applied sciences

MDPI

Article

Multidimensional Data Analysis for Enhancing In-Depth Knowledge on the Characteristics of Science and Technology Parks

Olga Francés *, José Abreu-Salas , Javi Fernández, Yoan Gutiérrez and Manuel Palomar

Research Group on Natural Language Processing and Information Systems, Department of Software and Computing Systems, University of Alicante, 03080 Alicante, Spain; ji.abreu@ua.es (J.A.-S.); javifm@ua.es (J.F.); ygutierrez@ua.es (Y.G.); mpalomar@dlsi.ua.es (M.P.)
* Correspondence: olga.frances@ua.es

Featured Application: This work expands the understanding of science and technology parks (STPs) and can also guide STP decision makers by incorporating data science and artificial intelligence tools, as well as guiding policy makers on how to promote innovation.

Abstract: The role played by science and technology parks (STPs) in technology transfer, industrial innovation, and economic growth is examined in this paper. The accurate monitoring of their evolution and impact is hindered by the lack of uniformity in STP models or goals, and the scarcity of high-quality datasets. This work uses existing terminologies, definitions, and core features of STPs to conduct a multidimensional data analysis that explores and evaluates the 21 core features which describe the key internal factors of an STP. The core features are gathered from a reliable and updatable dataset of Spanish STPs. The methodological framework can be replicated for other STP contexts and is based on descriptive techniques and machine-learning tools. The results of the study provide an overview of the general situation of STPs in Spain, validate the existence and characteristics of three types of STPs, and identify the typical features of STPs. Moreover, the prototype STP can be used as a benchmark so that other STPs can identify the features that need to be improved. Finally, this work makes it possible to carry out classifications of STPs, in addition to prediction and decision making for innovation ecosystems.

Keywords: science and technology parks; multidimensional data analysis; core features; dataset; classification; decision making; machine learning

Citation: Francés, O.; Abreu-Salas, J.; Fernández, J.; Gutiérrez, Y.; Palomar, M. Multidimensional Data Analysis for Enhancing In-Depth Knowledge on the Characteristics of Science and Technology Parks. *Appl. Sci.* **2023**, *13*, 12595. https://doi.org/10.3390/app132312595

Academic Editor: José Machado

Received: 31 October 2023
Revised: 13 November 2023
Accepted: 19 November 2023
Published: 22 November 2023

1. Introduction

Science and technology parks (STPs) are key drivers of technology transfer, industrial innovation, economic growth, and international competitiveness. The International Association of Science Parks (IASP) has 350 STP members in 75 countries around the world, although there are around 1300 STPs worldwide, according to Sanz [1]. The precise number may have increased given the global trend in favour of establishing STPs. STPs differ in size, scope, and specialisation, serving as innovation hubs for diverse industries and research areas. The growth of STPs underscores their role in supporting innovation, research, and technology transfer on a global scale.

Despite the growing prevalence of STPs worldwide, challenges persist in terms of the lack of a universally accepted and precise definition of what constitutes an STP. This has led to terminological ambiguities and inconsistent models, which, compounded by inadequate data, hampers the evaluation of an STP's actual impact and effectiveness. These challenges have triggered scholarly debate on the role of STPs as drivers of innovation, economic growth, and local development.

Moreover, the growing availability of data and the advent of artificial intelligence (AI) have introduced new opportunities along with complexities when it comes to assessing and managing STPs. The application of AI tools for data collection, cleaning, and analysis is becoming increasingly vital for understanding the multifaceted nature of these innovation ecosystems. This evolving landscape calls for a systematic approach to redefine and categorise STPs, establish standardised metrics, and harness the power of AI for comprehensive evaluation.

The specific problems detected in the field of STPs are the lack of: (i) critical standards in the definition, classification, and valuation of STPs; (ii) standardised and manageable sources of information in the field of STPs; and (iii) specific technologies for analysing STP ecosystems and enabling informed decision making. Problems (i) and (ii) have been recently addressed in a previous work by the authors [2]. Following on from this line of research, the present study addresses problem (iii) with the development of technologies oriented towards an in-depth analysis and informed decision making for STPs. This problem has not been explored in the literature to date.

Recently, Francés et al. [2] have developed different resources that are of crucial interest for this study, which are summarised next: the definition of three types of STPs; the core features that define the key internal factors that affect STPs; and a methodology to build standardised datasets of STPs applied to the use case of Spanish STPs.

On the one hand, short definitions of the three main STPs models were discussed and provided (Table 1). These definitions were considered for the expert labelling for the case study of Spanish STPs, also described by Francés et al. [2].

Table 1. Type of STPs and their main characteristics [2].

Type of STP	Definition
Science Park (SP)	STP promoted, owned, and managed wholly or partly by universities (or higher education institutions or research institutions) where the university plays a central role in the SP dynamics, and controls and substantially influences the performance of the SP.
Technology Park (TP)	STP where the university does not hold a central role in the dynamics and decision making of the STP; even when the university holds some shares in the STP, it signs "soft" collaboration agreements or shares the location with the STP. In conclusion, control and influence in the STP depend on a driver other than the university.
Hybrid Park (HP)	STP with an equitable and balanced control and influence of both the university and other drivers (typically regional government, private sector, etc.). It is not possible to discern a priori whether the model is closer to science or technology parks.

As explained, slight differences can be noted between the characteristics of an STP, as well as regarding their performance and impact. Francés et al. [2] also analyses the internal and external key factors affecting STPs identified in the literature. Internal factors can be malleable and external factors are totally outside the "control" of the STP decision-making management process. Hence, this work, in line with the dataset presented in [2], focuses on internal factors alone. Nevertheless, the methodology presented will allow the incorporation of as many factors (endogenous and exogenous) as required. Table 2 presents each internal key factor together with their corresponding descriptive core feature(s) that were considered in the study. The selection of the core features describing each factor is the product of consulting several sources including the literature, the authors' managerial experience of STPs, and the availability and access to data. There are a total of 21 core features that are broad and sufficiently representative to describe the 10 internal key factors, and these are presented in Table 2. They can be not only used in the case study of Spanish

STPs, but also applied to other regions, as these features universally describe STP internal factors. Moreover, other complementary features can be added.

Table 2. List of core features associated to each key factor affecting STPs [2].

Internal Key Factor	Core Feature		Definition	Source of Information
F5. STP promotors F6. STP management F7. University/STP interactions	Aggregating feature	Fe01. Type of STP	STP model according to Table 1 and author's expertise. Promotors, STP ownership, management, and university linkage are considered in this classification.	Expert labelling
F8. STP age	Individual features	Fe02. STP age	STP years in operation.	Primary sources
F9. STP size	Individual features	Fe03. STP size	Total surface area of the STP (m^2).	Primary sources
F12. Number of companies	Individual features	Fe04. Companies	Total population of companies in the STP.	Primary sources
F13. Revenue	Individual features	Fe05. Turnover	Total amount of the cost of goods or services billed by the entities of an STP during a year (in millions of €).	Primary sources
F14. Employees in companies	Individual features	Fe06. Employment	Total employees in the companies of the STP.	Primary sources
F15. Company profile	Complementary features	Fe07. International companies	Number of STP firms in which more than 10% of the capital is owned by a foreign company or is a branch or subsidiary of a foreign company.	Primary sources
F15. Company profile	Complementary features	Fe08. Incubated companies	Number of companies less than 3 years old incubated at the STP.	Primary sources
F15. Company profile	Complementary features	Fe09. Average company size 1	=Fe06/Fe04	Knowledge features
F15. Company profile	Complementary features	Fe10. Average company size 2	=Fe05/Fe04	Knowledge features
F15. Company profile	Complementary features	Fe11. Internationalisation	=Fe07/Fe04	Knowledge features
F15. Company profile	Complementary features	Fe12. Productivity	=Fe05/Fe06	Knowledge features
F15. Company profile	Complementary features	Fe13. Incubation ratio	=Fe08/Fe04	Knowledge features
F16. R&I	Complementary features	Fe14. Employment R&D	Employees dedicated to R&D activities in the companies of the STP.	Primary sources
F16. R&I	Complementary features	Fe15. Investment R&D	Expenditure on R&D made by the STP companies (in millions of €).	Primary sources
F16. R&I	Complementary features	Fe16. Filed patents	National patents filed by the STP entities on a yearly basis.	Primary sources
F16. R&I	Complementary features	Fe17. Granted patents	National patents granted to the STP entities on a yearly basis.	Primary sources
F16. R&I	Complementary features	Fe18. Innovative profile 1	=Fe14/Fe06	Knowledge features
F16. R&I	Complementary features	Fe19. Innovative profile 2	=Fe15/Fe05	Knowledge features
F16. R&I	Complementary features	Fe20. Patents ratio 1	=Fe16/Fe04	Knowledge features
F16. R&I	Complementary features	Fe21. Patents ratio 2	=Fe17/Fe04	Knowledge features

The proposed multidimensional data analysis explores 21 core features gathered in the developed STP dataset. They describe the critical internal factors of an STP and are relevant for management and decision making. The work provides relevant data on the current situation of Spanish STPs and their core features. It also offers empirical evidence to support the existence of different types of STPs, including the differences between them and their typical and standard features.

This research will be of interest to policy makers, industry leaders, practitioners, and universities, since it represents an important step in understanding the true nature of STPs, in the context of a real-world situation. Data science researchers and professionals in this field would find the AI-tool-driven data analysis useful for a better understanding of STPs. This insight can improve the management of innovation ecosystems and the orientation of policies and strategies to support global innovation and knowledge ecosystems.

The paper is organised as follows: Section 2 presents the objectives; Section 3 describes the resources and techniques used; Section 4 deals with the evaluation and results of the study; Section 5 presents the discussion and validation; and, finally, Section 6 describes the conclusions.

2. Objectives

The main objective of this work is to deepen the understanding of STPs in general through a case study of the Spanish STPs, specifically developing an exploratory and critical analysis of the STPs' core features and an intrinsic evaluation of the objectives of the study. Moreover, this work also expects to verify that this Spanish STP dataset and dashboard developed by Francés et al. [3], PCT Observer, have a practical and valid application.

Specific objectives (SOs) of the work will be:

- SO1: To have an overview of the general situation of STPs in Spain;
- SO2: To validate that different types of STPs exist with distinct characteristics and to establish whether they fit the three types proposed in the literature (SPs, TPs, and HPs);
- SO3: To know the main and typical features of the different types of STPs.

This paper addresses the objectives using the methodology schematically represented in Figure 1. After the objective statement, different resources and techniques are selected (Section 3) and applied (Section 4). The validation of the objectives is presented in Section 5.

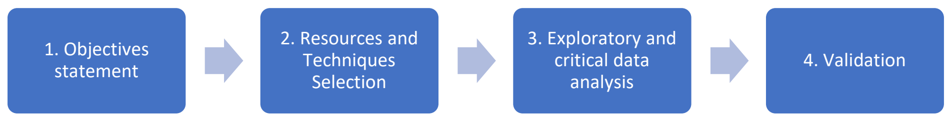


Figure 1. Methodology used.

3. Resources and Techniques

The multidimensional data analysis is applied to the updatable dataset developed in the previous work [2]. Nevertheless, the analysis can be applied to any STP dataset, although these datasets are not readily available nor accessible. The process can be developed with any dataset with a tabular format and the following fields as columns: the STP name, year of operation, and selected features.

A summary of the process for the data analysis, including the previous stages coming from the dataset methodology [2], is presented in Figure 2:

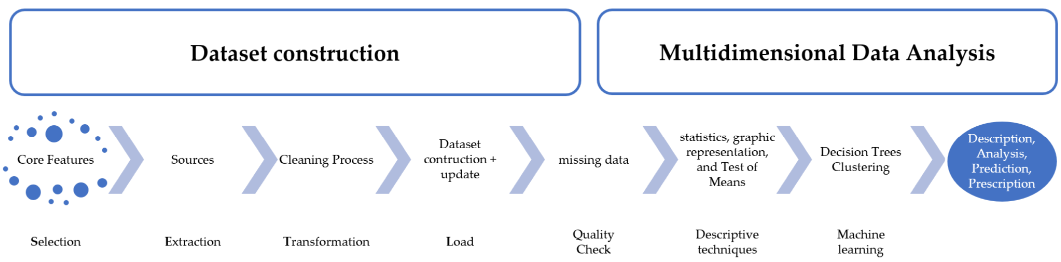


Figure 2. Main stages from the dataset construction and from the multidimensional data analysis.

The dataset is obtained after a cleaning process in which the features and STPs that are not of sufficient quality to be part of a deep analysis are removed. Likewise, once the dataset was built, a quality check was performed with 71% of the core features, being comprehensive and robust enough in the case of the Spanish STPs. All this indicates that the analysis, particularly the one carried out with machine-learning techniques, can only be performed with features of sufficient quality, i.e., for example, with no more than 16% of the data missing to avoid wrongly grouping STPs according to the data that is missing, rather than by the similarity of the features under study. These processes are also detailed in [2].

The descriptive techniques (Section 3.2.1) most typically used (statistics and graphs) serve to obtain an overview of the STPs and detect outliers. The test of means can detect the essential features with significant differences according to the type of STPs, which will be the features on which the machine-learning analysis (decision trees and clusters) will focus, only if these meet the aforementioned quality criteria. Otherwise, these core features are not included to avoid introducing bias into the analysis. Finally, the evidence-based grouping of STPs and the typical features from each type of STP can be obtained.

The achievement of the main objectives is based on an intrinsic evaluation of the general and specific objectives (SOs) outlined in Section 2. To this end, the specific methodology used includes the availability and selection of a series of resources and techniques that will enable the SOs to be addressed and, where possible, validated.

3.1. Resources

For an in-depth understanding of STPs, where a lack of uniformity in the models is observed as explained in Section 1, different resources need to be selected and accessible. This work takes the following resources into consideration:

- A defined terminology in the field of STPs, and the critical factors and core features affecting STPs summarised in Table 2 [2].
- Expert labelling for the case study of Spanish science and technology parks, developed and described by the Francés et al. study [2], where three types of STPs are considered—science parks, technology parks, and hybrid parks—for which short definitions are presented in Table 1.
- A dataset built by Francés et al. [2] in which the internal core features affecting the performance of STPs are presented. It presents a tabular format and covers a 16-year period (2004–2019), and 21 selected core features are considered, together with the STP name (STP identifier) and the year, resulting in a dataset with 23 columns. The rows of the dataset correspond to the STP name–year combination that constitutes each instance of the dataset, amounting to 603 instances in total.

3.2. Techniques

In order to develop the intrinsic evaluation, it is necessary to select a series of techniques that, together with the selected available resources, allow the objectives of the work to be validated. In addition, a visualisation tool has been developed for these techniques and resources, which contains and facilitates the interpretation of the data and techniques used. The dashboard PCT Observer [3], where different available and appropriate tools were applied and represented, facilitates the exploratory and critical analysis of STPs' core features presented in Table 2 and, consequently, the validation of the objectives.

The techniques used in this work are divided into two groups: descriptive techniques and machine-learning techniques applied to data analysis.

3.2.1. Descriptive Techniques

- General statistics references with each core feature, grouped by STPs (all), SP type, TP type, and HP type.
- Multidimensional graphic representation, including the graphic representation of each core feature according to the year or the antiquity of the STPs. Moreover, the graphical

representation of three different core features can be obtained, again for all the STPs, for each one or per type (SPs, TPs, and HPs).

- A two-sample test of means allows us to determine whether one group of STPs is significantly different from others for each core feature and to explore significant differences between types of STPs. The permuted Brunner Munzel test [4] is used, which is more appropriate for small sample sizes, as is the case here. This nonparametric test evaluates the null hypothesis where, when values are drawn, one for each sample, the probability of obtaining larger values in both samples is the same.

3.2.2. Machine Learning Applied to Data Analysis

In this section, the relevant aspects of the two machine-learning techniques used for data analysis are described, with the rationale for using these techniques, as well as details of the implementation, selected.

- Decision Tree Learning: This supervised classification method classifies instances through a sequence of questions related to the value of one of the features describing the pattern. The learning consists of building a set of questions that can be represented as a decision tree, hence the name. The goal is that each question can be used not only to describe an instance but a set of them within the same class ensuring generalisation. This structure makes the decision tree an easily explicable machine-learning algorithm.

The rationale for the decision taken is in line with [5,6]. Decision trees are used to highlight core features and their interactions, describing each type of STP based on experimental and actual data. This technique enables the exploration of different feature sets that can explain with accuracy the STP categories. Despite the limitations of decision tree learning, such as not working well in cases where decision boundaries are not aligned with the data axis [7], its non-parametric nature makes it an attractive technique in the absence of assumptions about the properties of the data. Another potential drawback is they are prone to overfitting when the ratio of features to instances is high. To alleviate this problem, strategies were used, such as constraining the tree's depth or allowing pruning. Different subsets of the features were evaluated.

The optimised version of the CART algorithm [8,9] implemented in scikit-learn was used. This technique has been applied to different fields such as risk management and green building, e.g., [10,11].

- Cluster Analysis: K-means [12–14] was used, a well-known non-parametric technique for cluster analysis that has numerous fields of application such as healthcare, coronavirus, and urban hotspots, e.g., [15–17]. In k-means, each instance is assigned to its nearest centroid, which represents an average of the instances in the cluster. The centroids are updated iteratively until no instance changes its cluster. It may not be well-suited if the expected structure is other than spheric clusters. Not being tied to a specific algorithm, the suitability issue can be addressed by evaluating the results, identifying signals that show the algorithm is not performing as expected and proceeding to study alternatives. Different values for the number of clusters (K) and the silhouette score (s) were used to evaluate the cluster quality. Features were min–max scaled before analysis.

This approach was used to shed light on the structure of the data in contrast to the expert annotation, study the importance of the key indicators, and highlight possible anomalies. The scikit-learn implementation of the k-means was leveraged.

The study presents a knowledge-based theoretical framework supported by a robust dataset and a practical interface dashboard [3] that enables the visualising of different resources and techniques to allow an exploratory and critical analysis, together with intrinsic evaluation. This analysis permits the empirical contrast of the proposed objectives, including the confirmation of different types of STPs in Spain and the determination of their core features.

4. Evaluation and Results

The application of the resources and techniques presented in Section 3 allows for the exploratory and critical analysis to be divided into a descriptive statistical analysis (Section 4.1) that presents the general situation of Spanish STPs and prepares the information for the second key stage of the analysis: a data analysis based on machine-learning tools (Section 4.2).

To analyse the current state of the Spanish STPs, the most recent data range was selected for each STP and feature from the period 2015–2019.

4.1. Descriptive Analysis: General Situation of Spanish STPs

This analysis is associated with SO1 and provides singular information about STPs in Spain given the resources and the descriptive techniques presented in Section 3. It will also give essential information to be used in the machine learning for the data analysis.

A universe of forty-nine (n = 49) Spanish STPs is considered. General statistics references with each core feature, grouped by STPs (all), SP type, TP type, and HP type, has been obtained and are available at https://github.com/gplsi/data-stp (accessed on 27 October 2023).

Fifteen of the forty-nine STPs are considered science parks, thirty are technology parks, and four are hybrid parks. Most of the SPs are promoted by public universities (except for two cases). Most of the TPs are promoted by regional governmental institutions or agencies with the involvement of other public and private drivers. Five out of the forty-nine STPs considered in the study are sectorial parks, four of them technology parks, and one is a hybrid park. According to the study, technology parks are predominant in Spain, constituting 61% of the STP universe. This result is consistent with Albahari et al. [18] who argue that 56% of Spanish STPs are not promoted by universities, in contrast to other countries such as the UK with a clear predominance of university science parks.

The Andalusia, Catalonia, and Valencian community are the regions with the highest number of STPs, forming 47% of the total. The predominant type of STP in the Catalonia and Valencian Community is the science park, in contrast to Andalusia, with 100% of the technology park type.

The first TP in Spain emerged in the mid-1980s, while SPs emerged in the late 1990s. The average age of a TP is 20 years old and that of an SP is less than 16 years, and over 71% of the Spanish STPs in this study were established between 25 and 10 years ago. The number of STPs promoted in the last decade has decreased considerably. One possible reason is that the peak of STP promotion in Spain occurred around the year 2000, when many policies also supported the development of STP infrastructures. Another possible reason is that there are already many STPs in Spain and most provinces have at least one STP.

The exploratory analysis of the STP features determined that there is a specific STP that presents an unusual model since it computes the total number of companies and their turnover, employment, etc., of the Spanish city in which it is located. Therefore, this STP has not been included neither in the descriptive analysis nor in the machine-learning analysis, as it would cause an inference that would distort the analysis developed. Figure 3 shows the large divergence of this STP (that is, a TP) from the rest of the STP universe, with much higher values for such critical features as turnover, number of companies, and employment.

After the deep exploratory analysis developed in this study, more reflections could be considered. With antiquity, STPs (SPs, TPs, and HPs) present a tendency to grow (at different levels) because of the STP consolidation, as can be observed in the following Figure 4.

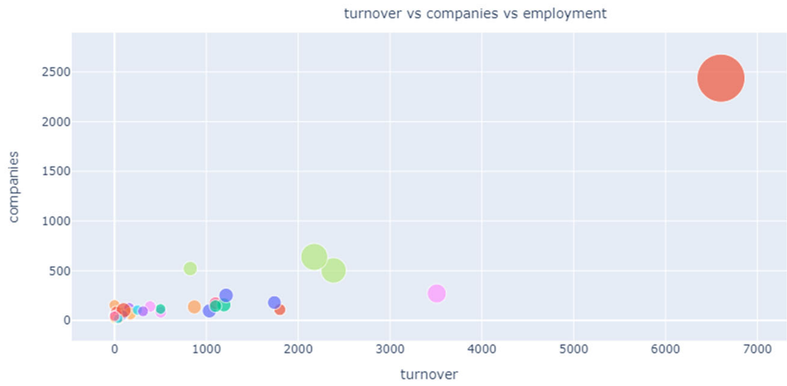


Figure 3. Turnover (Fe05) vs. companies (Fe04) vs. employment (Fe06) from the Spanish STP universe (49). Fe06 is represented proportionally to the size of the marker. Each coloured spot represents a different STP.

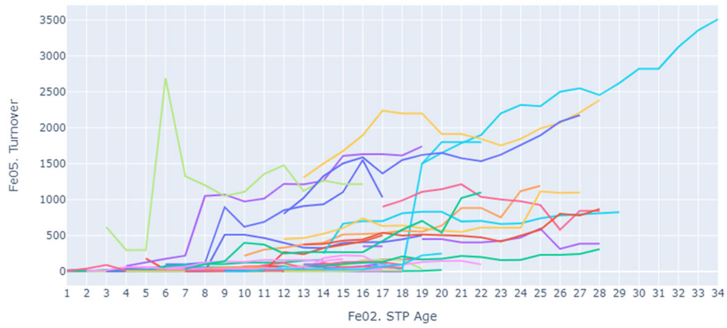


Figure 4. Turnover (Fe05) vs. antiquity (Fe02) in Spanish STPs. Each line represents a different STP from the universe.

As explained above, TPs emerged before SPs in Spain. Moreover, a group of three "big TPs" can be observed, which have experienced significant and steady growth, as well as remarkable performance in terms of general impact (over 2000 million € turnover and 11,000 employees), as can be seen in Figure 5 (in green).

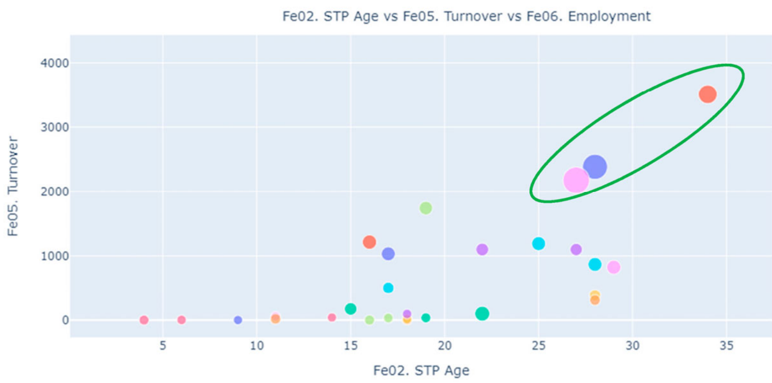


Figure 5. Age (Fe02) vs. turnover (Fe05) vs. employment (Fe06) in Spanish TPs. Fe06 is represented proportionally to the size of the marker. Each coloured spot represents a different STP. "Big TPs" are highlighted with the green oval.

Thus, TPs are larger projects in general terms compared to SPs, but there are considerable heterogeneities between them, with a subgroup of "big TPs" at one extreme.

Furthermore, the permuted Brunner Munzel test (Section 3.2.1) allows the detection of significant differences, in case, by type of STP. As presented in Table 3, there are eight features with no significant differences by type of STP. However, it is observed that SPs are smaller in terms of the following features: Fe02. STP age; Fe03. STP Size; Fe04. Companies; Fe05. Turnover; Fe06. Employment; Fe09. Average company size 1; Fe10. Average company size 2; Fe12. Productivity; and Fe15. Investment R&D. SPs are also smaller than HPs in terms of: Fe03. STP size; Fe06. Employment; and Fe09. Average company size 2. TPs present smaller figures compared to SPs only in terms of Fe13. Incubation ratio and Fe18. Innovative profile 1. HPs show hardly any differences compared to TPs, with Fe19. Innovative profile 2 being the only feature where TPs are smaller than HPs.

Table 3. Permuted Brunner Munzel test results for 20 core features per STP type.

Core Feature	Substantial Difference between SP and TP	Core Feature	Substantial Difference between SP and TP
Fe02. STP age	SP < TP	Fe12. Productivity	SP < TP
Fe03. STP size	SP < TP and SP < HP	Fe13. Incubation ratio	TP < SP
Fe04. Companies	SP < TP	Fe14. Employment R&D	NO
Fe05. Turnover	SP < TP	Fe15. Investment R&D	SP < TP
Fe06. Employment	SP < TP and SP < HP	Fe16. Filed patents	NO
Fe07. International companies	NO	Fe17. Granted patents	NO
Fe08. Incubated companies	NO	Fe18. Innovative profile 1	TP < SP
Fe09. Average company size 1	SP < TP	Fe19. Innovative profile 2	TP < HP
Fe10. Average company size 2	SP < TP and SP < HP	Fe20. Patents ratio 1	NO
Fe11. Internationalisation	NO	Fe21. Patents ratio 2	NO

4.2. Machine Learning Applied to Data Analysis

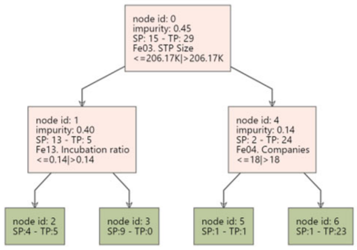
This analysis is associated with SO2 and SO3 and it is based on the resources and the machine-learning techniques applied to the data analysis (Section 3.2.2). The descriptive analysis presented in the previous section allows us to focus on the subsequent data analysis.

4.2.1. Decision Tree Analysis: Dominant Features by STP Type

Decision tree analysis is used for classification tasks. It could suggest the most dominant features for each type of STP, i.e., the critical factors describing each type of STP based on the experimental data. This supervised learning algorithm incorporates the STPs and their core features, using STP type as a decision factor, where it only considers SPs and TPs as part of the data training and includes HPs in the tree classification.

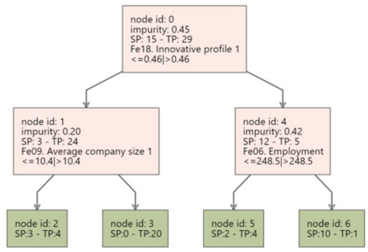
The core features considered in the decision tree analysis are the 12 presenting substantial differences by types of STP (Table 3). However, Fe15. Investment R&D and Fe19. Innovative profile 2 should be rejected, because over 16% of the data was missing, which could detract from the decision tree (force the aggrupation per missing data). Therefore, 10 STP core features have been considered (Fe02, Fe03, Fe04, Fe05, Fe06, Fe09, Fe10, Fe12, Fe13, and Fe18). Decision trees were grown for different subsets of the features, with different hyper-parameters related to tree depth and pruning.

Figure 6 presents two two-level decision trees, as more levels would not bring any further relevant information: Tree 1, for the 10 core features; and Tree 2, for the 9 core features excluding size.



- Leaf node 2:
  - Features:
    Fe03. STP Size <= 206,173 m^2,
    Fe13. Incubation ratio <= 0.14
  - Comprises: 4SPs, 5TPs
  - STP type tendency: mixed
- Leaf node 3:
  - Features:
    Fe03. STP Size <= 206,173 m^2,
    Fe13. Incubation ratio > 0.14
  - Comprises: 9SPs, 2HPs
  - STP type tendency: SP type
- Leaf node 5:
  - Features:
    Fe03. STP size > 206,173 m^2,
    Fe04. Companies <= 18
  - Comprises: 1SP, 1TP, 1HP
  - STP type tendency: mixed
- Leaf node 6:
  - Features:
    Fe03. STP size > 206,173 m^2,
    Fe04. Companies > 18
  - Comprises: 1SP, 23TPs, 1HP
  - STP type tendency: TP type

(**a**)



- Leaf node 2:
  - Features:
    Fe18. Innovative profile 1 <= 0.46,
    Fe09. Average company size 1 <= 10.40
  - Comprises: 3SPs, 4TPs, 1 HP.
  - STP type tendency: mixed
- Leaf node 3:
  - Features:
    Fe18. Innovative profile 1 <= 0.46,
    Fe09. Average company size 1 > 10.40
  - Comprises: 20 TPs.
  - STP type tendency: TP type
- Leaf node 5:
  - Features:
    Fe18. Innovative profile 1 > 0.46,
    Fe06. Employment <= 249
  - Comprises: 2SP, 4TP
  - STP type tendency: mixed
- Leaf node 6:
  - Features:
    Fe18. Innovative profile 1 > 0.46,
    Fe06. Employment > 249
  - Comprises: 10SP, 1TPs, 3HPs
  - STP type tendency: SP type

(**b**)

Figure 6. Decision tree and rule description for: (**a**) 10 core features with significative differences among STP types; and (**b**) 9 core features with significative differences among STP types.

Figure 6a,b presents an acceptable accuracy (0.86) that confirms a clear differentiation between training groups (TPs and SPs). While the trees only include TPs and SPs, in the detailed description, the given HPs are also classified according to their proximity to SPs or TPs.

STP size is the obtained root node from Figure 6a and, therefore, is a very determinant feature in STP classification, with TP size being significantly higher than SP size (Fe03). However, the size in terms of square meters is usually a stable metric in the short term. Therefore, the decision tree analysis in Figure 6b is developed again excluding the size factor and to explore other determinant core features.

Even when only the core features presenting substantial differences were considered in the decision tree analysis, the decision tree considering all the features has been also developed, obtaining the same results, which is very consistent because, as presented in Section 4.1, the rest of the core feature have no significant differences based on STP type.

Considering the decision tree analysis and in line with the previous descriptive analysis, the following STP type profiles can be described:

- The Spanish SP presents a surface area (Fe03. STP Size) under 200,000 m^2, Fe13. Incubation ratio > 0.14, and Fe18. Innovative profile 1 > 0.46 as the most significant feature.
- The Spanish TP presents a surface area over 200,000 m^2, innovative profile 1 less than 0.46, and average company size 1 > 10.40.
- Spanish HPs are between SPs and TPs but tend to be closer to the SP type, although this is not definitive.

However, there are a few TPs performing as SPs and vice versa. Specifically, based on the decision tree analysis, five TPs present features considered closer to the SP type and two SPs show features closer to the TP type.

4.2.2. Clustering Analysis: Natural Aggrupation

Clustering analysis is used to find out the natural aggrupation of STPs, since the whole taxonomy analysis from this chapter is based on the expert labelling in three STP types and observing and analysing analogies or differences between these groups. Through clustering, unobserved patterns can be revealed.

Many combinations of core features were used in this extensive clustering analysis, without considering, as in Section 4.2.1, Fe19. Innovative profile 2 and Fe14. Employment R&D due to their excessive missing data.

The developed dashboard allows the use of different combinations of features and a different number of clusters (K). The silhouette score is obtained for each combination, which shows the efficiency of a clustering technique and aims to interpret and validate the cluster analysis [19]. In the silhouette score (s), the interpretation is:

- 0.71–1.0: well-defined structure;
- 0.51–0.70: "reasonably" well-defined structure;
- 0.26–0.50: weak structure, may be artificial—it is suggested that we try other methods;
- <0.25: no structure has been discovered in the data.

Different combinations of features and number of clusters were applied and the evolution of the silhouette score was analysed. A sample of the most relevant combinations of features is presented in Table 4, where values of s over 0.51 are highlighted. Fe15 and Fe19 are not considered in any combination, as missing data accounts for over 16% for these features.

The most interesting approaches in this clustering analysis for the aim of this study, both of them based on impact, are as follows:

(a) Standard economic impact, where the typically considered features are Fe05. Turnover and Fe06. Employment (#4 in Table 4);
(b) Innovation and entrepreneurship focus impact, where the core features considered are Fe18. Innovative profile 1 and Fe13. Incubation ratio (#5 in Table 4).

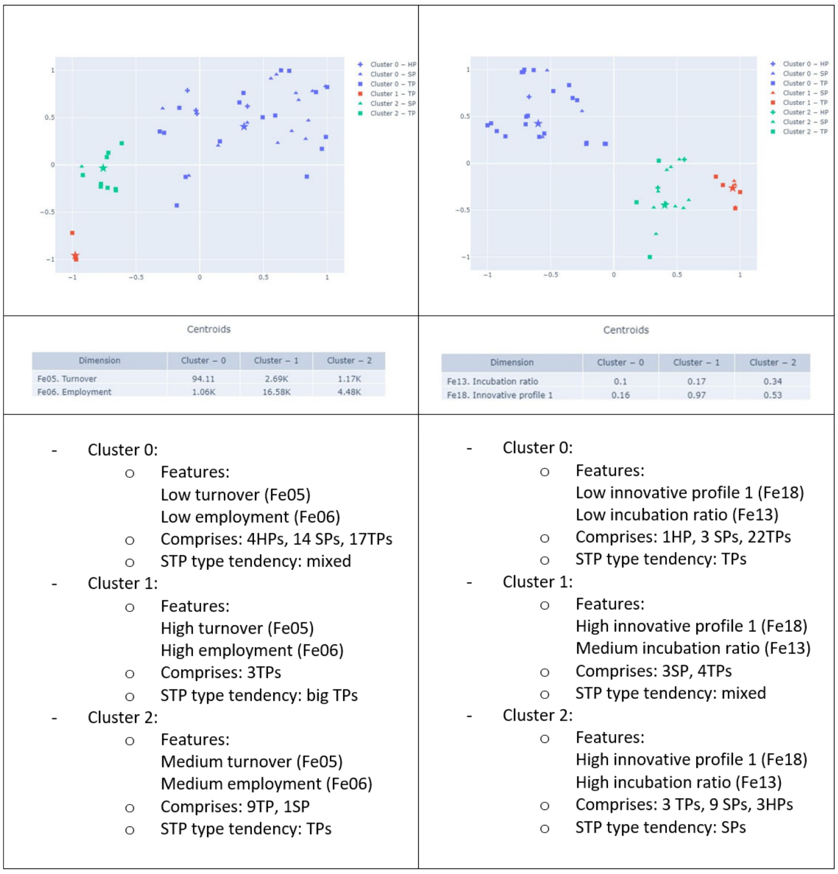
Figure 7a,b shows the results of the clustering analysis for both impact considerations. A description of the clusters formed is also included.

The standard economic impact (Figure 7a) delivers a maximum when K = 2. This clustering shows two groups, one being the larger STPs, that are three TPs. The next clustering analysis when K = 3 presents a remarkable, well-defined structure as indicated by s = 0.72. This delivers more information and, therefore, is the one represented in Figure 7a. This result suggests that the following three clusters can be defined considering the economic impact: the cluster 2 TP type with medium turnover and medium employment rates; the cluster 1 big TPs with higher rates; and cluster 0, a mixture of TPs, SPs, and HPs in the third, and numerous clusters with moderate rates for turnover and employment.

The clustering results for the impact in terms of innovation and entrepreneurship (Figure 7b) present a maximum when K = 4, but the fourth cluster was formed only by one TP with a high incubation ratio (Fe13), an outlier, presenting a value over 1. Apparently, this TP incubates more young companies than the companies belonging to it. Therefore, clustering when K = 3 for this case gives more information and presents a very similar and acceptable s = 0.56, which means a reasonably well-defined structure. These results suggest that the following three valid clusters can be defined considering the innovation and entrepreneurship impact: the cluster 2 SP type with remarkable innovation and entrepreneurship ratios, which contains 60% of the SPs; the cluster 0 TP type with low ratios, which contains 76% of the TPs; and cluster 1, a cluster with medium ratios and a mix of three SPs and four TPs.

Table 4. Clustering analysis for different combinations of core features, with K being the number of clusters considered and s being the silhouette score.

	Core Features Considered																					**K**								
#	Fe01	Fe02	Fe03	Fe04	Fe05	Fe06	Fe07	Fe08	Fe09	Fe10	Fe11	Fe12	Fe13	Fe14	Fe15	Fe16	Fe17	Fe18	Fe19	Fe20	Fe21	2	3	4	5	6	7	8	9	10
1																						0.30	0.28	0.22	0.22	0.21	0.22	0.23	0.20	0.19
2																						0.30	0.31	0.31	0.26	0.28	0.27	0.30	0.26	0.23
3																						0.30	0.28	0.32	0.34	0.34	0.32	0.31	0.34	0.37
4																						0.80	0.72	0.72	0.61	0.62	0.58	0.59	0.57	0.54
5																						0.50	0.56	0.57	0.44	0.39	0.35	0.37	0.37	0.41



(**a**) (**b**)

Figure 7. Clustering analysis for: (**a**) standard economic impact; and (**b**) innovation and entrepreneurship impact. Stars represent the centroids of each cluster.

In line with the decision tree analysis (Section 4.2.1), the clustering analysis conducted shows that some TPs tend to cluster according to their features with SPs and vice versa.

The centroids of each cluster represent the typical features and can indicate the distance of an STP from the prototype. This distance can support decision making in the management of an STP. For example, in the case of an STP—that is classified as an SP, focuses its objectives on fostering innovation and entrepreneurship (Figure 7b), and has, for instance, an incubation ratio of 0.22—this analysis suggests that it should set policies to reach the value of 0.34 to be in the range of its prototype SP.

Similarly, if the impact requiring analysis diverges from the two proposed main ones (i.e., economic and innovation–entrepreneurship), the analysis could be adjusted.

5. Discussion and Validation

In this section, the results in relation to the general and specific objectives outlined in Section 2 are interpreted and discussed separately:

- SO1: To have an overview of the general situation of STPs in Spain.

As presented in Section 4.1, a description of the current situation of the Spanish STPs is obtained from the resources and descriptive techniques applied. General statistics references with each core feature, grouped by STPs (all), SP type, TP type, and HP type, has been obtained and are available at https://github.com/gplsi/data-stp (accessed on 27 October 2023). Forty-nine STPs were considered in the study, with 61% of them grouped initially as technology parks (TPs), 31% grouped as science parks (SPs), and 8% as hybrid parks (HPs).

Most TPs were promoted by regional governmental institutions, whereas most SPs were promoted by public universities. A predominancy of technology parks promoted between 25 and 10 years ago is observed and the aggregated values of the 21 core features studied in this work are available.

- SO2: To validate that there are different types of STPs with distinct characteristics and whether they correspond to the three types proposed in the literature (SPs, TPs, and HPs).

As shown in Table 3, 12 of the 20 analysed core features are significantly different between STP types. Decision tree analysis (Section 4.2.1) also reveal with high accuracy rates that SPs and TPs present singular tendencies, and a typical profile for each type can be obtained. Clustering analysis (Section 4.2.2) also confirms this SP–TP difference in terms of innovation and entrepreneurial impact.

Regarding the distinct characteristics, the decision tree analysis shows that Fe03. Size and Fe18. Innovative profile 1 are determinants when differentiating SPs from TPs. Clustering analysis also identifies a natural aggrupation tendency for SPs when considering the innovation and entrepreneurship impact, with more than 60% of SPs performing highly in this respect. SPs seem to be clearly more R&D- and innovation-oriented in relative terms considering R&D employment, but no significantly different activity is observed in terms of patents. Fe19. Innovative profile 2, which represents the relative influence of R&D investment, was not available neither for decision tree analysis nor for clustering, due to over 16% of the data being missing for this feature. The TP group present higher values with respect to SPs in terms of Fe03. Size, Fe06. Employment, Fe05. Turnover, Fe09. Average company size 1, and Fe10. Average company size 2.

The data analyses using machine learning carried out in this study confirms notable differences between the SP type and the TP type. However, the HP group is not clearly defined. In the decision tree and clustering analysis, there is a slight tendency to be closer to the SPs. However, this group of HPs is a very small sample, and it is difficult to draw categorical or clear conclusions.

Based on the above, the results suggest that it is possible to establish a clear taxonomy of the two main groups: SPs and TPs. For both types, there are some cases that escape the trend where the behaviour is not typical of their group but tends to behave like the other group. The study suggests that few SPs perform as TPs and vice versa. Concretely, four SPs behave as TPs and five TPs behave as SPs. In addition, there is a third group of hybrid parks (HPs) with a very small initial sample (four) and which also shows intermediate behaviour between TPs and SPs. This study suggests that this initial sample of HPs would be bigger after using the proposed data analysis and intrinsic evaluation.

Therefore, it seems logical to assume that, for the three groups, although they generally remain intact in terms of what defines their characteristics/features, in the case of the hybrid group, the analysis conducted suggests that this grouping needs to extend to SPs and TPs that do not display the typical behaviour of their original group. The proportional groupings are presented next:

- 11 SPs, representing 73% of the initial sample of 15 SPs;
- 25 TPs, representing 86% of the initial sample of 29 TPs;
- 13 HPs, representing 100% of the initial sample of 4 HPs, plus 4 SPs with behaviour like the TP type, plus 5 TPs with behaviour like the SP type. Finally, there are three times the amount of HPs compared to the original sample.

The new HP category should be broadened to include not only a balance between the university and the government sector in the promotion and management of the STP, but also other unique situations that lead to a better balance between the features purely associated with SPs, including university influence, and those associated with TPs. Therefore, the original definition could be revised. Expert labelling is a good starting point for classifying STPs and the subsequent descriptive and data analysis presented in this work gives a complementary point of view of the real situation and allows a more accurate classification of the STP's type.

The representation of this situation is presented in Figure 8 below, where STP types from expert labelling are compared with the final classification after this work's analysis.

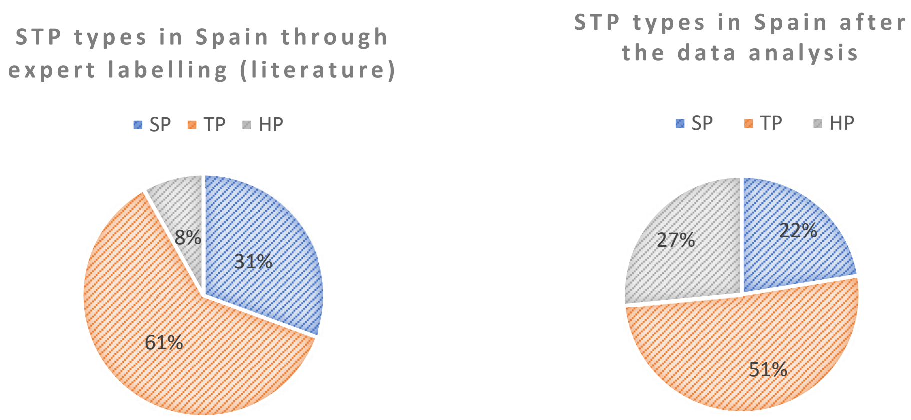


Figure 8. Spanish STP types through expert labelling and with data analysis.

Other studies have broadly analysed heterogeneity in the types of STPs [20–24]. Albahari et al. [20]. found that 20% of Spanish STPs are SPs, which is consistent with the results obtained in this study. Albahari et al. [18]. state that 37% of Italian ETPs would be TPs, which is significantly lower than in the Spanish case. Link and Scott [23], in a sample of 51 American STPs, found that 31% were SPs, which is also close to the Spanish results. At the other extreme, apparently, all UK STPs are university science parks, located in or near the institution itself [22].

The proposed in-depth analysis methodology allows for a more accurate classification of each STP, which, in turn, also enables the activation of informed decision-making mechanisms that can improve its performance (e.g., by benchmarking against its prototype STP to target the features that need to be improved).

This in-depth study is planned for other geographical regions. So far, it has only been possible to replicate it for the Argentinean case, where a comparative study has also been established for the Spanish case [24]. In Argentina, three types of STPs are also validated, the proportion being: 41% SPs, 17% HPsm and 42% TPs. Therefore, SPs are proportionally more numerous than in Spain. The prototype characteristics obtained for the Argentinean case focus on the size and number of companies. The SPs are significantly larger than the TPs, although with a smaller number of companies, more focused on biotechnology or chemical sectors compared to the TPs, which have a larger number of companies focused on ICT.

Therefore, SO2 has been achieved and validated, as this study suggests the presence of three STP types with distinct characteristics.

- SO3: To know the main and typical features of the different types of STPs.

SPs and TPs can be typified. With respect to a qualitative representation of the average features for SPs and TPs, the following Figure 9 is presented, where only the core features with significant differences among STP types are included:

Complementarily, a quantitative description with some relevant core features can be established based on the data analysis using machine learning (Section 4.2) and the following typical values can be presented:

- The typical Spanish SP presents a surface area (Fe03) under 200,000 m^2, the incubation ratio > 0.14 being the reference average of 0.34. The innovative profile 1 (Fe18) is the most significant feature and is over 0.46, which is the benchmark value. The typical average Spanish SP case has a turnover (Fe05) of 94 million € and around 1000 employees (Fe06).

- The typical Spanish TP presents a surface area (Fe03) over 200,000 m^2 and an innovative profile 1 (Fe18) less than 0.46, with the benchmark average at 0.16. Typically, the representative value of the incubation ratio is 0.1. The typical average Spanish TP case presents an annual turnover (Fe05) of 1170 million € and around 4500 employees (Fe06).

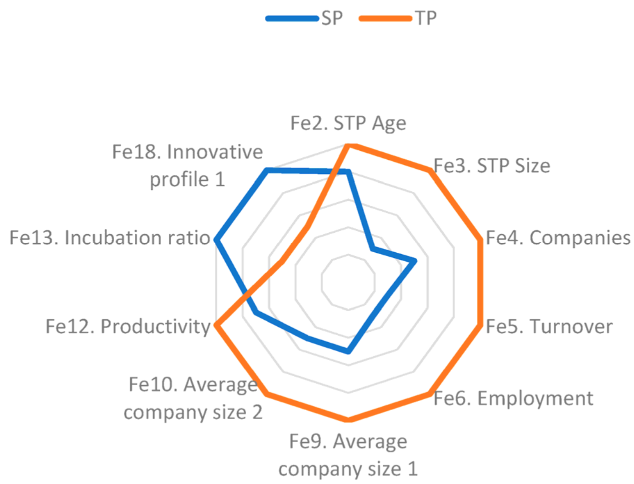


Figure 9. Qualitative representation of the core features for Spanish SPs and TPs.

The proposed multidimensional analysis enables a greater understanding of STPs and a practical application when establishing promotion policies and also when guiding the strategies of a specific STP. For example, an STP can be compared to the STPs in its region, or it can also determine which specific feature it needs to improve to better meet the STP model it aims to be or the specific impact it seeks to make. Although this analysis has focused on the economic impact and the innovation–entrepreneurship impact, the proposed methodology is replicable with other customised impacts (e.g., internationalisation or employment generation).

These applications are very relevant, since, to date, no in-depth analysis of these innovation ecosystems has been carried out using machine-learning techniques. The literature mostly shows case studies, exploratory statistical analysis limited to very specific features, and comparative studies of companies inside or outside STPs. Therefore, this study represents a significant advancement in knowledge related to STPs, as well as a tested methodology to assist in decision making and their overall management.

6. Conclusions

The combination of descriptive and machine-learning techniques has been applied to the Spanish STP case study, revealing the feasibility of the data analysis and their use for a better understanding of STPs and informing decision making. The proposed methodology provides an overview of the general situation of STPs in Spain, validates three main types of STPs, and provides valuable quantitative information about the core features. The methodology benefits from being highly applicable in guiding the management of STPs. Therefore, the study has successfully fulfilled the outlined research objectives.

The results obtained for the Spanish case are relevant because the maturity of the STP ecosystem in Spain is significant and includes STPs with a great variety in terms of the degree of university involvement. Therefore, the results can be useful for other regions with similar characteristics. Additionally, both the selected core features and the analysis methodology are common reference points for similar research that can be developed in other regions.

Moreover, the following concluding remarks can be made based on the data analysis conducted for the Spanish case study:

- The exploratory and a critical analysis presented in this study deepens the understanding of STPs. The methodology used is effective for achieving this goal. The previously developed STP dataset and the dashboard tool enable the data analysis and the visualisation of the techniques applied as an integrated tool to better analyse STPs.
- It is possible to establish a clear taxonomy of the two main groups, science parks and technology parks, based on their intrinsic features.
- A third expanded group of hybrid parks should be considered to include projects that not only have university and the governmental sector involvement, but also other outliers that cannot be strictly grouped as an SP or TP.
- The typical core features of the SPs and TPs of the Spanish use case have been obtained. SPs are more oriented toward R&D, innovation, and incubation, while TPs present higher values in size, employment, and turnover, for example.
- Relevant differences are clearly detectable in TPs, where a relevant subgroup of "big TPs" can be observed with very remarkable values in terms of turnover and number of employees.
- The descriptive and data analysis presented in this work enable greater accuracy in the classification of STP types.
- It is possible to predict the STP type and the distance from the "prototype STP" for each group, which could assist in STP management and decision making.

The study presents some limitations, as only essential features describing endogenous factors of STPs are initially considered. External factors, such as the regional economic situation or the university ecosystem, can strongly influence the performance of an STP, as described by Francés et al. [2]. These exogenous factors have not been considered in this study and can be incorporated to develop a more comprehensive analysis of the characteristics. In this sense, the availability of information sources in the field of STPs is crucial for achieving better results. Another limitation is that this extended analysis has only been performed for specific use cases, with Spanish STPs in this study, as well as a similar analysis for Argentinian STPs [24], where the sample of STPs and the number of features is more restricted. A wider universe of STPs from different countries and more features could be included. In this sense, the methodology enables the easy incorporation of information from other sources and facilitates the updating of the dataset and the replication of the data analysis.

Future work will deepen both prediction and prescription in the field of STPs using machine-learning tools. It is also possible to incorporate other features and extend the analysis to other geographical areas, making it possible to develop a holistic STP observatory.

Overall, it is important to keep fostering data culture and performance metrics within innovation ecosystems. Collecting as much information as possible will facilitate the processing and use of data as input for analysis tools, including AI-based systems. All this contributes to evidence-based decision making and the more efficient allocation of resources to better manage innovation ecosystems.

Author Contributions: Conceptualization, O.F., Y.G. and M.P.; Methodology, O.F., J.A.-S., Y.G. and M.P.; Software, J.A.-S. and J.F.; Validation, O.F.; Formal analysis, O.F.; Investigation, O.F.; Resources, O.F., J.A.-S. and J.F.; Data curation, J.A.-S. and J.F.; Writing—original draft, O.F.; Writing—review & editing, Y.G. and M.P.; Supervision, Y.G. and M.P.; Project administration, M.P.; Funding acquisition, Y.G. and M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research work has been partially funded by the Generalitat Valenciana through the project NL4DISMIS: Natural Language Technologies for dealing with dis- and misinformation with grant reference (CIPROM/2021/21); the Ministry of Science and Innovation, PID2021-123956OB-I00, CORTEX; PID2021-122263OB-C22 COOLANG; and the R&D project CLEARTEXT TED2021-130707B-I00.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. The data were obtained from the Spanish Association of Science and Technology Parks (APTE), the owner of the data, and are available by signing a confidentiality agreement with APTE.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sanz, L. Los Parques Científicos como espacios multidisciplinares. *Encuentros Multidiscip.* **2011**, *13*, 2–9.
2. Francés, O.; Fernández, J.; Abreu, J.; Gutiérrez, Y.; Palomar, M. How to build a standardised dataset for Science and Technology Parks from heterogenous sources. *Data Knowl. Eng.* 2023; *under review*.
3. Francés, O.; Abreu, J.; Gutiérrez, Y.; Fernández, J.; Palomar, M. PCT Observer Tablero de Parques Científicos/Tecnológicos. 2021. Available online: https://rua.ua.es/dspace/handle/10045/122837 (accessed on 31 October 2023).
4. Brunner, E.; Munzel, U. The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. *Biom. J. J. Math. Methods Biosci.* **2000**, *42*, 17–25. [CrossRef]
5. McArdle, J.J. Exploratory data mining using decision trees in the behavioral sciences. In *Contemporary Issues in Exploratory Data Mining in the Behavioral Sciences*; Routledge: London, UK, 2013; pp. 25–69. [CrossRef]
6. Hennessy, E.A.; Finch, A.J. Adolescent recovery capital and recovery high school attendance: An exploratory data mining approach. *Psychol. Addict. Behav.* **2019**, *33*, 669. [CrossRef] [PubMed]
7. Bishop, C.M.; Nasrabadi, N.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006; Volume 4, p. 738.
8. He, Z.; Wu, Z.; Xu, G.; Liu, Y.; Zou, Q. Decision tree for Sequences. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 251–263. [CrossRef]
9. Liu, J.; Feng, W.; Zhang, Y.; He, F. Improvement of PBFT Algorithm Based on CART. *Electronics* **2023**, *12*, 1460. [CrossRef]
10. Li, X.; Yi, S.; Cundy, A.B.; Chen, W. Sustainable decision-making for contaminated site risk management: A decision tree model using machine learning algorithms. *J. Clean. Prod.* **2022**, *371*, 133612. [CrossRef]
11. Motaghifard, A.; Omidvari, M.; Kazemi, A. Forecasting of safe-green buildings using decision tree algorithm: Data mining approach. *Environ. Dev. Sustain.* **2023**, *25*, 10323–10350. [CrossRef]
12. Li, M.; Frank, E.; Pfahringer, B. Large scale K-means clustering using GPUs. *Data Min. Knowl. Discov.* **2023**, *37*, 67–109. [CrossRef]
13. Nie, F.; Li, Z.; Wang, R.; Li, X. An effective and efficient algorithm for K-means clustering with new formulation. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 3433–3443. [CrossRef]
14. Zhao, X.; Nie, F.; Wang, R.; Li, X. Robust fuzzy k-means clustering with shrunk patterns learning. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 3001–3013. [CrossRef]
15. Bhopale, A.; Zanwar, S.; Balpande, A.; Kazi, J. Optimised Clustering Based Approach for Healthcare Data Analytics. *Int. J. Next-Gener. Comput.* **2023**, *14*. [CrossRef]
16. Kadali, D.K.; Mohan, R.N.V.; Padhy, N.; Satapathy, S.; Salimath, N.; Sah, R.D. Machine learning approach for corona virus disease extrapolation: A case study. *Int. J. Knowl. Based Intell. Eng. Syst.* **2022**, *26*, 219–227. [CrossRef]
17. Li, Y.; Zhou, X.; Gu, J.; Guo, K.; Deng, W. A Novel K-Means Clustering Method for Locating Urban Hotspots Based on Hybrid Heuristic Initialization. *Appl. Sci.* **2022**, *12*, 8047. [CrossRef]
18. Albahari, A.; Catalano, G.; Landoni, P. Evaluation of national Science Park systems: A theoretical framework and its application to the Italian and Spanish systems. *Technol. Anal. Strateg. Manag.* **2013**, *25*, 599–614. [CrossRef]
19. Rousseeuw, P. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
20. Albahari, A.; Barge-Gil, A.; Pérez-Canto, S.; Modrego, A. Technology parks versus science parks: Does the university make the difference? *Technol. Forecast. Soc. Change* **2017**, *116*, 13–28. [CrossRef]
21. Liberati, D.; Marinucci, M.; Tanzi, G.M. Science and technology parks in Italy: Main features and analysis of their effects on the firms hosted. *J. Technol. Transf.* **2015**, *41*, 694–729. [CrossRef]
22. Siegel, D.S.; Westhead, P.; Wright, M. Assessing the impact of university science parks on research productivity: Exploratory firm-level evidence from the United Kingdom. *Int. J. Ind. Organ.* **2003**, *21*, 1357–1369. [CrossRef]
23. Link, A.N.; Scott, J.T. The economics of university research parks. *Oxf. Rev. Econ. Policy* **2007**, *23*, 661–674. [CrossRef]
24. Francés, O.; Abreu, J.; Gutiérrez, Y.; Palomar, M. Estado de los Parques Tecnológicos en Argentina y estudio comparativo con la situación española. In Proceedings of the XX Congreso Latino-Iberoamericano de Gestión Tecnológica y de la Innovación ALTEC 2023, Paraná, Argentina, 22 September 2023.