VERİMLİLİK DERGİSİ
JOURNAL OF PRODUCTIVITY

Machine Learning-Based Feature Selection Analysis of Academic Spin-Off Survival in Technoparks Located in Türkiye*

Başak Apaydın Avşar[1], Mehmet Yılmaz[2]

ABSTRACT

Purpose**:** This study aims to identify the key determinants influencing the survival of academic spin-off (ASO) firms operating in Technology Development Zones (TDZs) in Türkiye. It contributes to the limited empirical evidence on the long-term sustainability of university-originated ventures in emerging innovation ecosystems.

Methodology**:** An original dataset covering all ASOs active between 2021 and 2024 was analysed using Mutual Information, Random Forest importance, Recursive Feature Elimination (RFE), and a Genetic Algorithm (GA). Class imbalance was addressed through SMOTE applied only to the training set, and predictor contributions were interpreted using SHAP.

Findings: RFE achieved the highest predictive performance (Accuracy = 0.9837; ROC-AUC = 0.9958). The number of ongoing projects emerged as the strongest predictor of ASO survival, reflecting the regulatory requirement for maintaining at least one active project. Additionally, R&D expenditures, public R&D support, and incubation participation enhance firms' financial resilience and increase the likelihood of continued operation.

Originality: This study is the first data-driven research to examine ASO survival in Türkiye using multiple feature selection techniques combined with explainable artificial intelligence. The findings offer evidence-based insights for policymakers seeking to strengthen the sustainability of academic entrepreneurship.

Keywords**:** Academic Spin-Offs, Firm Retention, Machine Learning, SHAP, SMOTE.
JEL Codes**:** C81, L26, M13, O31, R11.

Türkiye'deki Teknoparklarda Yer Alan Akademik Spin-Off'ların Hayatta Kalma Durumunun Makine Öğrenmesi Tabanlı Özellik Seçimi Analizi

ÖZET

Amaç**:** Bu çalışma, Türkiye'deki Teknoloji Geliştirme Bölgeleri'nde (TGB) faaliyet gösteren akademik spin-off (ASO) firmalarının hayatta kalmasını etkileyen temel belirleyicileri ortaya koymayı amaçlamaktadır. Araştırma, üniversite kökenli girişimlerin sürdürülebilirliğine ilişkin sınırlı ampirik literatüre katkı sunmaktadır.

Yöntem**:** 2021–2024 döneminde TGB'lerde aktif olan tüm ASO'ları kapsayan veri seti; Karşılıklı Bilgi (Mutual Information), Rastgele Orman önem düzeyi, Özyinelemeli Özellik Eleme (RFE) ve Genetik Algoritma (GA) yöntemleri kullanılarak analiz edilmiştir. Sınıf dengesizliği yalnızca eğitim verisine uygulanan SMOTE yöntemiyle giderilmiş; değişkenlerin etkileri SHAP ile yorumlanmıştır.

Bulgular**:** En yüksek tahmin performansı RFE yöntemiyle elde edilmiştir (Doğruluk = 0,9837; ROC-AUC = 0,9958). Devam eden proje sayısı, mevzuat gereği proje sürekliliğinin zorunlu olması nedeniyle ASO hayatta kalmasının en güçlü belirleyicisi olarak öne çıkmaktadır. Ar-Ge harcamaları, kamu Ar-Ge destekleri ve kuluçka programlarına katılım ise firmaların finansal dayanıklılığını artırarak hayatta kalma olasılığını yükseltmektedir.

Özgünlük**:** Bu çalışma, Türkiye'de akademik spin-off hayatta kalmasını çoklu özellik seçimi yöntemleri ve açıklanabilir yapay zekâ teknikleriyle bütüncül biçimde inceleyen ilk veri odaklı araştırmadır. Sonuçlar, akademik girişimciliğin sürdürülebilirliğini artırmaya yönelik politika yapıcılar için önemli kanıta dayalı çıkarımlar sunmaktadır.

Anahtar Kelimeler**:** Akademik Spin-Off, Firma Kalıcılığı, Makine Öğrenmesi, SHAP, SMOTE.
JEL Kodları**:** C81, L26, M13, O31, R11.

---

* This study was derive from the Doctoral Thesis titled "Predicting the Survival of Entrepreneurial Firms in Technology Development Zones in Türkiye Using Artificial Intelligence and Machine Learning Techniques" conducted by Başak Apaydın Avşar at Ankara University Institute of Science under the supervision of Prof. Dr. Mehmet Yılmaz.

[1] Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, Türkiye
[2] Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara, Türkiye

Corresponding Author: Başak Apaydın Avşar, bapaydin89@gmail.com

DOI: 10.51551/verimlilik.1840976
Research Article |Submitted: 12.12.2025 |Accepted: 12.01.2026
Cite: Apaydın Avşar, B. and Yılmaz, M. (2026). "Machine Learning-Based Feature Selection Analysis of Academic Spin-Off Survival in Technoparks Located in Türkiye", *Verimlilik Dergisi,* 60(1), 277-294.

1. INTRODUCTION

Universities have undergone a profound transformation in the shift toward knowledge-based economies, evolving from traditional educational institutions into central actors within national innovation systems. One of the most visible outcomes of this transformation is the emergence of academic spin-off (ASO) firms, which commercialise university-generated knowledge, convert scientific research into marketable technologies and play an important role in regional economic development (Djokovic and Souitaris, 2008; Hossinger et al., 2020). Although international research has extensively examined the antecedents of academic entrepreneurship at the individual, organisational and regional levels, considerably less attention has been devoted to understanding the survival of spin-offs, an essential dimension for evaluating their long-term economic contribution (Soetanto and van Geenhuizen, 2019; Rodeiro-Pazos, 2021).

In Türkiye, academic entrepreneurship was institutionally formalised with Law No. 4691 on Technology Development Zones (TDZs) (Official Gazette, 2001), enacted in 2001 to strengthen university–industry collaboration, promote technology-intensive firms and accelerate the commercialisation of research outputs. Subsequent policies, including Law No. 5746 on Supporting R&D and Design Activities (Official Gazette, 2008) and the TÜBİTAK 1513 Technology Transfer Offices Support Programme (TÜBİTAK, 2025), expanded the innovation infrastructure and contributed to the increasing formation of ASOs nationwide. However, despite this strong institutional environment, empirical knowledge regarding the survival patterns of ASOs in Türkiye remains limited. Recent indicators underscore the importance of this question. However, despite this strong institutional environment, empirical knowledge regarding the survival patterns of ASOs in Türkiye remains limited. Recent indicators underscore the importance of this question. By 2025, Türkiye hosted 113 active TDZs and a total of 12,235 firms operating within these zones, including 2,252 academic-partnered firms. In the same period, the aggregate sales of all TDZ firms exceeded 1 trillion TRY, exports reached 15.5 billion USD and the total number of patents rose to 2,424 (Republic of Türkiye Ministry of Industry and Technology, 2025). While these figures demonstrate rapid growth, they do not reveal which ASOs survive over time or which characteristics most strongly determine their persistence.

In the Turkish Technology Development Zone context, firm survival is not solely a market-driven outcome but is also shaped by compliance-based institutional conditions. Remaining active in a Technology Development Zone is closely linked to regulatory requirements, most notably the obligation to maintain at least one ongoing R&D project and to employ a minimum level of R&D personnel as defined by the legal framework. These requirements give rise to threshold-based structures in which firm status remains unchanged as long as minimum compliance conditions are satisfied, but may shift abruptly once these conditions are no longer met. In addition, firm-level financial, R&D and project-related indicators display highly skewed distributions and interact in complex ways that are unlikely to be adequately captured by linear or semi-parametric modelling assumptions. This institutional and empirical structure motivates the use of flexible, non-linear machine learning methods capable of capturing interaction effects, nonlinearities and regulatory thresholds inherent in the TDZ environment.

The existing literature reveals several important gaps. First, empirical studies on ASO survival remain heavily concentrated in Europe and North America, leaving Türkiye's TDZ ecosystem largely unexplored. Second, most prior research relies on traditional econometric approaches such as logistic regression, Cox proportional hazards models and small-scale categorical analyses, which impose linearity assumptions and have limited ability to capture nonlinear and multidimensional relationships. Third, although machine learning techniques have increasingly been applied to entrepreneurship research, studies that combine multiple feature selection algorithms such as Mutual Information, Random Forest importance, Recursive Feature Elimination and Genetic Algorithms in examining ASO survival are virtually nonexistent. Fourth, despite growing interest in high-performing algorithms, the literature rarely employs interpretable artificial intelligence tools to clarify how predictors influence survival outcomes. Finally, the interaction between these determinants and Türkiye's pronounced regional heterogeneity in innovation capacity, institutional infrastructure and university industry linkages remains insufficiently explored. Addressing these gaps is essential for generating evidence-based insights that can guide more effective policy-making.

Against this backdrop, the purpose of this study is to identify the determinants of academic spin-off survival in Türkiye's Technology Development Zones using a comprehensive machine learning framework. To achieve this aim, the study addresses two research questions:

(1) Which factors are most important in predicting the survival of academic spin-offs?

(2) Which machine learning method achieves the highest predictive accuracy in estimating the determinants of academic spin-off survival?

The originality of this research lies in integrating multiple feature selection algorithms with interpretable machine learning techniques to provide the first systematic and data-driven analysis of ASO survival in

Türkiye. By clarifying the mechanisms that underlie survival probabilities, the study contributes to both methodological advancement and practical policy design within emerging entrepreneurial ecosystems. Applying multiple feature selection techniques is central to this study, as it allows identification of the most influential determinants of academic spin-off survival, reduces redundancy among correlated predictors, and enhances both interpretability and policy relevance of the findings.

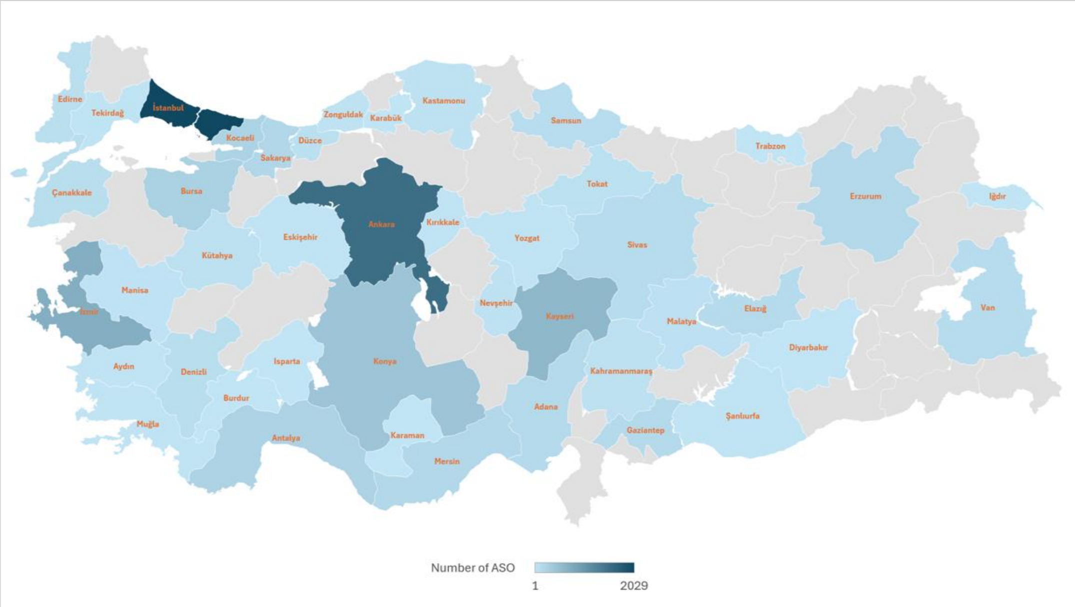

Figure 1. Academic spin-off density by province in Türkiye

Figure 1 presents the provincial distribution of academic spin-off density in Türkiye and reveals a clear spatial concentration pattern. The darkest provinces on the map, namely İstanbul, Ankara and İzmir, represent the country's leading innovation centres. These provinces host major research universities, strong R&D infrastructures and mature technology ecosystems, collectively providing a fertile environment for spin-off creation. Ankara and İstanbul, in particular, accommodate a high number of Technology Development Zones, public research institutions and advanced laboratories, resulting in the highest concentration of ASO activity. Several Central Anatolian provinces such as Konya, Eskişehir, Kayseri and Kocaeli also display moderate levels of ASO density, reflecting the presence of technically oriented universities, established industrial clusters and expanding university industry collaboration mechanisms. In contrast, many eastern and southeastern provinces exhibit lower ASO density, signalling persistent regional disparities in research capabilities, access to finance and institutional support structures. Overall, the map shows that ASO activity in Türkiye is heavily concentrated in regions with strong innovation ecosystems, high research intensity and greater absorptive capacity. This spatial distribution underscores the importance of regionally differentiated policies aimed at strengthening institutional capacities, enhancing technology transfer mechanisms and increasing the engagement of universities in local innovation systems.

This study is organized as follows: Section 2 reviews the theoretical and empirical literature on academic spin-offs, firm survival and machine learning applications. Section 3 explains the methodological framework, including data sources, variable construction, preprocessing steps and feature selection procedures. Section 4 presents the empirical results and discusses model performance and SHAP based interpretations. Section 5 concludes by summarising key findings and offering policy implications for strengthening the sustainability of academic entrepreneurship in Türkiye.

2. LITERATURE REVIEW

The literature on academic spin-offs (ASOs) has evolved along several complementary dimensions shaped by individual motivations, institutional frameworks and broader innovation systems. Within the entrepreneurial university paradigm, Etzkowitz (2003) conceptualises research groups as organisational units that increasingly resemble firms by adopting structures and routines that facilitate entrepreneurial activity. At the individual level, studies show that role identity transformation and localised social learning processes significantly influence researchers' engagement in commercialisation activities (Bercovitz and Feldman, 2008; Jain et al., 2009). Institutional conditions also play a central role in ASO formation. Universities with strong entrepreneurial orientations, effective technology transfer offices and high-quality research environments generate more spin-offs and achieve stronger technology transfer outcomes (O'Shea et al., 2008; Powers and McDougall, 2005). Recent conceptual work highlights that ASOs constitute a distinctive organisational form that cannot be fully explained by any single theory of the firm. Prokop (2023) proposes a pluralistic conceptual framework that combines insights from resource based, knowledge based, dynamic capabilities and transaction cost perspectives to explain ASO boundaries, decision making mechanisms and growth constraints. Complementary research shows that internal university regulations influence ASO creation and performance. Evidence from Italian universities indicates that rules related to general procedures, monetary incentives and the distribution of entrepreneurial risk shape institutional capacity to generate new ventures. Monetary incentives support spin-off formation and restrictive rules on contract research can discourage academic entrepreneurship (Muscio et al., 2016). Collectively, these studies show that ASOs are embedded within institutional structures that influence their emergence and long-term sustainability.

A substantial body of empirical research has examined the determinants of ASO survival. Founders' human capital has important effects. University specific and entrepreneurial experience increase the likelihood of survival, while some types of industry experience are associated with higher exit risk (Criaco et al., 2014). Incubation environments also influence performance outcomes. University based incubation affects whether ASOs adopt exploration oriented or exploitation oriented strategies and these strategies shape innovation outcomes (Soetanto and Jack, 2016). Spatial and relational proximity to universities is another determinant of long-term performance. The effect is not linear. Very strong proximity may lead to diminishing returns and the magnitude of the effect depends on contextual conditions such as entrepreneurial orientation and market hostility (Soetanto and van Geenhuizen, 2019). Firm level determinants have also been identified. Firm size, intellectual property strategies and export orientation influence survival patterns. There is evidence that a minimum efficient scale reduces the marginal effect of size on failure risk (Rodeiro Pazos, 2021).

Table 1. Selected empirical studies on university spin-off survival and performance

Author(s)	*Topic*	*Key Findings*
Prokop (2023)	Academic Spinoff Theory of the Firm	Proposes a multidimensional theoretical framework to explain ASO boundaries, decision processes, and growth constraints.
Rodeiro-Pazos (2021)	Firm size, patents, exports	Firm size and export intensity influence survival; a minimum efficient scale reduces size-related failure risk.
Civera et al. (2020)	Opportunity vs. necessity motives	Necessity-driven academic spin-offs exhibit higher survival probabilities, whereas opportunity-driven spin-offs show faster post-entry growth.
Soetanto and van Geenhuizen (2019)	University proximity	Balanced spatial and relational proximity enhances long-term performance; effects are non-linear and context-dependent.
Civera et al. (2019)	Internationalisation	ASOs internationalise earlier and more intensively, and the international orientation of their parent university reinforces this process.
Fackler et al. (2016)	Parent organisation effects	Parent organisation size, quality, and capabilities significantly influence spin-off survival.
Visintin and Pittino (2014)	Founding team diversity	Team diversity in skills and backgrounds improves early performance.
Criaco et al. (2014)	Founders' human capital	Academic and entrepreneurial human capital increases survival; some industry experience types weaken survival.
Czarnitzki et al. (2014)	Performance premium	The findings indicate that university spin-offs achieve higher employment growth compared to industry start-ups.
Rodríguez-Gulías et al. (2017)	University–regional spillovers	University and regional knowledge spillovers enhance ASO growth and performance.
Soetanto and Jack (2016)	Incubation & innovation strategy	Incubation support and strategic orientation jointly shape innovation performance.
Rodríguez-Gulías et al. (2016)	Growth determinants	ASO growth depends on internal resources and the surrounding university–regional ecosystem.
Muscio et al. (2016)	University rules & incentives	Monetary incentives strengthen spin-off creation; overly restrictive internal regulations hinder entrepreneurial activity.
Zhang (2009)	VC-backed USOs	Venture Capital (VC)-backed spin-offs exhibit higher survival, though growth outcomes are mixed.

Recent studies have examined heterogeneity in ASO motivations. Using the distinction between opportunity driven and necessity driven entrepreneurship, Civera et al. (2020) show that these groups follow different post entry trajectories. Necessity driven ASOs tend to show more stable survival patterns, while opportunity driven ventures exhibit faster early-stage growth. University engagement and technology transfer support may work differently for these groups, although the mechanisms depend on institutional context. Internationalisation is another relevant factor. Civera et al. (2019) find that academic spin-offs internationalise earlier and generate a larger share of foreign sales compared with non academic start-ups. Their affiliation with globally oriented parent universities contributes to these outcomes. These results indicate that early and intensive foreign market exposure is associated with better long-term performance.

Despite these contributions, relatively few studies have analysed ASO survival using advanced analytical approaches. Much of the empirical literature relies on regression based models and Cox type hazard models (Wennberg et al., 2011). These approaches are informative for identifying average relationships but have limited ability to capture complex interactions, multicollinearity and non-linear patterns that characterise ASO behaviour. Studies in related areas show that machine learning methods can model multidimensional dependencies more flexibly and improve predictive accuracy (Wang et al., 2019). Applications that combine feature selection with explainable artificial intelligence are still limited in ASO research. Overall, the literature shows that ASO survival is influenced by interconnected factors at the individual, institutional and environmental levels. This complexity highlights the relevance of supervised learning and explainable artificial intelligence techniques for identifying interactions among these determinants. A synthesis of influential empirical studies examining spin-off performance, survival and long-term development is presented in Table 1.

3. METHODOLOGY

This study employs a supervised machine learning framework to identify the determinants of academic spin-off (ASO) survival within Turkish Technology Development Zones (TDZs). The methodological workflow consists of four core stages: (i) data preprocessing, (ii) feature selection, (iii) model training and validation, and (iv) model interpretation using explainable artificial intelligence tools. The combined use of resampling, feature selection and ensemble based learning is consistent with best practices recommended in the applied machine learning literature (Kuhn and Johnson, 2019: 27-80 ; Géron, 2022: 68-75).

In this study, ASO survival status is defined based on official administrative records of the TDZs. A firm is coded as surviving if it remained active within the zone as of the end of 2024. Firms that exited the zone or became inactive due to not submitting a new project within the legally required period are classified as non surviving. Since no consistent data exist regarding off park activities, off zone continuity is not included in the survival definition.

3.1 Data Preprocessing

The dataset comprises all academic spin-offs operating in Turkish TDZs between 2021 and 2024. Firm level information was obtained from the Entrepreneur Information System (GBS), which maintains official administrative records for companies located within Turkish Technology Development Zones. The final dataset includes a total of 7,973 firm-level observations covering the 2021–2024 period. The dependent variable indicates whether a firm remained active within the TDZ (1) or became inactive or exited (0). Firms that left the zone or did not submit a new project within the legally required period were classified as non surviving. Since no consistent data exist regarding off park activities, survival is defined strictly based on TDZ administrative status.

Independent variables include demographic, structural and performance related characteristics such as firm age, scale, incubation status, origin, foreign partnership, employment composition, sales, exports, R&D revenue, R&D expenditures, public support amounts, tax incentives, project counts and intellectual property indicators. A detailed description of these variables is presented in Table 2. No explanatory variables were excluded prior to modelling; instead, all available variables were retained and subsequently evaluated through supervised feature selection methods within the modelling pipeline. Categorical variables are encoded using a hybrid approach. Binary categories such as incubation, origin and foreign partnership are label encoded. Multinomial variables such as firm scale are one hot encoded to avoid imposing artificial ordinality. Numerical features are normalised via Min Max scaling to reduce model sensitivity to scale differences, which is a standard practice shown to enhance optimisation and convergence in supervised learning (Han et al., 2012: 83-117). The dataset does not contain missing observations for the variables included in the analysis; therefore, no imputation procedure was required.

Because the dependent variable is imbalanced, the Synthetic Minority Over Sampling Technique (SMOTE) is applied exclusively within the training folds. The original class distribution of active and inactive firms is reported using descriptive statistics and is visually presented in Figure 2 (Active = 4,771; Inactive = 3,202), together with the post-SMOTE class distribution. SMOTE generates synthetic minority class samples by interpolating between neighbouring observations (Chawla et al., 2002) and has been shown to improve classifier sensitivity under imbalance (Fernández et al., 2018: 98-114). To avoid information leakage and preserve the integrity of model evaluation, SMOTE is not applied to the test data, which is consistent with widely accepted recommendations in the imbalanced learning literature (Branco et al., 2016).

This approach is particularly important in the present context, as inactive firms constitute the minority class but represent the primary outcome of interest from both a predictive and policy perspective. Accordingly, SMOTE is used as a controlled training-time adjustment to improve minority-class sensitivity, while model evaluation and interpretation are conducted on the original data distribution. While SMOTE constitutes the primary imbalance-handling technique in the main modelling pipeline, the sensitivity of the results to alternative resampling strategies is explicitly examined through robustness analyses reported in Section 4.5.

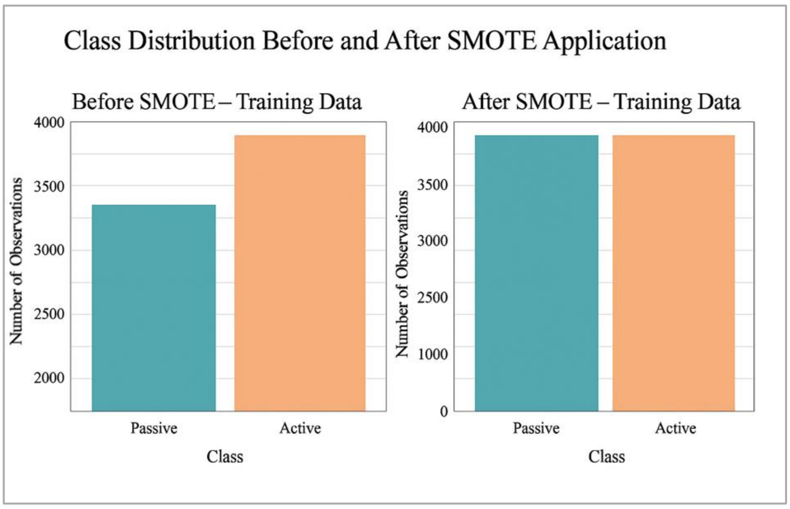

Figure 2. Changes in class distribution before and after SMOTE application

Table 2. Variables used in the analysis

Variable	*Type*	*Description / Measurement*
Firm Age (Lifespan)*	Numerical	Total number of years the firm has operated within the TDZ.
Scale	Categorical	Firm size category: micro, small, medium, large.
Incubated	Binary	Indicates whether the firm is or was supported by an incubation centre.
Origin	Binary	Domestic vs foreign origin of the firm.
Foreign Partnership	Binary	Indicates whether the firm has foreign equity partners.
Total Employment	Numerical	Total number of active employees.
Total Sales	Numerical	Total domestic sales revenue (excluding exports, measured in TRY).
Total Exports	Numerical	Export revenue reported in USD within administrative records.
R&D Sales	Numerical	Domestic revenue derived from R&D based activities (TRY).
R&D Exports	Numerical	Export revenue from R&D based products or services (USD).
R&D Expenditures	Numerical	Total annual R&D investment (TRY).
Public R&D Support Amount	Numerical	Total public R&D grant support received (TRY).
Ongoing Projects	Numerical	Number of ongoing R&D or innovation projects.
Completed Projects	Numerical	Number of completed R&D or innovation projects.
IPR Count (FSMH)	Numerical	Total number of registered and applied-for intellectual property rights.
Tax and Social Security Incentives	Numerical	Total tax exemption and Social Security premium support amount (TRY).
Active (Target Variable)	Binary	1 = Active within TDZ, 0 = Inactive or exited.

Note: *For active firms, this variable represents a right-censored observation, indicating their duration within the technopark up to the dataset's end date (31.12.2024).

Table 3 presents the descriptive statistics for the 17 variables included in the empirical analysis. The dataset covers academic spin-offs operating in Turkish Technology Development Zones between 2021 and 2024. The numerical variables display substantial variation, reflecting differences in firm capacity, employment structure and financial performance across ASOs. The average Firm Age of approximately 4.8 years indicates that the ASO population is predominantly young and consistent with the early-stage nature of academic entrepreneurship in Türkiye. Financial variables such as total sales, exports, R&D sales, R&D expenditures, public support amounts and tax incentives exhibit highly skewed distributions, with very large maximum values driven by a small number of exceptionally high-performing firms. Overall, these descriptive patterns highlight the structural diversity of Türkiye's academic spin-off ecosystem and support the use of machine learning models capable of capturing nonlinear, heterogeneous and interaction-driven relationships.

Table 3. Descriptive statistics of numerical variables

Variable	*Mean*	*Std. Dev.*	*Min*	*25%*	*Median*	*75%*	*Max*
Total Sales	6.19M	81.85M	0	0.91K	169.58K	1.36M	6.22B
Total Exports	0.15M	2.31M	0	0	0	0	132M
R&D Sales	3.03M	21.97M	0	0	1K	412K	1.08B
R&D Exports	0.11M	1.64M	0	0	0	0	85.82M
R&D Expenditures	3.45M	38.26M	0	39.28K	179.18K	895.89K	2.76B
Public R&D Support	0.49M	7.03M	0	0	0	159.99K	496.78M
Tax & Social Security Incentives	2.56M	121.94M	0	0	18.28K	221.07K	10.85B

Figures 3 and 4 present the sectoral and scale distributions of academic spin-offs in the dataset. Figure 3 shows the distribution of ASOs across the five most common sectors, highlighting the dominance of software and ICT related activities and the comparatively limited presence of firms operating in engineering intensive and manufacturing oriented domains. This pattern is consistent with the broader structure of the Turkish TDZ ecosystem and provides contextual background for interpreting firm level performance outcomes. Figure 4 displays the size distribution of ASOs, indicating that the vast majority operate at the micro scale, with only a small number classified as small, medium or large enterprises. Together, the sectoral and scale distributions help explain the heterogeneity observed in the numerical indicators reported in Table 3 and offer important context for the modelling framework used in the subsequent analysis.

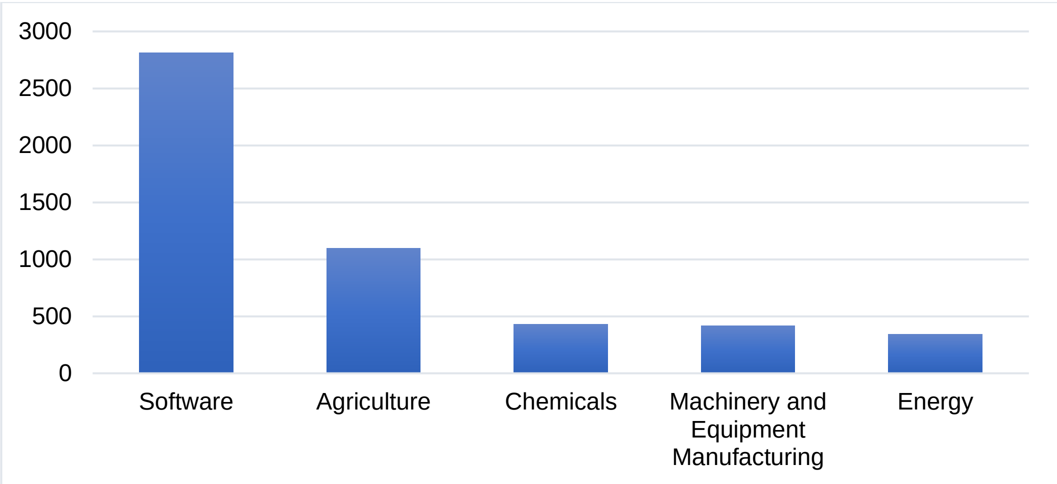

Figure 3. Sectoral distribution (top five sectors) of academic spin-offs in Türkiye

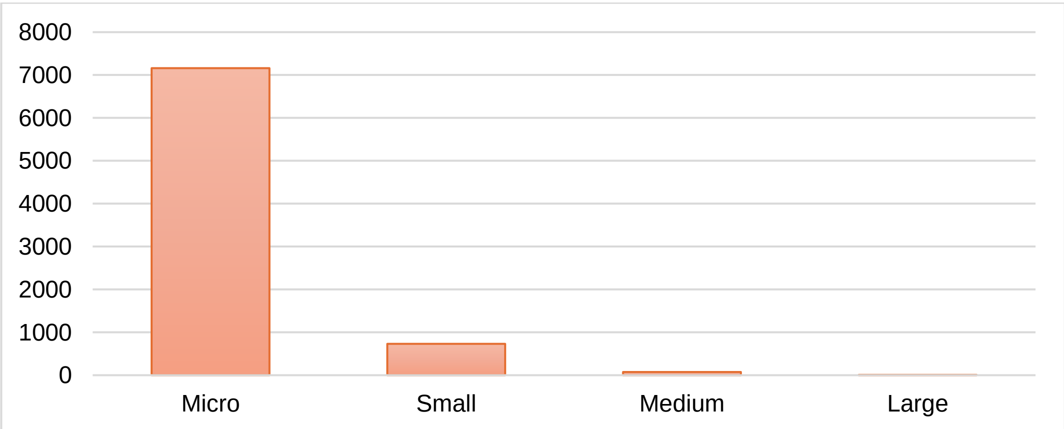

Figure 4. Distribution of academic spin-offs by firm scale in Türkiye

3.2 Feature Selection

The firm-level dataset used in this study comprises financial, R&D-related and project-based indicators that are conceptually and institutionally interconnected. Variables such as total sales, exports, R&D sales, R&D expenditures, public support amounts, tax incentives and project counts reflect overlapping dimensions of firm activity within Technology Development Zones. To illustrate this structure, Figure 5 presents a descriptive correlation matrix of firm-level numerical variables, indicating moderate to high pairwise correlations among several indicators. In this context, feature selection is employed as a precautionary modelling strategy to reduce redundancy among predictors and to obtain more stable and interpretable importance rankings. More importantly, feature selection enhances policy relevance by allowing the analysis to focus on a parsimonious set of actionable mechanisms—such as project continuity, R&D intensity and incubation support—rather than distributing explanatory weight across a large set of closely related indicators.

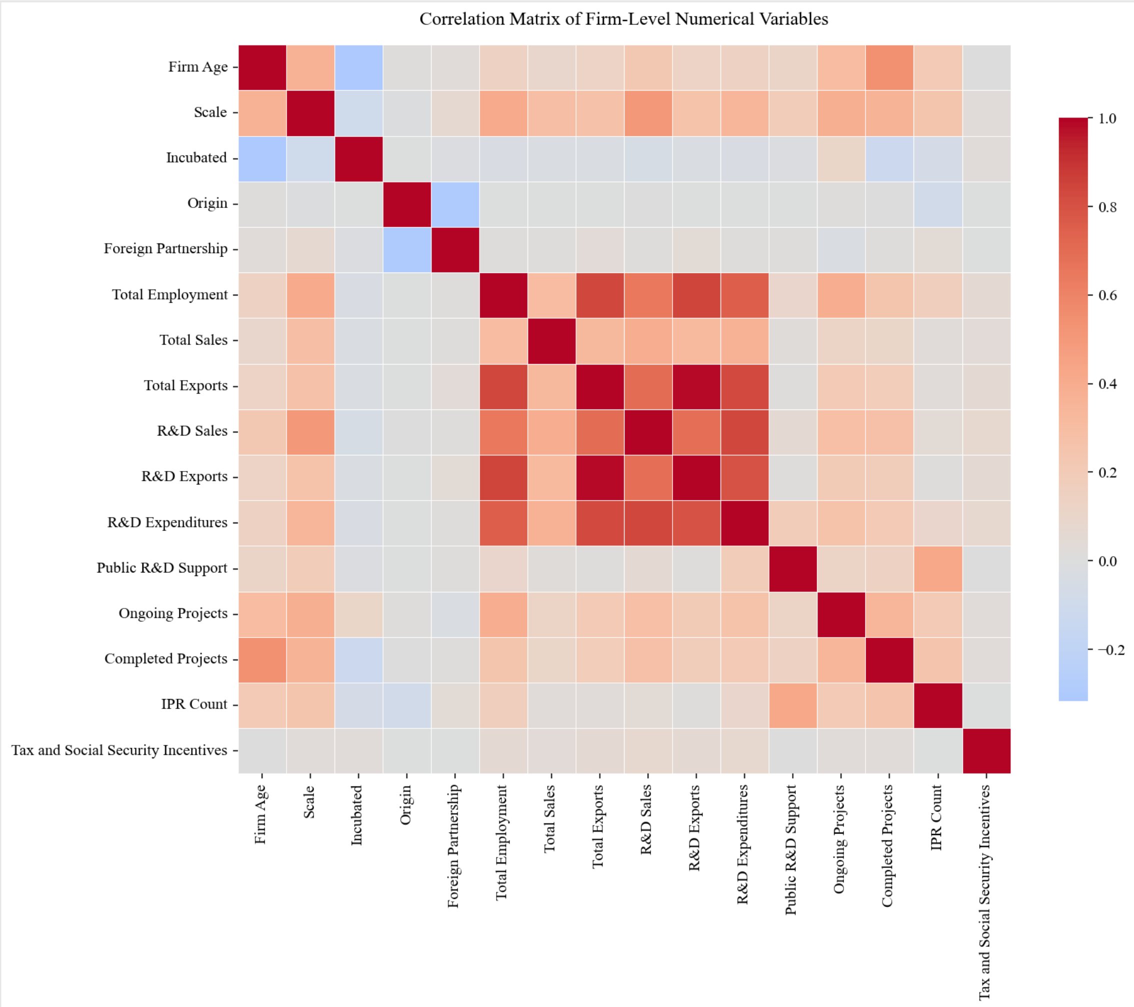

Figure 5. Correlation matrix of firm-level numerical variables

To identify the most informative predictors and reduce model complexity, four supervised feature selection methods are implemented: Recursive Feature Elimination (RFE), Mutual Information (MI), Random Forest importance and a Genetic Algorithm (GA) wrapper. These methods represent three complementary families of feature selection strategies (filter, embedded and wrapper approaches), a distinction widely adopted in the literature to improve the robustness of variable relevance assessment (Guyon and Elisseeff, 2003). Employing multiple approaches allows for methodological triangulation and strengthens the identification of predictors associated with ASO survival.

In addition to comparing feature selection methods, hyperparameter tuning was performed for the final Random Forest classifier trained on the optimal feature subset identified by Recursive Feature Elimination (RFE). The number of trees (n_estimators) and maximum tree depth (max_depth) were tuned using GridSearchCV over a limited grid (n_estimators = {300, 500}; max_depth = {5, 10, None}). To prevent overfitting and information leakage, hyperparameter optimisation was nested within the outer five-fold stratified cross-validation procedure. For each outer training fold, grid search was conducted using an inner three-fold cross-validation, while the corresponding test fold remained completely unseen during optimisation. All other preprocessing steps and feature selection procedures were applied using fixed settings as described above.

3.2.1 Recursive Feature Elimination with Random Forest

Recursive Feature Elimination (RFE) iteratively trains a model, ranks features according to an importance metric and removes the least influential predictors until a reduced subset is obtained. The method was originally introduced for gene selection in cancer classification, where it produced compact and discriminative feature sets (Guyon et al., 2002). Subsequent work shows that RFE can be combined with various learning algorithms, including tree-based models, improving robustness in settings characterised by nonlinear relationships and interaction effects (Darst et al., 2018). In this study, RFE is applied using a Random Forest classifier. Feature importance is computed using the mean decrease in impurity, and at each iteration approximately ten per cent of the lowest-ranked predictors are removed. Although RFE itself is model-agnostic, the use of a tree ensemble helps capture complex decision structures more effectively than linear models, a point widely discussed in the machine learning literature (Hastie et al., 2009: 587-602).

3.2.2 Mutual Information Based Feature Selection

Mutual Information (MI) quantifies the reduction in uncertainty about one variable gained from observing another and detects both linear and nonlinear associations between features and the target variable (Cover and Thomas, 2006: 19-30). As a filter method, MI has been widely applied in economic, biomedical and business analytics contexts to uncover complex dependency structures in high-dimensional data (Vergara and Estévez, 2014). In this study, features with MI scores above the global mean MI value are retained. This threshold is computationally efficient and produces an interpretable subset without relying on model-specific assumptions.

3.2.3 Random Forest Importance Based Selection

Random Forests, introduced by Breiman (2001), construct multiple decision trees on bootstrapped samples and aggregate their predictions. In addition to strong predictive performance, Random Forests provide internal measures of variable importance. The most common metric, Mean Decrease in Impurity, aggregates the impurity reductions attributable to each feature across all splits and trees. In this study, a Random Forest is trained on the full set of predictors, and features with importance values below the mean importance score are removed. This threshold acts as a pragmatic mechanism to manage model complexity while retaining variables that meaningfully contribute to impurity reduction. Ensemble-based importance measures naturally capture nonlinearities and interaction effects, which is particularly relevant in firm-level datasets where relationships are rarely linear (Biau and Scornet, 2016).

3.2.4 Genetic Algorithm Based Wrapper Selection

Genetic Algorithms (GAs) are population-based optimisation heuristics inspired by biological evolution and widely used for feature selection due to their ability to explore complex combinatorial search spaces (Holland, 1975; Siedlecki and Sklansky, 1989; Xue et al., 2016). Here, a binary-coded GA is applied as a wrapper around a Random Forest classifier. Each chromosome represents a candidate feature subset, with 1 indicating inclusion and 0 exclusion. For each chromosome, a Random Forest is trained on a stratified 70–30 train–validation split of the preprocessed data, and the fitness value is defined as the ROC AUC on the validation set. The GA starts with a randomly generated population and iteratively applies selection, one-point crossover and bit-flip mutation over 15 generations, ensuring that at least one feature is always selected. Chromosomes with higher validation ROC AUC values are more likely to be chosen for reproduction. The best-performing chromosome observed during the evolutionary process is retained as the final feature subset. This approach leverages the global search ability of evolutionary optimisation and the nonlinear modelling capacity of Random Forests, enabling the discovery of feature interactions that may not be identified through greedy or univariate selection methods (Chuang et al., 2011; Xue et al., 2016).

3.2.5 Evaluation of Feature Selection Methods

Each feature selection method is implemented within a pipeline in which preprocessing and SMOTE oversampling are applied only within the training folds to avoid information leakage. The resulting feature subsets are evaluated using a Random Forest classifier. Performance is assessed through five-fold stratified cross-validation on the full dataset. The main evaluation metric is the mean ROC-AUC, complemented by Accuracy, F1-score, Precision and Recall to capture multiple dimensions of classification quality, especially for the minority (inactive) class.

As reported in Table 4, all four methods achieve very high discriminative performance (ROC-AUC ≥ 0.990), indicating that the observed firm-level indicators collectively contain strong information on ASO survival. Among them, RFE attains the highest ROC-AUC (0.9958) and Accuracy (0.9837), followed closely by Mutual Information (ROC-AUC = 0.9956; Accuracy = 0.9831) and the Genetic Algorithm (ROC-AUC = 0.9951; Accuracy = 0.9826). The Random Forest importance method performs slightly lower (ROC-AUC = 0.9907; Accuracy = 0.9775) yet still demonstrates strong predictive capability. Given its superior ROC-AUC and Accuracy, RFE is selected as the optimal feature subset for training the final Random Forest model and for the subsequent interpretability analysis.

Table 4. Cross-validated performance of feature selection methods

Method	*Accuracy*	*F1 Score*	*ROC-AUC*	*Precision*	*Recall*
RFE	0.9837	0.9862	0.9958	0.997	0.9757
Mutual Information	0.9831	0.9857	0.9956	0.9979	0.9738
Random Forest Importance	0.9775	0.981	0.9907	0.9914	0.9709
Genetic Algorithm (GA)	0.9826	0.9853	0.9951	0.9962	0.9746

Figure 6 presents the ROC–AUC curve of the RFE feature selection method, which achieved the highest predictive accuracy, illustrating the classification performance of the proposed model. The curve demonstrates a high predictive capability, with an AUC value of 0.996, indicating that the selected features effectively capture the critical information required to predict firm survival in Technology Development Zones. The relatively high AUC further suggests that firm survival aligns with structural characteristics closely linked to institutional, financial, and operational factors.

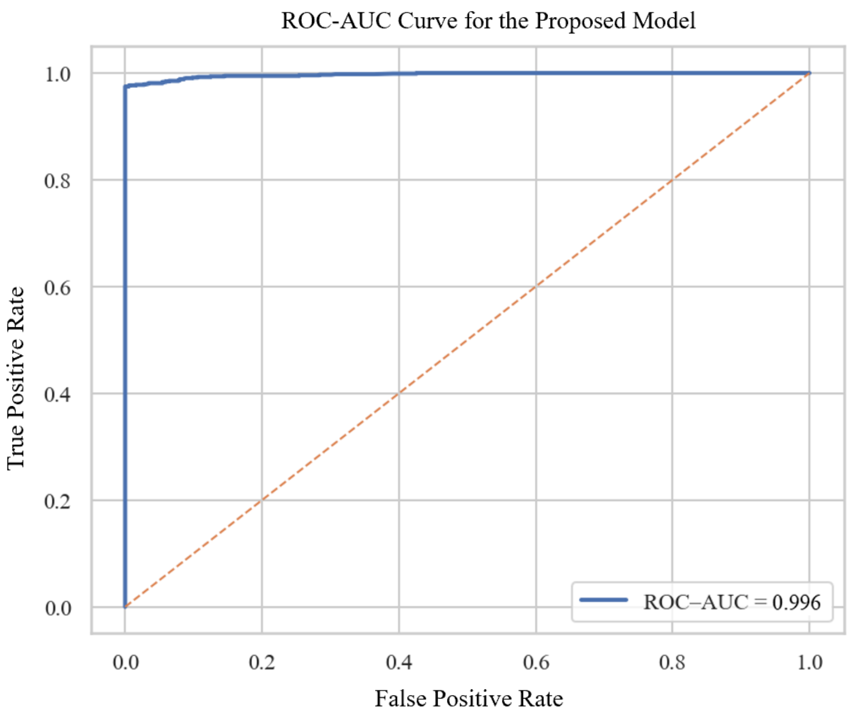

Figure 6. ROC-AUC Curve of the Proposed Model

4. RESULTS and DISCUSSION

4.1 Descriptive Overview and Class Balancing

Before modelling, the dataset exhibited a pronounced imbalance between active and inactive academic spin-offs, with active firms forming the majority class. To reduce bias during training, the Synthetic Minority Over-Sampling Technique (SMOTE) was applied exclusively within the training folds. This procedure generated synthetic minority observations while preserving the integrity of the untouched test folds. Visual comparison of pre- and post-SMOTE class distributions confirms that oversampling effectively mitigates imbalance effects and improves the model’s sensitivity to the minority group of inactive firms.

4.2 Performance Comparison of Feature-Selection Methods

Table 4 presents the comparative performance of the four supervised feature-selection approaches evaluated using a Random Forest classifier. All methods achieve high predictive accuracy (Accuracy ≥ 0.977), indicating that the explanatory variables collectively capture the structural determinants of firm survival. RFE delivers the strongest performance, achieving the highest Accuracy (0.9837) and ROC-AUC (0.9958), followed closely by Mutual Information (Accuracy = 0.9831; ROC-AUC = 0.9956) and the Genetic Algorithm (Accuracy = 0.9826; ROC-AUC = 0.9951). The Random Forest Importance method performs slightly lower (Accuracy = 0.9775; ROC-AUC = 0.9907) yet still demonstrates robust predictive capability. These findings align with established evidence that wrapper and hybrid strategies often outperform simple univariate filters by capturing complex feature interactions and model-dependent relationships. Given its superior performance, RFE is selected as the optimal feature subset for training the final Random Forest model and for conducting SHAP-based interpretability analysis.

4.3 Feature-Selection Results and Theoretical Interpretation

Table 5 summarises the features selected by each of the four supervised feature-selection methods. Presenting the results side by side allows a clear comparison of which variables are consistently identified across different algorithms and which ones appear only under specific methods. This comparative structure strengthens the transparency and robustness of the modelling approach. Based on these results, the RFE subset—which achieved the strongest predictive performance—is used for training the final Random Forest model and for conducting the SHAP-based interpretation.

Table 5. RFE-selected features used in the final model

Feature	*RFE*	*Mutual Info*	*Random Forest Importance*	*Genetic Algorithm*
Firm Age	✓	✓	✓	✓
Ongoing Projects	✓	–	–	–
Completed Projects	✓	–	–	–
Incubated	✓	–	✓	✓
Total Sales	✓	✓	✓	✓
R&D Sales	✓	✓	✓	✓
R&D Expenditures	✓	✓	✓	–
Public R&D Support	✓	✓	✓	✓
Tax & Social Security Incentives	✓	✓	✓	–
Scale (Medium, Large)	–	–	–	✓
Export Revenue	–	–	–	✓
Origin	–	–	–	✓
Foreign Partnership	–	–	–	✓

4.4 SHAP-Based Model Interpretation

Figure 7 presents the SHAP beeswarm plot derived from the final Random Forest model trained on the RFE-selected feature set. The ranking of mean absolute SHAP values shows that the most influential predictors of academic spin-off survival are Ongoing Projects, Incubated, R&D Expenditures, Firm Age, Total Sales, Public R&D Support, Tax and Social Security Incentives, Completed Projects and R&D Sales.

Ongoing Projects emerges as the strongest contributor because firms with a greater number of active R&D projects exhibit significantly higher survival probabilities. This reflects both the organisational benefits of maintaining continuous project activity, such as sustained learning, technological capability and improved access to external funding, and the practical requirement in TDZs that firms must maintain an active project to remain in the zone. Incubated also shows a strong positive effect. Firms operating as incubation firms typically benefit from mentoring, subsidised workspace and administrative and technical support, which together strengthen early-stage resilience. R&D Expenditures and Firm Age further improve survival prospects. Higher R&D spending signals long-term technological commitment, and older firms benefit from accumulated experience, routines and legitimacy.

Financial indicators such as Total Sales, Public R&D Support and Tax and Social Security Incentives contribute by reducing liquidity pressures and enabling firms to sustain their investment in human capital and R&D activities. Completed Projects and R&D Sales display more moderate but still meaningful effects, suggesting that past innovation outputs and the commercialisation of research activities support ongoing operations, although their influence is weaker than that of active project intensity. The distribution of SHAP values shows that most firms cluster around moderate predicted survival effects. A smaller subset demonstrates strongly positive or negative impacts depending on project volume, R&D capability and the use of institutional support mechanisms. Overall, the SHAP analysis reveals a consistent pattern in which project continuity, R&D intensity and policy-based financial support jointly underpin the survival of academic spin-offs in Turkish Technoparks.

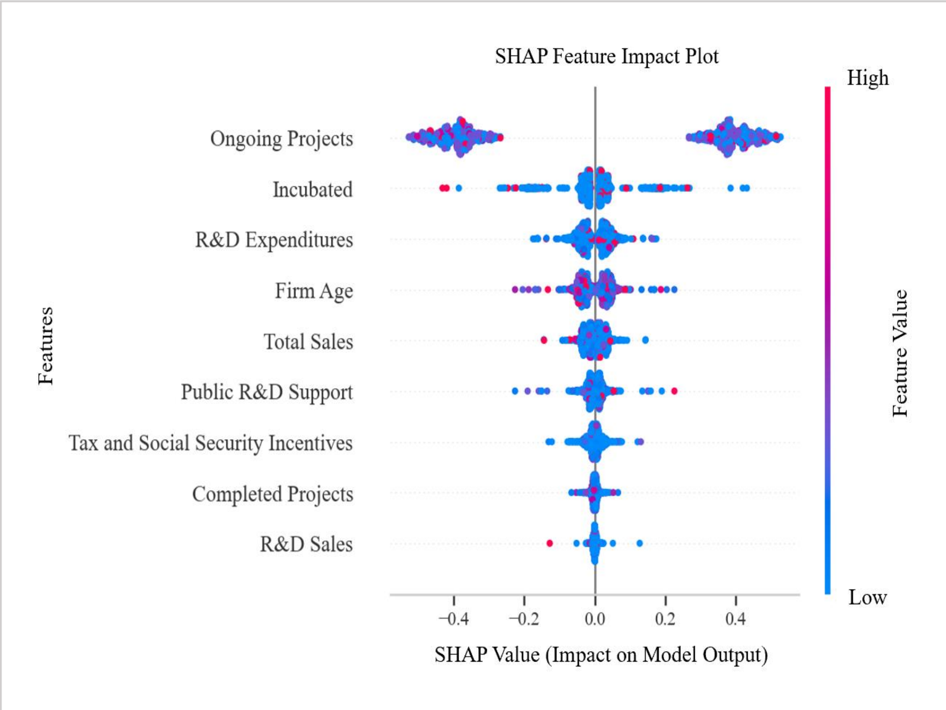

Figure 7. SHAP Feature Importance for Academic Spin-Off Survival

It is important to note that SHAP values reflect conditional associations learned by the model rather than causal effects. In the context of Turkish Technology Development Zones, several influential predictors, such as ongoing projects, incubation status and eligibility for public support, are closely linked to regulatory compliance and institutional design rather than purely discretionary firm behaviour. Accordingly, SHAP-based interpretations should not be treated as direct policy levers, but rather as indicators of institutional and organisational mechanisms associated with firm persistence within the TDZ framework. From a policy perspective, these results highlight the structural conditions under which academic spin-offs are more likely to remain active, rather than prescribing isolated interventions.

4.5 Robustness Checks: Class Imbalance and Minority-Class Performance

Building on the preprocessing strategy outlined in Section 3.1 and the initial class balancing described in Section 4.1, additional analyses were conducted to assess the robustness of the proposed modelling framework under class imbalance by comparing alternative imbalance-handling strategies. Specifically, SMOTE, SMOTE-Tomek and random undersampling were considered, as these methods represent conceptually distinct and widely used resampling paradigms in the imbalanced learning literature (Chawla et al., 2002; Batista et al., 2004; He and Garcia, 2009). Previous studies emphasise that classification performance under imbalance is influenced not only by class proportions but also by factors such as minority-class sample size and class overlap, implying that no single resampling strategy is universally optimal (Batista et al., 2004). Accordingly, this design allows an assessment of whether the results are sensitive to different imbalance-handling philosophies rather than to a specific technique.

In all cases, resampling was applied exclusively within the training folds of a nested cross-validation framework, while model evaluation was performed on the original, non-synthetic test folds to avoid information leakage. Table 6 reports the outer-fold average performance of the Random Forest model trained on the RFE-selected feature set under each resampling strategy. In addition to overall performance metrics (Accuracy, F1-score, ROC-AUC, Precision and Recall), minority-class performance for inactive firms is explicitly evaluated using class-specific Precision, Recall and F1-score.

Across all three imbalance-handling approaches, overall predictive performance remains highly stable, with Accuracy values around 0.98 and ROC-AUC values exceeding 0.995. Importantly, minority-class performance also exhibits a high degree of consistency. Recall values for inactive firms range between 0.991 and 0.996, while minority-class F1-scores remain close to 0.98 across all sampling strategies. These results indicate that the model effectively captures exit dynamics and identifies inactive firms with high reliability.

The similarity of both overall and minority-class performance across SMOTE, SMOTE-Tomek and undersampling suggests that the findings are not sensitive to the choice of imbalance-handling method. Rather than serving to inflate predictive accuracy through synthetic observations, resampling functions as a stabilising mechanism during training by mitigating majority-class dominance and supporting balanced learning. Overall, these robustness checks provide strong evidence that the identified determinants of academic spin-off survival reflect genuine structural relationships inherent in the data rather than artefacts of a specific resampling technique.

Table 6. Robustness analysis of class imbalance handling strategies: Overall and minority-class performance (outer-fold averages)

Sampler	*Accuracy*	*F1*	*ROC-AUC*	*Precision*	*Recall*	*Precision (Inactive)*	*Recall (Inactive)*	*F1 (Inactive)*
SMOTE	0.9829	0.9856	0.9959	0.996	0.9755	0.9647	0.9941	0.9791
SMOTE-Tomek	0.9822	0.985	0.9951	0.9938	0.9763	0.9657	0.9909	0.9781
Random Under-Sampling	0.9829	0.9856	0.9955	0.997	0.9744	0.9633	0.9956	0.9791

Notes: Performance metrics are averaged over the outer folds of a nested cross-validation procedure. Resampling methods are applied exclusively during training. Minority-class metrics refer to inactive firms and are computed on the original, non-synthetic test folds.

4.6 Discussion

The findings of this study provide clear evidence that academic spin-off survival in Turkish Technology Development Zones is shaped by a coherent set of mechanisms centred on project continuity, internal R&D capacity and institutional support structures. The consistent dominance of project-related and R&D-intensity variables across all feature-selection algorithms and SHAP analyses demonstrates that sustained engagement in research activities is the primary engine of firm persistence. This pattern aligns with previous work highlighting absorptive capacity, ongoing innovation and funding intensity as critical conditions for survival within knowledge-intensive environments (Rodeiro-Pazos, 2021; Soetanto and Jack, 2016). Firms with multiple ongoing projects appear particularly resilient, suggesting that the ability to maintain a continuous project pipeline helps shield academic spin-offs from technological stagnation, financial volatility and early-stage vulnerability.

A second important finding is the strong positive effect of incubation. The SHAP results indicate that incubated firms exhibit substantially higher survival prospects, consistent with international evidence showing that incubation programmes reduce early-stage liabilities by providing structured mentorship, subsidised infrastructure and administrative and technical support (Soetanto and van Geenhuizen, 2019). In the context of Türkiye's TDZs, incubation may help mitigate capability gaps and market uncertainties that typically challenge young technology ventures.

Moreover, the influence of fiscal and R&D-related incentives shows that Türkiye's policy architecture under Law No. 4691 contributes meaningfully to firm continuity. Tax exemptions, social-security incentives and public R&D support consistently display positive contributions to survival outcomes. This suggests that the incentive structure effectively eases liquidity constraints and enables sustained investment in R&D and human capital. Although structural characteristics such as scale, export orientation and intellectual property outputs are less dominant than project-based variables, their selection in the broader Genetic Algorithm subset implies that market positioning, internationalisation and intangible assets complement internal R&D capacity, particularly as firms mature. Overall, the results point to a multidimensional survival mechanism in which project continuity, organisational capability and policy-driven institutional support jointly sustain the persistence of academic spin-offs in Turkish technoparks.

From a methodological perspective, the robustness analyses further reinforce the credibility of these findings. The stability of both overall performance and minority-class performance across different imbalance-handling strategies indicates that the identified relationships are not sensitive to the specific resampling approach employed. In particular, the consistently high recall and F1-scores for inactive firms show that the model effectively captures exit dynamics without artificially inflating predictive accuracy through synthetic observations. This methodological robustness supports the interpretation that the determinants of academic spin-off survival identified in this study reflect genuine structural patterns inherent in firm behavior and the institutional design of Turkish Technology Development Zones.

While the models are evaluated using five-fold stratified cross-validation to enhance robustness, the findings should be interpreted within the institutional context of Turkish Technology Development Zones. The consistency of results across multiple feature-selection methods and performance metrics suggests that the identified determinants are not sensitive to alternative train–test splits. Although the model is expected to generalise to future cohorts of firms operating under similar regulatory and institutional conditions, its direct applicability to different national or policy contexts may be limited. Future research could assess external validity by applying the framework to other technopark systems or by using temporally separated training and test cohorts.

5. CONCLUSION and POLICY IMPLICATIONS

This study employed multiple machine-learning-based feature-selection techniques within a robust, cross-validated modelling framework to identify the main factors influencing the survival of academic spin-offs in Turkish Technology Development Zones. Combining Recursive Feature Elimination, Mutual Information, Random Forest importance and a Genetic Algorithm within a cross-validated Random Forest model produced a concise but highly informative set of predictors, whose relevance remained stable across alternative imbalance-handling strategies. These results provide a consistent empirical structure that aligns with theoretical expectations from the resource-based view and absorptive capacity perspectives, both of which emphasise the role of sustained learning, technological capability and resource endowments in firm persistence.

The findings show that survival is driven by four main mechanisms. The first is project continuity. Firms that maintain a steady flow of R&D projects are more likely to survive because continuous activity supports technological capability, access to external funding and organisational learning. This mechanism is also structurally reinforced by TDZ regulations, indicating that policy design and organisational behaviour operate jointly. The second mechanism is internal R&D intensity. Higher R&D expenditures and the commercialisation of research outputs strengthen firms' learning capacity and improve resilience, reflecting both capability accumulation and long-term strategic commitment. The third mechanism is incubation. Incubated firms show clearly higher survival rates, highlighting the importance of mentoring, infrastructural support and administrative facilitation in reducing early-stage uncertainty. The fourth mechanism is financial incentives. Public R&D funding, tax exemptions and social-security incentives reduce financial pressure and help firms sustain investment in human capital and technology, functioning as external stabilisers during vulnerable phases of venture development.

These mechanisms jointly suggest that academic spin-off survival depends not only on firm-level capabilities but also on the institutional environment created by TDZ policies. Strengthening project continuity, expanding incubation capacity and targeting fiscal incentives according to firm maturity can support more stable growth. Encouraging export-oriented innovation and internationalisation may further increase resilience, particularly for firms transitioning beyond the start-up phase. In addition, the observed nonlinear relationships suggest the presence of threshold effects, implying that policy interventions may yield the greatest impact when they are designed to push firms beyond critical capability or investment levels rather than relying on homogeneous support schemes.

This study has several limitations. Our measure of survival relies exclusively on TDZ administrative records, and therefore does not capture firms that continue operating outside the zones. The dataset covers the period from 2021 to 2024 and therefore does not reflect long-term dynamics or delayed policy effects. The model describes statistical associations in the data but does not identify causal relationships. Future studies could use longer observation windows, link TDZ records to external datasets such as patent, export or investment data and compare different groups of academic entrepreneurs or firms inside and outside TDZs to obtain a deeper understanding of survival pathways and the influence of policy interventions. Expanding the methodology toward causal machine-learning or survival-analysis frameworks could also provide more robust insight into underlying mechanisms.

Acknowledgements
I would like to express my sincere gratitude to the Scientific and Technological Research Council of Türkiye (TÜBİTAK) for the support provided through the 2214-A International Research Fellowship Program for Ph.D. Students and the 2211-C Domestic Priority Areas Ph.D. Scholarship Program, which offered valuable opportunities that contributed to the advancement of my doctoral research. I am also deeply grateful to my advisor, Mehmet Yılmaz, for his invaluable guidance, continuous support, and encouragement throughout my Ph.D. studies.

Author Contributions

Başak Apaydın Avşar: Literature Review, Conceptualization, Methodology, Data Curation, Analysis, Writing-original draft Writing-review and editing *Mehmet Yılmaz*: Conceptualization, Methodology, Modelling, Analysis, Validation, Writing – review and editing

Conflict of Interest

No potential conflict of interest was declared by the authors.

Funding

Any specific grant has not been received from funding agencies in the public, commercial, or not-for-profit sectors.

Compliance with Ethical Standards

It was declared by the authors that the tools and methods used in the study do not require the permission of the Ethics Committee.

Ethical Statement

It was declared by the authors that scientific and ethical principles have been followed in this study and all the sources used have been properly cited.

REFERENCES

Batista, G.E.A.P.A., Prati, R.C. and Monard, M.C. (2004). "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data", *ACM SIGKDD Explorations Newsletter*, 6(1), 20-29. https://doi.org/10.1145/1007730.1007735

Bercovitz, J. and Feldman, M. (2008). "Academic Entrepreneurs: Organizational Change at the Individual Level", *Organization Science*, 19(1), 69-89. https://doi.org/10.1287/orsc.1070.0295

Biau, G. and Scornet, E. (2016). "A Random Forest Guided Tour", *Test*, 25(2), 197-227. https://doi.org/10.1007/s11749-016-0481-7

Branco, P., Torgo, L. and Ribeiro, R.P. (2016). "A Survey of Predictive Modelling Under Imbalanced Distributions", *ACM Computing Surveys*, 49(2), 1-50. https://doi.org/10.1145/2907070

Breiman, L. (2001). "Random Forests", *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002). "SMOTE: Synthetic Minority Over-Sampling Technique", *Journal of Artificial Intelligence Research*, 16, 321-357. https://doi.org/10.1613/jair.953

Chuang, L.Y., Tsai, S.-W. and Yang, C.-H. (2011). "Chaotic Binary Genetic Algorithm for Feature Selection", *Expert Systems with Applications*, 38(10), 13367-13377. https://doi.org/10.1016/j.eswa.2011.04.147

Civera, A., Meoli, M. and Vismara, S. (2019). "Do Academic Spinoffs Internationalize?", *Journal of Technology Transfer*, 44(2), 381-403. https://doi.org/10.1007/s10961-018-9683-3

Civera, A., Meoli, M. and Vismara, S. (2020). "Engagement of Academics in University Technology Transfer: Opportunity and Necessity Academic Entrepreneurship", *European Economic Review*, 123, 103376. https://doi.org/10.1016/j.euroecorev.2020.103376

Cover, T.M. and Thomas, J.A. (2006). “Elements of Information Theory”, (2nd ed.), Wiley.

Criaco, G., Minola, T., Migliorini, P. and Serarols-Tarrés, C. (2014). "To Have and Have Not: Founders’ Human Capital and University Start-Up Survival", *Journal of Technology Transfer*, 39(4), 567-593. https://doi.org/10.1007/s10961-013-9312-0

Czarnitzki, D., Rammer, C. and Toole, A.A. (2014). "University Spin-Offs and the Performance Premium", *Small Business Economics*, 43(2), 309-326. https://doi.org/10.1007/s11187-013-9538-0

Darst, R.P., Malecki, K.M.C. and Engelman, C.D. (2018). "Using Recursive Feature Elimination in Random Forest to Account for Correlated Variables in High-Dimensional Data", *BMC Research Notes*, 11(1), 364. https://doi.org/10.1186/s13104-018-3593-1

Djokovic, D. and Souitaris, V. (2008). "Spinouts from Academic Institutions: A Literature Review with Suggestions for Further Research", *Journal of Technology Transfer*, 33(3), 225-247. https://doi.org/10.1007/s10961-006-9000-4

Etzkowitz, H. (2003). "Research Groups as Quasi-Firms: The Invention of the Entrepreneurial University", *Research Policy*, 32(1), 109-121. https://doi.org/10.1016/S0048-7333(02)00009-4

Fackler, D., Schnabel, C. and Schmucker, A. (2016). "Spinoffs in Germany: Characteristics, Survival, and the Role of Their Parents", *Small Business Economics*, 46(1), 93-114. https://doi.org/10.1007/s11187-015-9673-x

Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B. and Herrera, F. (2018). “Learning from Imbalanced Data Sets”, Springer.

Géron, A. (2022). “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow”, O’Reilly.

Guyon, I. and Elisseeff, A. (2003). "An Introduction to Variable and Feature Selection", *Journal of Machine Learning Research*, 3, 1157-1182.

Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. (2002). "Gene Selection for Cancer Classification Using Support Vector Machines", *Machine Learning*, 46(1-3), 389-422. https://doi.org/10.1023/A:1012487302797

Han, J., Kamber, M. and Pei, J. (2012). “Data Mining: Concepts and Techniques”, (3rd ed.), Morgan Kaufmann.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). “The Elements of Statistical Learning”, Springer.

He, H. and Garcia, E.A. (2009). "Learning from Imbalanced Data", *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284. https://doi.org/10.1109/TKDE.2008.239

Holland, J.H. (1975). “Adaptation in Natural and Artificial Systems”, University of Michigan Press.

Hossinger, S.M., Chen, X. and Werner, A. (2020). "Drivers, Barriers and Success Factors of Academic Spin-Offs: A Systematic Literature Review", *Management Review Quarterly*, 70(1), 97-134. https://doi.org/10.1007/s11301-019-00161-w

Jain, S., George, G. and Maltarich, M. (2009). "Academics or Entrepreneurs? Investigating Role Identity Modification of University Scientists Involved in Commercialization Activity", *Research Policy*, 38(6), 922-935. https://doi.org/10.1016/j.respol.2009.02.007

Kuhn, M. and Johnson, K. (2019). "Applied Predictive Modeling", Springer.

Muscio, A., Quaglione, D. and Ramaciotti, L. (2016). "The Effects of University Rules on Spinoff Creation: The Case of Academia in Italy", *Research Policy*, 45(7), 1386-1396. https://doi.org/10.1016/j.respol.2016.04.011

Official Gazette. (2001). "Law No. 4691 on Technology Development Zones, Number: 24454", https://www.resmigazete.gov.tr/eskiler/2001/07/20010706.htm, (Accessed: 11.09.2025).

Official Gazette. (2008). "Law No. 5746 on Supporting R&D and Design Activities, Number: 26814", https://www.resmigazete.gov.tr/eskiler/2008/03/20080312.htm, (Accessed: 11.09.2025).

O'Shea, R.P., Allen, T.J., Chevalier, A. and Roche, F. (2005). "Entrepreneurial Orientation, Technology Transfer and Spinoff Performance of U.S. Universities", *Research Policy*, 34(7), 994-1009. https://doi.org/10.1016/j.respol.2005.05.011

Powers, J.B. and McDougall, P.P. (2005). "University Start-Up Formation and Technology Licensing with Firms That Go Public", *Journal of Business Venturing*, 20(3), 291-311. https://doi.org/10.1016/j.jbusvent.2003.12.008

Prokop, D. (2023). "The Academic Spinoff Theory of the Firm", *International Journal of Entrepreneurship and Innovation*, 24(4), 233-243. https://doi.org/10.1177/146575032110660

Republic of Türkiye Ministry of Industry and Technology. (2025). "Technology Development Zones Statistics", https://sanayi.gov.tr, (Accessed: 06.09.2025).

Rodríguez-Gulías, M.J., Fernández-López, S. and Rodeiro-Pazos, D. (2016). "Growth Determinants in Entrepreneurship: A Longitudinal Study of Spanish Technology-Based University Spin-Offs", *Journal of International Entrepreneurship*, 14(3), 323-344. https://doi.org/10.1007/s10843-016-0185-9

Rodríguez-Gulías, M.J., Rodeiro-Pazos, D. and Fernández-López, S. (2017). "The Effect of University and Regional Knowledge Spillovers on Firms' Performance: An Analysis of the Spanish USOs", *International Entrepreneurship and Management Journal*, 13(1), 191-209. https://doi.org/10.1007/s11365-016-0399-2

Rodeiro-Pazos, D. (2021). "Size and Survival: An Analysis of University Spin-Offs", *Technological Forecasting and Social Change*, 171, 120953. https://doi.org/10.1016/j.techfore.2021.120953

Siedlecki, W. and Sklansky, J. (1989). "A Note on Genetic Algorithms for Large-Scale Feature Selection", *Pattern Recognition Letters*, 10(5), 335-347.

Soetanto, D. and Jack, S. (2016). "The Impact of University-Based Incubation Support on the Innovation Strategy of Academic Spin-Offs", *Technovation*, 25-40. https://doi.org/10.1016/j.technovation.2015.11.001

Soetanto, D. and van Geenhuizen, M. (2019). "Life After Incubation: The Impact of Entrepreneurial Universities on the Long-Term Performance of Their Spin-Offs", *Technological Forecasting and Social Change*, 141, 263-276. https://doi.org/10.1016/j.techfore.2018.10.021

TÜBİTAK. (2025). "1513 Technology Transfer Offices (TTO) Support Programme", https://tubitak.gov.tr/en/funds/sanayi/ulusal-destek-programlari/1513-technology-transfer-office-support-program, (Accessed: 10.09.2025).

Vergara, J.R. and Estévez, P.A. (2014). "A Review of Feature Selection Methods Based on Mutual Information", *Neural Computing and Applications*, 24(1), 175-186. https://doi.org/10.1007/s00521-013-1368-0

Visintin, F. and Pittino, D. (2014). "Founding Team Composition and Early Performance of University-Based Spin-Off Companies", *Technovation*, 34(1), 31-43. https://doi.org/10.1016/j.technovation.2013.09.004

Wang, P., Li, Y. and Reddy, C.K. (2019). "Machine Learning for Survival Analysis: A Survey", *ACM Computing Surveys*, 51(6), 1-36. https://doi.org/10.1145/3214306

Wennberg, K., Wiklund, J. and Wright, M. (2011). "The Effectiveness of University Knowledge Spillovers: Performance Differences Between University Spinoffs and Corporate Spinoffs", *Research Policy*, 40(8), 1128-1143. https://doi.org/10.1016/j.respol.2011.05.014

Xue, B., Zhang, M., Browne, W.N. and Yao, X. (2016). "A Survey on Evolutionary Computation for Feature Selection", *IEEE Transactions on Cybernetics*, 46(3), 239-261. https://doi.org/10.1109/TCYB.2015.2501371

Zhang, J. (2009). "The Performance of University Spin-Offs: An Exploratory Analysis Using Venture Capital Data", *Journal of Technology Transfer*, 34(3), 255-285. https://doi.org/10.1007/s10961-008-9088-9