY)	Continuous	Predictor
F_INC_SGK	Mean employer social security premium support (TRY)	Continuous	Predictor
F_RD_REV	Mean R&D revenue (TRY)	Continuous	Predictor
F_M2	Mean office space occupied in the park (m^2)	Continuous	Predictor
F_AGE_LOG	Log-transformed firm age in years	Continuous	Predictor
F_TENURE_LOG	Log-transformed years of park tenancy	Continuous	Predictor
F_UNI_COL	Any university collaboration during the panel period	Binary	Predictor
F_TUBITAK	Number of TÜBİTAK-supported projects	Count	Predictor
L_INN_SCR	Mean LLM innovation score across the firm's projects	Ordinal, 1-5	Predictor

3.3 Phase I. Automated Feature Extraction Via Large Language Models

Each official project text, comprising a formal title, an executive summary, and designated keywords, is scored by a large language model against a fixed five-level rubric of technological complexity. Within this qualitative rubric, a score of 1 denotes basic routine software implementation and a score of 5 denotes fundamental research-grade technological novelty. The automated scoring is performed utilizing Kimi (Moonshot AI) through a standardized prompt that embeds the evaluation rubric, the domain taxonomy, and a strictly formatted structured output schema. This extraction is applied under default model decoding settings to ensure deterministic outputs. The full translated prompt text is provided in Appendix A to guarantee experimental reproducibility. A single language model family scores all texts so that the measurement instrument remains perfectly constant across the entire sample space. The model simultaneously assigns each project one discrete domain label from a predefined taxonomy of 18 categories, including artificial intelligence, enterprise software, cybersecurity, and advanced manufacturing among others.

The methodological justification for employing algorithmic extraction over manual human coding rests on recent extensive validation studies in computational social science. Large language models have been empirically shown to consistently match or outperform trained crowd workers on complex text-annotation tasks [32]. They produce highly valid automated perceptual judgments in applied marketing settings [33] and successfully support highly reproducible qualitative data annotation in management research provided a fixed rubric and strictly documented prompts are consistently used [35]. Specific zero-shot benchmarks further confirm highly reliable classification performance for native Turkish text [37], which constitutes the primary administrative language of all project narratives in the evaluated registry.

The firm-level feature L_INN_SCR is calculated as the arithmetic mean of the technological complexity scores across all active projects belonging to a specific firm. This specific aggregation metric appropriately retains statistical variation in both the volume and the technical ambition of a corporate project portfolio. Simultaneously, the extracted domain labels characterize the overall technological composition of the park and are reported descriptively. A rigorous reliability protocol accompanies the automated extraction procedure to verify measurement consistency. A random subsample of 140 projects, corresponding exactly to 25 percent of all available texts extracted with a fixed randomization seed, is scored a second time under the identical analytical protocol. Test-retest agreement between these two independent passes reaches a quadratic-weighted Cohen kappa of 0.862, falling well within the statistical range universally accepted for almost-perfect agreement, alongside a Spearman correlation of 0.875. Criterion validity is subsequently assessed against an objective external empirical benchmark. The firm-level language model score correlates positively with the absolute count of competitively awarded national research projects, displaying a Spearman coefficient of 0.303. The automated score thus successfully tracks an independent empirical signal of research quality rather than merely restating basic firm size or operational age metrics.

3.4 Phase II. Predictive Modeling and Explainable Artificial Intelligence

Four distinct supervised machine learning classifiers are estimated on the standardized analytical sample. Standard logistic regression serves as a highly interpretable linear baseline. The empirical competitiveness of regularized logistic models against complex tree-based ensembles is well documented in applied corporate scoring and risk classification tasks [38, 39]. The support vector machine, random forest, and XGBoost algorithms provide progressively more flexible nonlinear classification alternatives. The XGBoost framework in particular has demonstrated robust empirical performance in firm-level bankruptcy and risk prediction when equipped with optimized hyperparameters [41].

The empirical exporter share of 31.6 percent constitutes a moderate class imbalance. Methodologically, this structural imbalance is addressed through algorithmic class weighting rather than synthetic data resampling techniques. Statistical literature indicates that synthetic resampling offers severely limited mathematical benefits at distribution ratios milder than the 30 to 70 threshold, and model explanation stability remains optimally preserved at this exact level of natural imbalance [44]. All candidate models are rigorously evaluated utilizing repeated stratified 5-fold cross-validation with 10 independent repetitions, yielding 50 distinct out-of-sample performance scores per algorithm. This specific resampling protocol is strongly recommended for limited administrative datasets because it mathematically reduces both the statistical variance and the optimistic bias inherent in small-sample out-of-sample performance estimates [43]. Hyperparameters are fixed a priori at conservative default values rather than tuned dynamically on the internal evaluation folds.

The classifier achieving the highest mean area under the curve proceeds to the explanation stage. Because the four specific financial incentive amounts are mathematically correlated by legislative construction, multicollinearity is rigorously assessed statistically prior to model interpretation. Variance inflation factors are computed on the transformed feature set, the full correlation structure is reported in Section 4.1, and the L2 regularization penalty of the logistic model provides critical additional stabilization for the coefficient estimates that directly feed the subsequent explanation stage.

To transition from accurate algorithmic prediction to actionable managerial insight, the framework utilizes Shapley Additive Explanations values. These values assign each feature a mathematically consistent marginal contribution to an individual prediction averaged over all possible feature coalitions [42]. The basic ensemble averaging equation originally utilized for model training has been deliberately omitted from this methodology to maintain focus on the core explanatory and prescriptive mathematical innovations. For any given feature j, the exact SHAP value is computed algebraically as defined in Eq. (1).

$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!\,(M-|S|-1)!}{M!} [f(S \cup \{j\}) - f(S)] \quad (1)$$

In Eq. (1), N represents the complete set of all M features, S represents a feature coalition that explicitly excludes feature j, f(S) represents the algorithmic model output conditioned solely on the features present in S, and ϕj represents the computed marginal contribution of feature j to the final prediction for a given firm. Positive SHAP values systematically push the classification probability toward the exporter class. For the winning linear model, these values are computed exactly utilizing the dedicated linear explainer and are expressed mathematically on the margin scale of the logit function.

3.5 Phase III. Objective Weighting and Multi-Criteria Prioritization

Global feature importance is mathematically converted into objective criterion weights through the mean absolute SHAP value calculated across all tenant firms in the analytical sample. These derived values are subsequently normalized so that the final criterion weights sum exactly to one. This normalization procedure is formalized in Eq. (2).

$$w_j = \frac{s_j}{\sum_{k=1}^{M} s_k} \quad (2)$$

In Eq. (2), wj denotes the normalized weight of criterion j, and M represents the total number of evaluated criteria. Criterion functional orientation is objectively assigned by the mathematical sign of the Spearman correlation between the actual feature values and their corresponding SHAP attributions. A negative statistical sign indicates a cost orientation, whereas all remaining criteria are computationally treated as benefit criteria. This data-driven weighting scheme methodologically replaces two highly flawed conventional alternatives with a single outcome-linked mathematical rule. Traditional subjective expert surveys embed severe cognitive judgment bias and logical inconsistency into resource allocation algorithms. Conversely, objective dispersion-based weights such as standard entropy reflect only random statistical variability in the dataset rather than actual empirical relevance to the targeted economic outcome. Deriving criterion weights directly from a validated predictive machine learning model successfully resolves both of these profound analytical weaknesses [47].

The tenant firms are subsequently ranked utilizing the fuzzy TOPSIS methodology, which explicitly extends the classical ideal-solution operational logic to accommodate triangular fuzzy numbers [45]. Each min-max normalized criterion value of a firm is computationally fuzzified into a triangular number utilizing a designated half-width parameter, which is strictly truncated to the unit interval.

This fuzzification step formally represents and mitigates the inherent measurement imprecision prevalent in administrative self-reported indicators. The triangular fuzzy transformation is defined in Eq. (3).

$$\tilde{x}_{ij} = \left(\max\left(0, x_{ij} - \delta\right),\ x_{ij},\ \min\left(1, x_{ij} + \delta\right)\right) \tag{3}$$

The mathematical sensitivity of the final priority ranking to this specific parameter choice is rigorously examined utilizing alternative half-widths of 0.05 and 0.20, while a crisp TOPSIS counterpart without any fuzzification is computed as an additional methodological benchmark. The weighted fuzzy decision matrix is calculated by multiplying each triangular coordinate by its corresponding crisp criterion weight derived from the SHAP analysis as shown in Eq. (4).

$$\tilde{v}_{ij} = w_j \cdot \tilde{x}_{ij} = \left(w_j l_{ij},\ w_j m_{ij},\ w_j u_{ij}\right) \tag{4}$$

In Eq. (4), lij, mij, and uij represent the lower, modal, and upper components of the fuzzified input xij. The fuzzy positive-ideal and negative-ideal solutions for each criterion are subsequently determined from the extreme upper and lower geometric components of the weighted matrix. These computational definitions apply for benefit criteria. For the single cost criterion, the mathematical extremes are inverted. The spatial distances between these triangular fuzzy numbers are computed employing the specialized vertex method. For any two triangular numbers a and b, the absolute distance is calculated as shown in Eq. (5).

$$d\left(\tilde{a}, \tilde{b}\right) = \sqrt{\frac{(l_a - l_b)^2 + (m_a - m_b)^2 + (u_a - u_b)^2}{3}} \tag{5}$$

Aggregating these geometric distances across all evaluated criteria yields the absolute mathematical separation of each individual firm from the theoretical positive-ideal and negative-ideal solutions. The ultimate closeness coefficient is then calculated to order the tenant firms from highest to lowest operational priority, as detailed in Eq. (6).

$$CC_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{6}$$

A firm achieving a closeness coefficient approaching unity sits mathematically near the positive-ideal empirical profile of a successful exporter and maximally far from the negative-ideal non-exporter profile. The resulting numerical coefficient generates a strictly ordinal ranking framework. This final prioritized list constitutes the definitive prescriptive output of the entire analytical pipeline, providing park management with an auditable, transparent, and empirically grounded operational tool for targeted resource allocation.

4. Results

4.1 Descriptive Statistics and Feature Extraction Outcomes

The empirical analysis begins with a comprehensive examination of the structural characteristics of the tenant firms alongside the automated text extraction results. Table 2 presents the detailed descriptive statistics for the twelve continuous and categorical features utilized in the predictive models across the 114 firms comprising the analytical sample. The descriptive profile reveals a highly skewed tenant distribution characteristic of emerging technology parks. The typical tenant is structurally small in operational scale. Mean total employment stands at 12.6 personnel with a maximum observation of 369.7, indicating that a few large anchor tenants coexist with a long statistical tail of micro firms. Mean research and development revenue equates to 30.1 million Turkish Liras against a sample maximum of 2.20 billion, revealing an identical concentration pattern on the commercial output side. Administrative incentive use is equally uneven across the tenant

population. The mean corporate income tax exemption is 3.28 million, yet the median firm receives only 0.16 million, while the largest annual average exceeds 132 million. Institutional linkages such as formal university collaboration remain structurally rare at 4.4 percent of the sample firms, and competitively awarded national grants average only 0.17 projects per firm. The technopark ecosystem therefore combines a small number of mature and heavily subsidized tenants with a broad foundational base of early-stage software and engineering ventures.

Table 2
Descriptive statistics of model features

Variable	N	Min	Median	Max	Mean	SD	Skew.
F_EMP_TOT	114	1.000	5.800	369.667	12.630	38.676	7.834
F_INC_CORP	114	0.000	0.157	132.576	3.277	13.298	8.459
F_INC_PAY	114	0.000	0.454	101.479	2.798	12.566	7.073
F_INC_VAT	114	0.000	0.000	33.856	1.269	4.626	5.911
F_INC_SGK	114	0.000	0.072	19.544	0.476	2.035	7.985
F_RD_REV	114	0.000	1.761	2196.158	30.091	206.323	10.289
F_M2	114	0.543	36.000	2716.000	96.645	278.177	7.703
F_AGE_LOG	114	0.000	1.884	3.867	1.884	1.002	-0.080
F_TENURE_LOG	114	0.042	1.494	3.150	1.573	0.968	0.070
F_UNI_COL	114	0.000	0.000	1.000	0.044	0.206	4.455
F_TUBITAK	114	0.000	0.000	6.000	0.167	0.677	6.310
L_INN_SCR	114	1.400	2.708	4.000	2.748	0.665	0.249

Note: Monetary features are measured in million Turkish Lira. Statistics are computed on the original administrative units prior to the logarithmic transformation applied for the predictive modeling phase. Firm age and park tenure variables represent the natural logarithms of their respective values in years.

All monetary and operational count variables are strongly right-skewed, exhibiting skewness coefficients between 5.911 and 10.289. This pronounced statistical asymmetry dictates the logarithmic transformation applied to these specific variables prior to predictive modeling. The two log-transformed demographic variables encompassing firm tenure and firm age are approximately symmetric with skewness values of 0.070 and -0.080 respectively. This confirms that the logarithmic transformation successfully normalizes the extreme values. The algorithmically derived innovation score is approximately normally distributed around a firm-level mean of 2.748 with a standard deviation of 0.665.

Table 3 details the Spearman correlation structure of the transformed model inputs alongside the computed variance inflation factors. The correlation matrix confirms the expected mathematical dependence within the administrative incentive framework. Research and development revenue correlates strongly with the corporate income tax exemption at 0.84. The employer social security premium support correlates with the payroll relief at 0.81 and with total employment at 0.74. Furthermore, park tenure correlates with general firm age at 0.84. All mentioned coefficients are statistically significant at the 0.001 level. These large magnitudes are direct mechanical consequences of the national incentive legislation, under which every

individual policy instrument scales proportionally with the eligible wage bill or with qualifying commercial income. Multicollinearity nevertheless remains completely within accepted econometric bounds. The largest observed variance inflation factor is 4.55 for park tenure, followed by 4.33 for firm age, and all remaining values fall safely below 3.6. Every coefficient is therefore estimated with tolerable mathematical precision, and the specific regularization penalty of the applied logistic model provides critical additional stabilization for the parameters.

Table 3
Spearman correlation matrix and variance inflation factors

	(1)	(2)	(3)	(4)	(5)	(6)
F_EMP_TOT	1					
F_INC_CORP	0.42***	1				
F_INC_PAY	0.56***	0.33***	1			
F_INC_VAT	0.16	0.51***	0.17	1		
F_INC_SGK	0.74***	0.41***	0.81***	0.16	1	
F_RD_REV	0.45***	0.84***	0.39***	0.56***	0.44***	1
F_M2	0.68***	0.44***	0.49***	0.12	0.59***	0.33***
F_AGE_LOG	0.40***	0.46***	0.17	0.19*	0.30**	0.36***
F_TENURE_LOG	0.29**	0.57***	0.21*	0.29**	0.28**	0.50***
F_UNI_COL	-0.04	-0.05	-0.00	-0.10	0.05	-0.06
F_TUBITAK	0.24*	0.22*	0.19*	0.08	0.20*	0.20*
L_INN_SCR	0.00	-0.29**	-0.10	-0.26**	-0.16	-0.26**

	(7)	(8)	(9)	(10)	(11)	(12)	VIF
F_EMP_TOT							2.58
F_INC_CORP							3.59
F_INC_PAY							2.17
F_INC_VAT							1.57
F_INC_SGK							2.89
F_RD_REV							3.14
F_M2	1						2.58
F_AGE_LOG	0.56***	1					4.33
F_TENURE_LOG	0.44***	0.84***	1				4.55
F_UNI_COL	0.02	0.07	0.07	1			1.17
F_TUBITAK	0.24**	0.20*	0.24**	0.23*	1		1.43
L_INN_SCR	-0.04	-0.08	-0.08	0.12	0.30**	1	1.33

Note. Correlations are computed on the logarithmically transformed model inputs. Statistical significance: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Variance inflation factors are computed on the standardized feature set to evaluate potential multicollinearity.

Figure 2 visually illustrates the overall technological composition of the park derived from the 557 algorithmically assigned domain labels. Enterprise software and enterprise resource planning projects form the largest categorical block at 23.3 percent, followed closely by manufacturing and industrial applications at 19.9 percent and financial technology at 11.7 percent. Artificial intelligence and machine learning projects account for 7.9 percent of the total portfolio, while defense systems alongside embedded internet of things applications each contribute 5.0 percent. The remaining twelve categories span diverse applications including health biotechnology, energy systems, and data analytics, which collectively constitute just over one quarter of the project portfolio. This structural composition perfectly mirrors the national macroeconomic profile of Turkish technology development zones, where computer programming structurally dominates total firm counts.

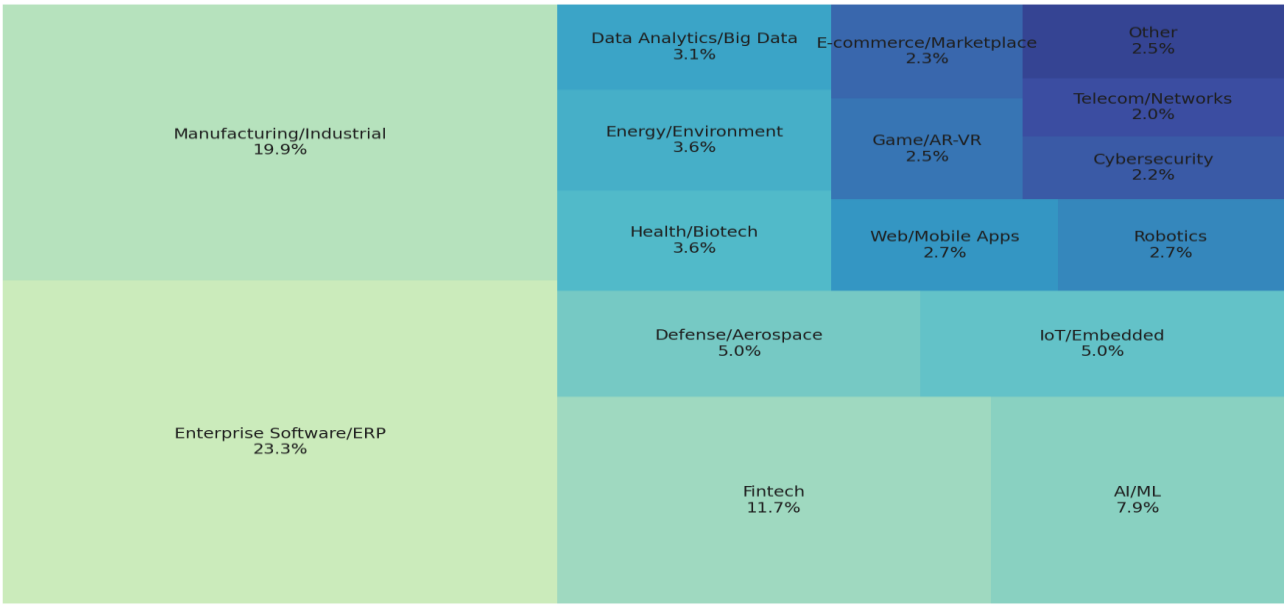


Fig. 2. Distribution of LLM-assigned technology domains across 557 R&D projects

The individual project-level innovation scores statistically concentrate in the exact middle of the five-level evaluation rubric. The calculated arithmetic mean is 2.78 with a standard deviation of 0.81. The specific distribution counts 9 projects at level 1, 226 at level 2, 203 at level 3, 115 at level 4, and only 4 projects achieving the maximum level 5. The substantial probability mass concentrated at levels 2 and 3 is highly realistic for a commercial tenant population predominantly focused on applied product development. The extremely thin upper tail accurately identifies a remarkably small group of genuinely research-grade technological efforts.

Table 4 empirically illustrates exactly what the automated extraction methodology captures and evaluates utilizing three anonymized project examples. Project A receives the maximum possible score because implementing a federated learning architecture under strict data privacy constraints and statistical network heterogeneity constitutes fundamental research-grade work. Project B scores a 4 because designing real-time seismic and acoustic signal processing algorithms for a safety-critical rescue device represents high-complexity applied engineering, despite lacking foundational algorithmic novelty. Project C scores a 2 because consolidating routine administrative human resources on a single digital platform characterizes standard enterprise software, and the designated algorithm correctly discounts vaguely described artificial intelligence additions utilized primarily for corporate marketing purposes.

Table 4
Sample LLM-based feature extractions from anonymized project texts

Case	Domain	Score	LLM reasoning	Translated text snippet
Project A	AI/ML	5	A federated learning platform tackling distributed neural network training, data privacy, and statistical heterogeneity with blockchain is research-grade work.	federated learning platform addressing data privacy and distributed neural network training
Project B	IoT/Embedded	4	Real-time seismic-acoustic signal processing with AI-based life-sign detection and classification in a safety-critical rescue device is high-complexity applied work.	AI-supported algorithms for life sign detection, noise filtering, signal classification

Table 4
Continued

Case	Domain	Score	LLM reasoning	Translated text snippet
Project C	Enterprise Software/ERP	2	A modular SME business-management platform consolidating tasks, HR, and expenses is a typical enterprise software product with vaguely described AI additions.	bringing all in-company processes together on a single platform

4.2 Predictive Model Performance and Validation

Table 5 details the out-of-sample classification performance across 50 repeated stratified cross-validation runs for the four candidate machine learning classifiers alongside a naive baseline prevalence metric. Regularized logistic regression attains the highest mean score across all six evaluated performance metrics. The linear model achieves an area under the receiver operating characteristic curve of 0.810 with a 95 percent confidence interval spanning 0.788 to 0.831. Furthermore, the model records an overall accuracy of 0.738, a minority class recall of 0.765, an F1 score of 0.651, and a precision-recall area under the curve of 0.737 against a strict base rate prevalence of 0.316. The support vector machine ranks second with an area metric of 0.765, while the random forest acts as the strongest nonlinear ensemble alternative at 0.751. XGBoost statistically trails the benchmark group at 0.684. A Nadeau and Bengio corrected resampled statistical test executed on the paired fold-level predictions confirms that the predictive advantage of the logistic regression over the random forest algorithm is statistically significant, and the uncorrected Wilcoxon signed-rank test similarly returns a probability value strictly less than 0.001.

Table 5
Out-of-sample classification performance across 50 stratified cross-validation runs

Model	Accuracy	Precision	Recall	F1	AUC-ROC	PR-AUC
Logistic Regression	0.738 (0.079)	0.576 (0.110)	0.765 (0.126)	0.651 (0.095)	0.810 (0.076)*	0.737 (0.100)
SVM	0.733 (0.082)	0.573 (0.114)	0.735 (0.146)	0.635 (0.104)	0.765 (0.090)	0.646 (0.112)
XGBoost	0.701 (0.096)	0.548 (0.171)	0.565 (0.191)	0.538 (0.149)	0.684 (0.094)	0.563 (0.125)
Random Forest	0.714 (0.077)	0.568 (0.154)	0.518 (0.190)	0.522 (0.144)	0.751 (0.083)	0.625 (0.116)
Majority baseline	0.684 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.500 (0.000)	0.316 (0.000)

Note: Cells report mean values with standard deviations in parentheses calculated over 50 evaluation runs. Metrics are computed specifically for the exporter class, which constitutes 31.6 percent of the sample. Statistical significance against the nonlinear benchmark is explicitly indicated by an asterisk.

The empirical advantage of the regularized linear model aligns with algorithmic benchmarking in small samples. Highly flexible ensembles struggle to estimate complex interaction surfaces from few observations, requiring cross-validation to maintain unbiased estimates [43]. The fold-level standard deviations reinforce this finding. Random forest and XGBoost display recall standard deviations of 0.190 and 0.191 respectively, contrasted against 0.126 for the logistic regression. This stability is consistent with corporate classification findings where regularized linear models frequently rival advanced tree-based methods [38]. Operationally, the validated recall of 0.765 means the algorithm detects roughly three of every four actual exporting firms, satisfying the administrative requirement to discover entities with latent export potential.

Evaluated against a static baseline predicting all firms as non-exporters, which yields an accuracy of 0.684 and an area under the curve of 0.500, the optimized linear model lifts predictive accuracy to 0.738 and the primary discrimination metric to 0.810. Because only 31.6 percent of the evaluated firms actually export, an operational decision support tool must clear this baseline by a wide margin

to demonstrate practical utility. Achieving an area metric of 0.810 with 114 tenant firms and twelve predictors approaches the predictive performance reported in national-scale exporter prediction studies involving tens of thousands of observations [10].

Two experimental checks validate the construction of the predictor set. First, replacing the firm-level arithmetic mean of project scores with the firm maximum score yields an area metric of 0.808. The resulting feature weight vector correlates with the baseline at 0.986, and the downstream firm priority ranking correlates at 0.996. These metrics confirm that the predictive outcomes do not depend on the specific aggregation rule, thereby addressing concerns that averaging penalizes firms combining ambitious research with routine implementations. Second, an ablation experiment completely removing the language model innovation score from the feature matrix reduces the area metric from 0.810 to 0.802. Although this difference does not reach statistical significance at this restricted sample size, the partial correlation of the language score conditional on existing research revenue remains robust at 0.173. The language model feature therefore operates as a complementary information channel. Its contribution to the predictive framework is empirically supported by its high test-retest reliability, its positive correlation with external benchmark grants, and its stable rank in the final attribution structure.

4.3 Feature Attribution and Explanatory Insights

Explaining the internal logic of the optimal predictive model directly converts statistical accuracy into actionable managerial knowledge. Table 6 reports the classical econometric parameter estimates derived from the winning regularized logistic model, which is fitted utilizing balanced frequency weights on the fully scaled feature set. The mathematical signs of the unpenalized estimates agree completely with the penalized predictive model on eleven of the twelve coefficients, and the calculated odds ratios perfectly reproduce the underlying attribution logic. Because all continuous input features are strictly min-max scaled to the unit interval, the odds ratios represent the exact multiplicative effect on the calculated odds of exporting when shifting a firm mathematically from the sample minimum to the sample maximum. For example, moving a firm from the absolute minimum to the absolute maximum park tenure observed in the sample multiplies the predicted odds of exporter status by 14.8. Individual logistic coefficients are nevertheless estimated with inherently limited precision given the restricted sample size, and none individually clears the 5 percent statistical significance level in strict isolation, although the joint model discriminates highly effectively out of sample.

The calculated Shapley values mathematically decompose each individual firm-level classification into additive feature contributions [42]. Table 7 reports the resulting global average importance of every evaluated feature together with the normalized criterion weight and mathematical orientation explicitly carried into the final phase. Throughout this specific subsection, a positive attribution value systematically pushes the final prediction toward exporter status, and a negative value pushes it away. The designated functional orientation follows directly from the mathematical sign of the Spearman correlation between the scaled feature values and their respective attributions. Benefit criteria represent features whose higher values raise the predicted export probability, whereas cost criteria function in the strict opposite mathematical direction.

Figure 3 visualizes the attribution summary plot specifically for the positive exporter class, where each singular data point represents one firm, the horizontal position dictates the calculated attribution value, and the visual color encodes the original feature value from low to high. Figure 4 subsequently presents this identical global importance data mathematically aggregated as normalized percentile shares. High numerical values of institutional tenure, research revenue, corporate tax exemption amounts, general firm age, and the algorithmic innovation score all cluster

densely at positive attribution values. Conversely, the value added tax exemption strictly reverses this statistical pattern, with high categorical values concentrating intensely at negative attribution values. The top five most critical features jointly account for exactly 71.4 percent of total model importance. The algorithmic explanation is therefore highly concentrated rather than statistically diffuse, which is precisely the mathematical property that renders the derived absolute weights highly usable for multi-criteria prioritization.

Table 6
Classical logistic regression estimates for exporter status

Feature	Coef.	SE	z	p	Odds ratio
F_EMP_TOT	1.528	2.133	0.72	0.474	4.61
F_INC_CORP	0.742	1.180	0.63	0.530	2.10
F_INC_PAY	0.298	1.224	0.24	0.808	1.35
F_INC_VAT	-0.841	0.728	-1.16	0.248	0.43
F_INC_SGK	0.554	1.382	0.40	0.689	1.74
F_RD_REV	1.318	1.154	1.14	0.253	3.74
F_M2	-1.036	2.097	-0.49	0.621	0.36
F_AGE_LOG	0.744	2.424	0.31	0.759	2.10
F_TENURE_LOG	2.695	1.939	1.39	0.165	14.81
F_UNI_COL	0.769	1.317	0.58	0.560	2.16
F_TUBITAK	3.668	2.613	1.40	0.160	39.16
L_INN_SCR	1.263	1.151	1.10	0.273	3.53

Note: Estimates are derived from a binomial generalized linear model applying balanced frequency weights on the scaled feature set. The constant term is computed at -3.992. No individual coefficient reaches standard statistical significance alone, supporting the methodological reliance on joint algorithmic attribution.

Table 7
Global SHAP importance and derived MCDM criterion weights

Feature	Mean abs. SHAP	Weight	Share (%)	Orientation
F_TENURE_LOG	0.4280	0.2268	22.7	Benefit
F_RD_REV	0.2865	0.1518	15.2	Benefit
F_INC_CORP	0.2296	0.1217	12.2	Benefit
F_AGE_LOG	0.2184	0.1157	11.6	Benefit
L_INN_SCR	0.1854	0.0982	9.8	Benefit
F_INC_SGK	0.1503	0.0796	8.0	Benefit
F_INC_VAT	0.1370	0.0726	7.3	Cost
F_TUBITAK	0.0746	0.0396	4.0	Benefit
F_INC_PAY	0.0647	0.0343	3.4	Benefit
F_EMP_TOT	0.0636	0.0337	3.4	Benefit
F_UNI_COL	0.0372	0.0197	2.0	Benefit
F_M2	0.0118	0.0063	0.6	Benefit

Note: Weights represent the mean absolute attribution values mathematically normalized to sum exactly to one. The functional orientation is explicitly assigned by the statistical correlation between the raw feature values and their respective attributions.

The feature attribution strictly details the statistical drivers of export success. Institutional park tenure emerges as the single strongest predictive feature, capturing 22.7 percent of the total global importance. The data indicates that cumulative years of tenancy correlate strongly with export probability, an empirical pattern consistent with longitudinal observations of park tenant performance [2, 3]. General firm age ranks fourth overall at 11.6 percent, demonstrating that the predictive model successfully separates park-specific tenure from standard organizational maturity.

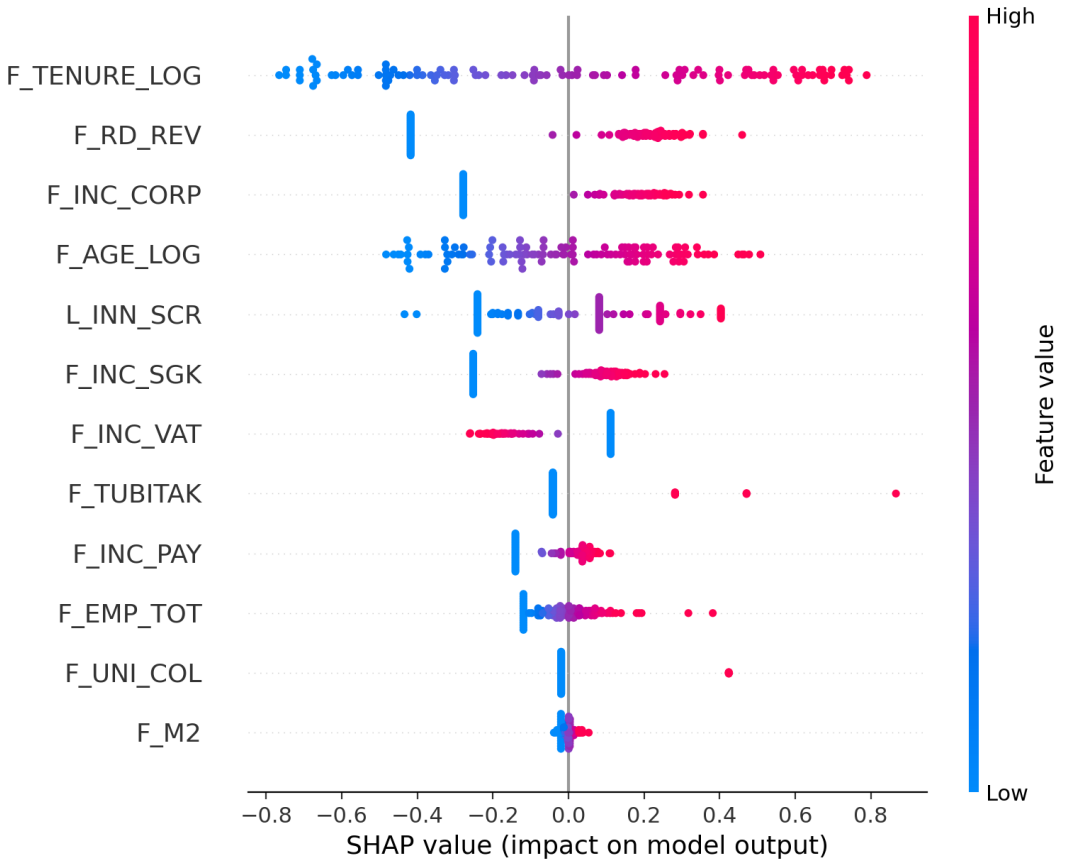


Fig. 3. SHAP summary plot for the exporter class

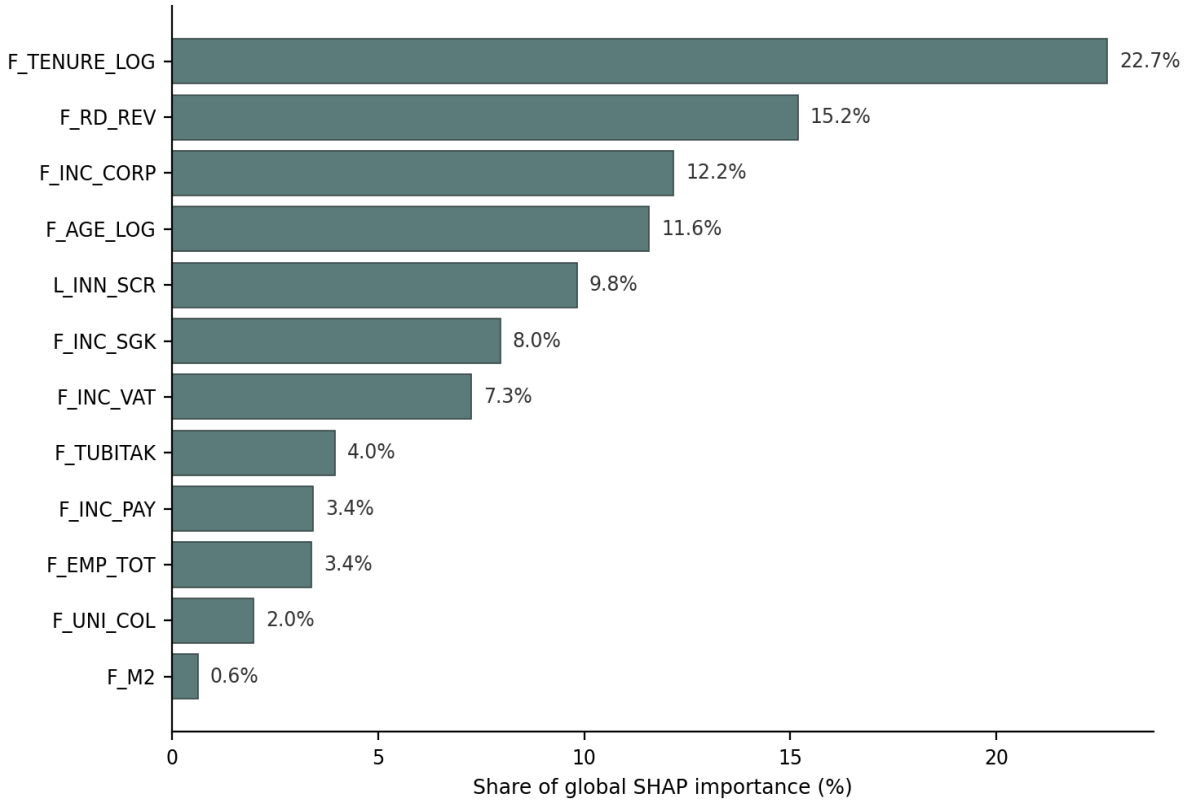


Fig. 4. Global feature importance shares

Total research revenue at 15.2 percent and the corporate income tax exemption volume at 12.2 percent collectively form the secondary predictive tier. Because the corporate tax exemption mathematically accrues in proportion to generated research earnings, both features function as direct empirical indicators of commercial research operation at scale. The employer social security premium support ranks sixth at 8.0 percent, scaling directly with the aggregate research wage bill. The feature attribution thus demonstrates that the absolute intensity of institutional incentive use is strongly associated with a higher probability of exporting.

The automated innovation score derived from text ranks fifth overall at 9.8 percent of global importance. This textual metric notably places ahead of total employment, which contributes only 3.4 percent to the predictive model. The finding indicates that the semantic content of a project portfolio predicts export success more effectively than the sheer volume of personnel employed, supporting prior text-as-data applications in innovation measurement [8]. The inclusion of unstructured project texts therefore provides discriminating statistical power completely independent of conventional headcount metrics.

The value added tax exemption operates as the sole cost-oriented criterion within the framework at 7.3 percent, indicating that higher exemption amounts correlate with lower predicted export probabilities. Because this specific tax exemption applies exclusively to software deliveries sold domestically, its absolute amount mechanically identifies tenant firms oriented strictly toward the domestic market. This empirical separation of domestic-oriented financial instruments from export success provides firm-level data relevant to broader evaluations of incentive heterogeneity and potential crowding out effects [6, 7].

4.4 Prescriptive Analytics and Firm Ranking

The structural transition from explainable prediction to prescriptive analytics represents a change in the operational question rather than a change of the underlying mathematical model. The identical normalized weights that successfully explain the logistic classifier become the exact criterion weights, and the same historical features that accurately predict empirical export success become the precise multidimensional axes along which all tenant firms are rigorously ranked. Table 8 presents the absolute top 15 of the 114 evaluated firms ordered by the weighted fuzzy procedure, explicitly reporting the exact geometric distances to the theoretical positive-ideal and negative-ideal mathematical solutions alongside the final closeness coefficient rounded to four decimals.

Table 8
Top 15 firms by SHAP-weighted Fuzzy TOPSIS closeness coefficient

Rank	Firm	Sector	D+	D-	CC	Exporter	Recommendation
1	Firm_083	Software	0.1320	0.8046	0.8590	Yes	High priority, scale-up support
2	Firm_088	Electronics	0.1473	0.7942	0.8436	Yes	High priority, scale-up support
3	Firm_085	Other	0.2257	0.7136	0.7597	Yes	High priority, scale-up support
4	Firm_021	Software	0.2375	0.7019	0.7472	Yes	High priority, scale-up support
5	Firm_053	ICT	0.2610	0.6820	0.7233	Yes	High priority, scale-up support
6	Firm_075	Engineering services	0.2623	0.6771	0.7208	Yes	High priority, scale-up support
7	Firm_006	Other	0.2633	0.6771	0.7200	No	Hidden gem, targeted export mentoring
8	Firm_078	Software	0.2674	0.6729	0.7156	Yes	High priority, scale-up support
9	Firm_059	Software	0.2776	0.6620	0.7045	Yes	High priority, scale-up support
10	Firm_081	Software	0.2795	0.6604	0.7026	Yes	High priority, scale-up support
11	Firm_060	Software	0.2842	0.6554	0.6975	Yes	High priority, scale-up support

Table 8
Continued

Rank	Firm	Sector	D+	D-	CC	Exporter	Recommendation
12	Firm_102	ICT	0.3024	0.6374	0.6782	No	Hidden gem, targeted export mentoring
13	Firm_014	Software	0.3036	0.6365	0.6771	Yes	High priority, scale-up support
14	Firm_086	Software	0.3076	0.6314	0.6724	No	Hidden gem, targeted export mentoring
15	Firm_099	Software	0.3085	0.6331	0.6723	Yes	High priority, scale-up support

Note: Geometric distances are computed utilizing the specified vertex method on the weighted triangular fuzzy numbers. The complete prioritized list of all 114 firms remains available for administrative deployment.

The empirical composition of the top priority ranks validates the ranking framework against the commercial outcome it was designed to predict. Twelve of the top fifteen firms are confirmed exporters, and the top six positions are exclusively occupied by exporters commanding closeness coefficients strictly above 0.71. The leading enterprise, a software firm designated as Firm 083, attains a coefficient of 0.8590. Sectorally, software firms occupy nine of the fifteen available positions, consistent with the domain structure of the overall park. However, electronics, engineering services, and communication firms also reach the top tier, proving that the algorithm does not mechanically reproduce sector membership. The compressed spacing of numerical coefficients from rank 6 down to rank 15, spanning only 0.0485 points, indicates that priority differences below the top tier remain marginal.

The three non-exporting firms located at ranks 7, 12, and 14 constitute the primary prescriptive value of the framework. Identified as latent high-potential firms, they combine advanced organizational profiles and strong text-derived innovation scores with absolutely no recorded international exports during the evaluation window. The predictive model places them within a few hundredths of a numerical point of confirmed exporters. The firm located at rank 7 actually outranks six confirmed historical exporters within the top 15 tier. Their empirical profiles indicate an underlying organizational export readiness that has not yet converted into measurable commercial activity.

Unlike existing technopark applications of multi-criteria decision making that typically rank macro parks and derive weights from subjective expert judgment [11], the methodology presented here provides an entirely firm-level and completely data-driven prioritization. Four structural checks support the stability of this prescriptive ranking. First, sequentially varying the fuzzification half-width parameter from the baseline of 0.1 to alternative values of 0.05 and 0.20 yields priority rankings that correlate with the baseline ordering at Spearman coefficients of 0.9999 and 0.9998. A crisp counterpart computed without any fuzzification similarly reproduces the baseline ordinal ranking at 0.9997, confirming that the triangular representation of measurement imprecision does not artificially drive the core results.

Second, re-deriving the initial criterion weights from the completely different random forest algorithm yields a new weight vector that correlates with the baseline logistic weights at only 0.664. Yet, the resulting prescriptive firm ranking generated from these new weights correlates strongly at 0.982, confirming that the final ordering remains highly stable irrespective of the specific machine learning model used for weight extraction. Third, traditional entropy and statistical weighting methodologies produce weight vectors that correlate with the computationally derived attributions at only 0.000 and 0.531 respectively. This divergence confirms that the objective weights successfully encode predictive contribution rather than mere statistical variability, overcoming the inherent limitations of pure dispersion-based schemes [47]. Despite this fundamental difference in weight generation, the final output rankings under the alternative weighting regimes correlate at 0.870 and

0.887 with the proposed predictive ordering. The prescriptive administrative conclusions therefore do not hinge arbitrarily on the chosen weighting scheme.

5. Discussion

5.1 Theoretical Implications

The empirical results align with the capability accumulation perspective developed in the theoretical framework. Institutional park tenure dominates the explanation of export success, and it operates independently of general firm age. Under the resource-based view [12], this statistical separation identifies two fundamentally distinct stocks of firm-specific capital. General age captures standard organizational maturity accumulated in any generic business environment. Conversely, park tenure captures path-dependent capabilities built inside the specific institutional context of the technology zone [15]. Firms accumulate this location-specific capital by learning to utilize specialized park services, absorbing the complex incentive regime, and integrating into localized knowledge networks. This finding provides direct within-park empirical evidence for a theoretical proposition that the existing evaluation literature has historically tested through broad comparisons between parks and their external surroundings. If spatial park effects genuinely exist, they should materialize precisely through such accumulated location-specific capabilities. The statistical dominance of tenure in a predictive model containing twelve competing administrative variables supports this theoretical premise [4].

The secondary tier of predictive features functions as a direct empirical proxy for realized absorptive capacity. Total research revenue and the corporate income tax exemption both scale mechanically with the historical volume of research and development activity. Absorptive capacity theory predicts this operational channel. Firms convert external public support into measurable commercial output in proportion to their accumulated prior related knowledge [18]. Consequently, the identical national incentive regime produces heterogeneous commercial returns across different tenant firms. The applied machine learning framework captured this underlying heterogeneity without being programmed with the national legislative rules. The absolute intensity of incentive use effectively signals the organizations possessing ongoing technical programs, marking them as the entities most statistically capable of competing in international export markets.

Furthermore, the automated innovation score derived from natural language models behaves as the economic theory of intangible resources suggests it should. Its standalone linear association with export success remains relatively weak, yet it ranks fifth within a stable algorithmic attribution structure. Its predictive signal strengthens once conventional financial metrics such as research revenue are statistically controlled. This pattern confirms that technological distinctiveness does not substitute for baseline commercial capability. Instead, it complements financial metrics, which characterizes the expected profile of a valuable but not yet fully commercialized intangible knowledge resource [12]. This complementary dynamic also clarifies the negative mathematical direction of the value added tax exemption. Because this specific exemption accrues exclusively on domestic software deliveries, it measures pure market orientation rather than underlying technical capability. Its cost orientation within the predictive model separates domestic-market firms from active exporters without implying any structural failure of the policy instrument itself.

5.2 Integration with the Empirical Evaluation Literature

The findings of this study extend the technology park evaluation canon in three distinct methodological directions. First, the traditional between-park comparative literature frequently reports only marginal average advantages for on-park locations compared to off-park equivalents [1, 21]. Leading systematic reviews attribute this statistical inconclusiveness to unmodeled

organizational heterogeneity [4]. The present results empirically demonstrate how this unmodeled heterogeneity manifests inside a single operational park. Park tenure and specific absorptive capacity proxies separate exporters from non-exporters with an out-of-sample area metric of 0.810. This high classification accuracy suggests that traditional average-effect estimates artificially compress the precise firm-level statistical variation that matters most for practical park administration.

Second, existing duration-based empirical evidence indicating that park tenants sustain stronger innovative output progressively over the firm life cycle receives a direct firm-level counterpart within this study [2, 3]. Institutional tenure accurately predicts commercial success even after general age, physical size, incentive intensity, and project content are controlled. This specific statistical isolation is the signature of park-specific capability accumulation rather than generic corporate maturation. While reverse causality caveats apply symmetrically to both literatures since successful exporters possess the financial resources to extend their physical tenancy, the underlying empirical reading remains associational and capability-driven in both contexts.

Third, the proposed framework connects two methodological literatures that have not previously interacted. Exporter prediction models currently exist primarily at the national macroeconomic scale, evaluating tens of thousands of manufacturing firms entirely without a prescriptive localized allocation layer [10]. Simultaneously, text-based innovation measurement literature demonstrates that narrative corporate content reveals critical latent innovation activity that remains invisible to conventional statistical indicators [8, 9]. The present methodological framework proves that both conceptual approaches can operate effectively at the localized park scale, and that their mathematical combination makes prescriptive resource allocation possible. The algorithmic prediction identifies the exact profile of a successful exporter, while the text-derived feature supplies the qualitative information that numerical administrative records omit.

The findings regarding administrative incentives also contribute directly to the ongoing economic debate regarding policy additionality. Empirical evidence from the Turkish context estimates relatively modest intensity gains from research tax incentives, exhibiting sub-unity multipliers [6]. Parallel park-level evidence from different emerging economies shows that generic public funding can crowd out firms possessing strong internal research capabilities [7]. The feature attribution framework deployed in this study sharpens that critical debate at the level of individual policy instruments. The volume of income-linked exemptions closely tracks export-relevant operational capability, while domestically oriented tax instruments isolate non-exporters. A uniform incentive expansion applied equally across all tenants therefore predictably misallocates public support. The generated objective ranking specifically resolves this inefficiency by identifying exactly where the marginal unit of public intervention carries the largest expected commercial return.

5.3 Managerial and Policy Implications

The transition from predictive explanation to prescriptive ranking generates specific policy recommendations for regional technology management. The firm-level prioritization framework operates directly as a strategic portfolio tool capable of directing differentiated administrative instruments to specific tenant clusters based on their algorithmic profiles. The identification of latent high-potential firms constitutes the primary managerial value of this system. These entities possess strong textual innovation scores and advanced structural profiles but lack recorded international exports. A conventional administrative allocation rule based strictly on historical export performance or operational headcount would pass over them entirely.

For these identified latent organizations, park management should prioritize targeted export mentoring and international market discovery programs. Practical interventions include providing funded participation in international trade fairs, facilitating subsidized matchmaking with foreign

commercial buyers, and delivering technical guidance on complex export documentation protocols. The algorithmic model indicates that the binding operational constraint for these entities is international market access rather than a fundamental lack of technological capability. National support agencies, including KOSGEB and TÜBİTAK, can efficiently target their limited internationalization grants exactly at this specific subgroup. Directing public financial support to these highly capable but domestically confined firms is most likely to generate genuine policy additionality rather than subsidizing inframarginal operations.

Conversely, for the confirmed exporters dominating the absolute top of the numerical ranking, comprehensive scale-up policy instruments are appropriate. Because their primary binding constraint is physical operational capacity rather than basic market entry, these top-ranked firms require advanced growth financing and physically expanded park office space. Finally, national science agencies can utilize the ordinal firm ranking as an objective supplementary screening layer during national project evaluations. By utilizing the framework to accurately flag high-ranking non-exporters, public agencies can efficiently route these specific organizations into advanced commercialization-oriented funding programs rather than fundamental research tracks. By shifting criterion weights from subjective expert elicitation toward objective predictive attributions, this framework provides public administrators with a completely data-driven and reproducible mechanism for allocating scarce innovation resources.

5.4 Limitations and Future Research Directions

Four specific limitations bound the current findings and establish a clear agenda for future academic research. First, the empirical setting remains restricted to a single major technology park. Comprehensive multi-park replication across fundamentally different sectoral profiles is required before these specific predictive parameters can be generalized to broader national contexts. Second, the analytical design is structurally contemporaneous. Features and target outcomes are measured over the identical five-year observation window. Consequently, the study functions as a rigorous profiling and classification exercise rather than a true causal forecasting model. Reverse causality cannot be completely excluded because highly successful exporting firms may simply accumulate higher research revenues and sustain longer institutional tenures over time. Future longitudinal research designs incorporating strictly lagged predictors to forecast actual export market entry and long-term survival would successfully convert this profiling framework into a definitive causal prediction tool. Third, the targeted commercial outcome is purely binary. Incorporating continuous export intensity volume metrics would considerably sharpen the economic interpretation of the individual policy incentives. Finally, the textual extraction phase depends entirely on a single proprietary large language model. Implementing a formal human expert validation panel alongside multi-model ensemble scoring would further strengthen the reliability of the initial feature extraction. Developing richer textual variables including embedding-based semantic representations and complex domain composition metrics offers a promising avenue for extending text-based evaluation within regional innovation ecosystems.

6. Conclusion

The question of whether science and technology parks efficiently allocate public resources remains critically important for regional economic development. The existing empirical evaluation literature primarily investigates average park effects, providing minimal actionable guidance regarding exactly which specific tenants should receive scarce public support. This study successfully bridges that prescriptive gap by developing a comprehensive hybrid decision support framework. By utilizing a large language model to extract technological complexity from unstructured project

narratives and applying explainable machine learning classifiers to predict export success, the framework accurately isolates the micro-level drivers of commercialization. The analysis reveals that localized institutional tenure and realized absorptive capacity function as the dominant predictors of export readiness. Furthermore, by translating algorithmic feature attributions into objective criteria weights for a multi-criteria prioritization model, the methodology eliminates reliance on subjective expert judgments. The resulting data-driven ranking system successfully identifies latent high-potential firms exhibiting strong export capabilities without recorded international sales. Methodologically, the research shifts technology park evaluation away from generic observational comparisons toward actionable prescriptive analytics. Ultimately, this framework provides park administrators and policy agencies with an auditable, reproducible, and empirically grounded tool to optimize the targeted allocation of public innovation resources.

Appendix

Appendix A. LLM Scoring Protocol

All 557 project texts were scored with Kimi (Moonshot AI), a large language model, through the fixed prompt reproduced below, applied under default decoding settings. Texts were processed in batches of 25, and each output record was validated for schema compliance before aggregation. The identical prompt was used for the test-retest reliability pass.

Rubric and instructions given to the model, translated from the operational prompt.

Score every project on two dimensions. First, assign exactly one domain label from the following taxonomy. AI/ML, IoT/Embedded, Enterprise Software/ERP, Web/Mobile Apps, Cybersecurity, Game/AR-VR, Health/Biotech, Energy/Environment, Manufacturing/Industrial, Fintech, Edutech, E-commerce/Marketplace, Defense/Aerospace, Telecom/Networks, Data Analytics/Big Data, Robotics, HR/Business Services, Other. Second, assign an integer technological complexity score from 1 to 5 using this rubric.

1. Routine implementation of well-established technology with no novelty, such as basic transactional websites, standard e-commerce setups, or simple content portals.

2. Conventional application of established technology with minor adaptation or customization, such as typical ERP or CRM modules, standard mobile apps, or simple integrations.

3. Moderate novelty requires meaningful engineering, such as integration of modern frameworks or sensors, applied data processing, or non-trivial system design.

4. High complexity or clear novelty, such as applied machine learning models, embedded hardware plus cloud integration, advanced algorithms, or real-time and safety-critical systems.

5. Cutting-edge, research-grade work with potential for genuine scientific or technological advance, such as novel algorithms, deep learning architecture, or frontier hardware-software co-design.

Score the technical ambition described in the text, not the business potential. When in doubt between two levels, choose the lower one. For each project also return one English sentence of at most 25 words justifying the score, and an English translation of the most informative 8-to-15-word fragment of the text. Return only structured records, one per project.

Appendix B. Classifier Configurations

All classifiers were estimated with fixed, a priori configurations under scikit-learn and XGBoost defaults unless listed. Every pipeline applies KNN imputation (k = 5) and min-max scaling fitted within the training folds only. Table B1 summarizes the classifier configurations.

Table B1
Classifier configurations

Classifier	Configuration
Logistic regression	class_weight = balanced, max_iter = 5000, L2 penalty
Support vector machine	RBF kernel, class_weight = balanced, probability estimates enabled
Random forest	n_estimators = 500, min_samples_leaf = 2, class_weight = balanced
XGBoost	n_estimators = 200, max_depth = 3, learning_rate = 0.05, min_child_weight = 3, reg_lambda = 2.0, subsample = 0.9, colsample_bytree = 0.9, scale_pos_weight = class ratio

The evaluation protocol is repeated stratified 5-fold cross-validation with 10 repeats (seed 42), producing 50 out-of-sample scores per model and metric. The winning logistic model is explained with the linear SHAP explainer on the margin scale, and the random forest benchmark is explained with the tree SHAP explainer.

Acknowledgement
This research was not funded by any grant.

Conflicts of Interest
The authors declare no conflicts of interest.

References
[1] Lindelöf, P., & Löfsten, H. (2002). Growth, management and financing of new technology-based firms—assessing value-added contributions of firms located on and off science parks. Omega, 30(3), 143–154. https://doi.org/10.1016/S0305-0483(02)00023-3
[2] Squicciarini, M. (2008). Science parks' tenants versus out-of-park firms: Who innovates more? A duration model. The Journal of Technology Transfer, 33(1), 45–71. https://doi.org/10.1007/s10961-007-9037-z
[3] Squicciarini, M. (2009). Science parks: Seedbeds of innovation? A duration analysis of firms' patenting activity. Small Business Economics, 32(2), 169–190. https://doi.org/10.1007/s11187-007-9075-9
[4] Albahari, A., Barge-Gil, A., Pérez-Canto, S., & Landoni, P. (2023). The effect of science and technology parks on tenant firms: A literature review. The Journal of Technology Transfer, 48(4), 1489–1531. https://doi.org/10.1007/s10961-022-09949-7
[5] T.C. Sanayi ve Teknoloji Bakanlığı. (2026). Teknoloji Geliştirme Bölgeleri İstatistiki Bilgiler (data as of end-February 2026). Republic of Türkiye Ministry of Industry and Technology, Ankara. https://www.sanayi.gov.tr/istatistikler/istatistiki-bilgiler/mi0203011501
[6] Taş, E., & Erdil, E. (2024). Effectiveness of R&D tax incentives in Turkey. Journal of the Knowledge Economy, 15(2), 6226–6272. https://doi.org/10.1007/s13132-023-01326-5
[7] Teng, T., Zhang, Y., Si, Y., Chen, J., & Cao, X. (2020). Government support and firm innovation performance in Chinese science and technology parks: The perspective of firm and sub-park heterogeneity. Growth and Change, 51(2), 749–770. https://doi.org/10.1111/grow.12372
[8] Bellstam, G., Bhagat, S., & Cookson, J. A. (2021). A text-based analysis of corporate innovation. Management Science, 67(7), 4004–4031. https://doi.org/10.1287/mnsc.2020.3682
[9] Kelly, B., Papanikolaou, D., Seru, A., & Taddy, M. (2021). Measuring technological innovation over the long run. American Economic Review: Insights, 3(3), 303–320. https://doi.org/10.1257/aeri.20190499
[10] Micocci, F., & Rungi, A. (2023). Predicting exporters with machine learning. World Trade Review, 22(5), 584–607. https://doi.org/10.1017/S1474745623000265
[11] Durak, İ., Arslan, H. M., & Özdemir, Y. (2022). Application of AHP–TOPSIS methods in technopark selection of technology companies: Turkish case. Technology Analysis & Strategic Management, 34(10), 1109–1123. https://doi.org/10.1080/09537325.2021.1925242
[12] Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99–120. https://doi.org/10.1177/014920639101700108
[13] Wernerfelt, B. (1984). A resource-based view of the firm. Strategic Management Journal, 5(2), 171–180. https://doi.org/10.1002/smj.4250050207

[14] Peteraf, M. A. (1993). The cornerstones of competitive advantage: A resource-based view. Strategic Management Journal, 14(3), 179–191. https://doi.org/10.1002/smj.4250140303
[15] Dierickx, I., & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. Management Science, 35(12), 1504–1511. https://doi.org/10.1287/mnsc.35.12.1504
[16] Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. Organization Science, 2(1), 88–115. https://doi.org/10.1287/orsc.2.1.88
[17] Levitt, B., & March, J. G. (1988). Organizational learning. Annual Review of Sociology, 14(1), 319–338. https://doi.org/10.1146/annurev.so.14.080188.001535
[18] Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly, 35(1), 128–152. https://doi.org/10.2307/2393553
[19] Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. Academy of Management Review, 27(2), 185–203. https://doi.org/10.5465/amr.2002.6587995
[20] Lane, P. J., & Lubatkin, M. (1998). Relative absorptive capacity and interorganizational learning. Strategic Management Journal, 19(5), 461–477. https://doi.org/10.1002/(SICI)1097-0266(199805)19:5<461::AID-SMJ953>3.0.CO;2-L
[21] Lindelöf, P., & Löfsten, H. (2003). Science park location and new technology-based firms in Sweden–implications for strategy and performance. Small Business Economics, 20(3), 245–258. https://doi.org/10.1023/A:1022861823493
[22] Lecluyse, L., Knockaert, M., & Spithoven, A. (2019). The contribution of science parks: A literature review and future research agenda. The Journal of Technology Transfer, 44(2), 559–595. https://doi.org/10.1007/s10961-018-09712-x
[23] Theeranattapong, T., Pickernell, D., & Simms, C. (2021). Systematic literature review paper: The regional innovation system–university–science park nexus. The Journal of Technology Transfer, 46(6), 2017–2050. https://doi.org/10.1007/s10961-020-09837-y
[24] Rui, C., Lokshin, B., & Mohnen, P. (2023). Heterogeneity in performance of science and technology parks in China: Is there "club" convergence? Papers in Regional Science, 102(6), 1145–1168. https://doi.org/10.1111/pirs.12759
[25] Akçomak, İ. S., & Taymaz, E. (2004). Assessing the effectiveness of incubators: The case of Turkey. ERC Working Paper 04/12, Middle East Technical University.
[26] Vaidyanathan, G. (2008). Technology parks in a developing country: The case of India. The Journal of Technology Transfer, 33(3), 285–299. https://doi.org/10.1007/s10961-007-9041-3
[27] Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. Journal of Economic Literature, 57(3), 535–574. https://doi.org/10.1257/jel.20181020
[28] Dell, M. (2025). Deep learning for economists. Journal of Economic Literature, 63(1), 5–58. https://doi.org/10.1257/jel.20241733
[29] Korinek, A. (2023). Generative AI for economic research: Use cases and implications for economists. Journal of Economic Literature, 61(4), 1281–1317. https://doi.org/10.1257/jel.20231736
[30] Arts, S., Hou, J., & Gomez, J. C. (2021). Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. Research Policy, 50(2), 104144. https://doi.org/10.1016/j.respol.2020.104144
[31] Bowen, D. E., III, Frésard, L., & Hoberg, G. (2023). Rapidly evolving technologies and startup exits. Management Science, 69(2), 940–967. https://doi.org/10.1287/mnsc.2022.4362
[32] Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. Proceedings of the National Academy of Sciences, 120(30), e2305016120. https://doi.org/10.1073/pnas.2305016120
[33] Brand, J., Israeli, A., & Ngwe, D. (2023). Using LLMs for market research. Harvard Business School (Working Paper No. 23–062). https://doi.org/10.2139/ssrn.4395751
[34] Li, P., Castelo, N., Katona, Z., & Sarvary, M. (2024). Frontiers: Determining the validity of large language models for automated perceptual analysis. Marketing Science, 43(2), 254–266. https://doi.org/10.1287/mksc.2023.0454
[35] Carlson, N. A., & Burbano, V. (2026). The use of LLMs to annotate data in management research: Foundational guidelines and warnings. Strategic Management Journal, 47(3), 699–725. https://doi.org/10.1002/smj.70023
[36] Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. Business & Information Systems Engineering, 66(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7
[37] Çelik, E., & Dalyan, T. (2023). Unified benchmark for zero-shot Turkish text classification. Information Processing & Management, 60(3), 103298. https://doi.org/10.1016/j.ipm.2023.103298
[38] Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

[39] Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. European Journal of Operational Research, 297(3), 1178–1192. https://doi.org/10.1016/j.ejor.2021.06.053

[40] Xia, Y., Liu, C., Li, Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. Expert Systems with Applications, 78, 225–241. https://doi.org/10.1016/j.eswa.2017.02.017

[41] Ben Jabeur, S., Stef, N., & Carmona, P. (2023). Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering. Computational Economics, 61(2), 715–741. https://doi.org/10.1007/s10614-021-10227-1

[42] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in neural information processing systems (Vol. 30). Curran Associates. https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

[43] Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. PLOS ONE, 14(11), e0224365. https://doi.org/10.1371/journal.pone.0224365

[44] Chen, Y., Calabrese, R., & Martin-Barragan, B. (2024). Interpretable machine learning for imbalanced credit scoring datasets. European Journal of Operational Research, 312(1), 357–372. https://doi.org/10.1016/j.ejor.2023.06.036

[45] Chen, C. T. (2000). Extensions of the TOPSIS for group decision-making under fuzzy environment. Fuzzy Sets and Systems, 114(1), 1–9. https://doi.org/10.1016/S0165-0114(97)00377-1

[46] Özsoy, V. S., Belgin, Ö., & Balkan, D. (2022). A novel approach for determining common weights in two division network DEA: A case study of science and technology parks in Turkey. Technology Analysis & Strategic Management, 34(10), 1124–1138. https://doi.org/10.1080/09537325.2021.1947491

[47] Ahmed, M., Dogru, A., Zhang, C., & Meng, C. (2025). Learning-based multi-criteria decision model for site selection problems. [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2504.04055