UNIVERSITY OF TARTU

Faculty of Social Sciences

School of Economics and Business Administration

Emre Kaan Sarikaya | Abhishek Giri

Decoding the Innovation Ecosystem in Estonia:
A Clustering Approach to the Innovation Ladder Model

Supervisor: Professor Kadri Ukraniski

Tartu, 2023

We have written this Master Thesis independently. Any ideas or data taken from other authors or other sources have been fully referenced.

Table of contents

Abstract

This study aims to validate the Innovation Ladder Model, developed by the Innovation Ladder Working Group, by applying it to Estonia's CIS 2018 dataset. The research methodology involved selecting relevant questions from the CIS 2018 dataset and grouping them according to the dimensions in the Innovation Ladder Model. Companies were then clustered with K-means algorithms for each group of questions, and their innovation levels were determined for each dimension, followed by another round of clustering to categorise Estonian companies' innovation stage and validate the Innovation Ladder Model and finally present sector analysis based on the Innovation Ladder model. The results of this study contribute to the understanding of innovation performance in Estonian companies and provide empirical evidence for the applicability of the Innovation Ladder Model in different contexts. By validating the Innovation Ladder Model, this research has implications for innovation management, policy-making, and strategic decision-making, as it offers a robust tool for assessing and comparing companies' innovation capabilities. Further research is encouraged to apply the model to other datasets and industries to enhance its generalizability and utility in different settings.

1. Introduction

In today's globalised economy, innovation plays a critical role as a driver of economic growth, productivity, and competitiveness. Businesses that embrace continuous innovation can gain a competitive advantage, adapt to evolving market conditions, and foster sustainable growth. Estonia, a small Baltic nation, has emerged as a leader in digital transformation and information and communication technology (ICT), with a strong focus on innovation (Reila, 2021). However, to effectively support and enhance the innovation capacity of businesses, it is essential to understand the factors that contribute to their innovation potential and the challenges and barriers they face. This knowledge forms the basis for crafting policies and strategies. The Innovation Ladder Working Group 2022, consisting of Estonian academicians and working professionals, is actively developing the Innovation Ladder Model (Appendix A), an open-source methodology, businesses can use to assess their current innovation state. It provides a framework to determine the next steps regarding policies, processes, and collaborations necessary to enhance their innovative output.

This thesis is organised into several sections, each serving a specific purpose. Section 2 presents a comprehensive literature review, focusing on critical theoretical and empirical studies related to the various stages of innovation in businesses and the innovation landscape in Estonia. Section 3 delves into the details of the Innovation Ladder Model, discussing the data utilised in this research study, data sources, and the research methodology employed. The section also describes the analytical techniques used to assess the innovative stages of Estonian firms. Moving forward, Section 4 presents the empirical findings obtained from the research. It highlights the significance of the Innovation Ladder and sheds light on the gradual progression of businesses towards higher stages of innovation with sector analysis. Finally, Section 5 concludes the thesis, summarising the main findings and proposing potential avenues for future research. It also suggests alterations to the current version of the Innovation Ladder Model. The ultimate goal of this study is to enhance knowledge regarding the stages of innovation in businesses and provide valuable insights to the Innovation Ladder Model Working Group 2022, facilitating model improvement and alignment with the dynamic realities of the real-world market.

2. Literature Review

Society places great importance on innovation as it helps to drive progress and improve our way of life. Innovation is broadly used to describe the practical application of new ideas. It can refer to developing new products or implementing technological advancements that help us better utilise existing resources. Innovation is one of the crucial factors for the economic growth of businesses, individuals, and countries (Romer, 1986).

Assessing a firm's innovation potential is crucial for identifying strengths and weaknesses and developing strategies to improve innovation capabilities. There has been a growing interest in assessing innovation potential in recent years, as it provides a basis for developing policies and strategies to foster innovation. *Innovation* is a complex and multifaceted concept that can be defined differently. Innovation is creating and implementing new or improved products, services, processes, or business models that generate economic value and improve the quality of life (Schumpeter, 1934). Teece (1986) posits that a firm's innovation ability is vital for maintaining its competitive advantage and ensuring long-term success. Innovation potential refers to the ability of a firm to generate and implement new or improved innovations (Tidd & Bessant, 2020). A firm's ability to innovate and create value in the future includes various components such as the firm's technology capabilities, human resources, organizational structure, and external environment. The assessment of innovation potential requires a holistic and dynamic approach considering all these components. Innovation potential comprises three main components: absorptive capacity, innovative capability, and innovation opportunity (Damanpour & Aravind, 2012). Absorptive capacity refers to a firm's ability to acquire, assimilate, and apply external knowledge and technologies (Cohen & Levinthal, 1990). Innovative capability refers to a firm's internal resources and capabilities that enable it to generate and implement new or improved innovations (Teece, 1998). Innovation opportunity refers to the external factors that create opportunities for innovation, such as market demand, technological developments, and regulatory changes (Damanpour & Aravind, 2012).

Assessing innovation potential is a complex and multidimensional process that involves multiple methods and approaches. Various methods and approaches are used for assessing innovation potential, depending on the context and objectives of the assessment. According to Gunday et al. (2011), surveys and interviews help collect qualitative and quantitative data on innovation potential from employees, customers, and stakeholders. One of the most commonly used methods for assessing innovation potential is the innovation audit, which involves a systematic and comprehensive review of a firm's innovation capabilities and processes (Tidd & Bessant, 2020). Other methods include innovation surveys, innovation indicators, and benchmarking (Damanpour & Aravind, 2012). Case studies and benchmarking can provide insights into other firms' best practices and performance in the same industry or region (Lööf & Heshmati, 2006). The choice of method depends on the purpose of the assessment, the level of analysis (i.e., firm, industry, or country), and the availability of data and resources. Each method has advantages and limitations, and no single method is universally applicable or superior to others (Damanpour & Aravind, 2012). Alhusen et al. (2021) researched determining indicators for measuring the innovation capabilities of German companies. Their research provides 47 new indicators to compare and contrast the firm's approach to innovation. The crucial point is that indicators must be observable in successive periods. However, dealing with different types of firms may lead to the interpretation of the results. (Boly et al. 2014) applied the IC measure framework with 15 innovation management practices to French companies. The most crucial step is categorizing companies and evaluating them in their league. Based on the previous findings, our initiation point combines traditional measurement methods with quantitative study methods. Cavdar and Aydin (2015) searched for the significant factors that affect technological development and innovation among the different sub-groups. Their research shows that patent applications are significant variables in the model.

Estonia, a small Baltic country in Europe, has dramatically emphasised innovation and technology since its independence from the USSR regime. The government has implemented policies encouraging innovation and promoting the startup spirit, from cryptographic authenticated National ID card in the early 2000s to introducing i-voting, an online voting system in 2005, the first country in the world to do so (Ehin et al., 2022). The Estonian economy is an excellent example of innovation driving economic growth and competitiveness. Since regaining its independence in 1991, Estonia's economy has grown over six times. Despite being a small country with a GDP of 37.9 billion euros (World Bank Open Data, 2021), Nowadays, Estonia is considered to emerge as a hub for innovation and entrepreneurship in the Baltic region. Given the small size of the Estonian economy and high social security costs in European companies, Estonian businesses cannot compete on price levels. Instead, high-quality services and innovation are critical to compete with low labour-cost countries (Hogeforster, 2014). Collaboration and networks are key factors that influence a firm's innovation potential (Powell et al., 1996). Estonia has a highly skilled workforce, a favourable business climate, and a supportive government policy that encourages innovation and investment in R&D. Tiits et al. (2015) suggest that Estonia's innovation system is supported by robust public and private institutions and a solid commitment to research and development (R&D). Such support includes R&D tax incentives, innovation grants, and initiatives like the Estonian Entrepreneurship Growth Strategy (Ministry of Economic Affairs and Communications, 2014). The country has a well-developed innovation ecosystem, a robust education system, a supportive business environment, and a vibrant startup scene (European Innovation Scoreboard, 2021). However, the innovation landscape in Estonia is still evolving, and several challenges must be addressed. Despite government support, R&D investments in Estonian firms remain relatively low compared to other European countries (Eurostat, 2021).

According to Eamets et al. (2014), challenges that Small and medium-sized enterprises face include more funding for innovation, limited access to talent and resources, and low collaboration among firms and research institutions. Despite recent positive developments, Innovation in Estonian firms faces several challenges, including a lack of access to financing, limited cooperation between academia and industry, and a need for more skilled workers (National Innovation Scoreboard, 2021). The literature on Innovation in Estonian firms highlights the need for more research on the factors influencing innovation potential and the best practices for assessing and improving innovation capabilities. Other than the researchers trying to determine the innovation potential and the factors affecting Innovation, Several companies worldwide specialise in assessing innovation and providing consultation services to help businesses increase their innovation potential and impact. Innovation360 is one such company with local offices in more than eight countries. Lean Business Ireland, a community leader in Ireland, also brings together all the key elements that support and build competitiveness. Estonia Tehnopol, a science and business park organisation based in Tallinn, Estonia, helps small and medium-sized businesses grow faster. These organisations analyse different business parameters to determine the current stage of innovation in businesses, identify areas for improvement, and offer consulting services. The Innovation Ladder Working Group thoroughly considers the models, assessment methodology and tools before arriving at the current version of the Innovation Ladder Model. These private and public organisations play a vital role in promoting innovation growth by providing companies with valuable insights into their innovation potential and guiding them in developing strategies to improve their innovation capabilities. Here is a brief info about some of them relevant to our study because of their methodology and how they help analyse companies' innovation potential and impact. Innovation360 Group provides consulting services for innovation management and assessment. Primarily they help small and medium businesses to improve their innovation capabilities through data-driven insights and tools (Innovation360 Group, 2021). Innovation360 Group uses their proprietary assessment tool, InnoSurvey®, to evaluate various dimensions of a company's innovation capabilities. Innovation360 Group's services aim to help companies become more innovative, competitive, and successful by leveraging their innovation potential (Innovation360 Group, 2023). i360 Assessment is a comprehensive innovation assessment tool to evaluate a company's capabilities and potential. The tool assesses over 100 innovation capabilities using a proprietary framework called the InnoSurvey® across all areas of a company's operations, including strategy, leadership, culture, and operations.

Another organisation working in the same direction is Lean Business Ireland, which provides support, training, and resources to Irish businesses to help them implement lean business principles and improve their competitiveness (About Us; Lean Business Ireland, 2022). Their relevant service is the Innovation Health Check, a comprehensive assessment of a company's innovation capabilities. The process involves reviewing the company's strategy, culture, and processes and evaluating its innovation performance. A team of experienced innovation practitioners conduct this assessment, tailored to the company's specific needs/stages. The Innovation Health Check gives businesses a report outlining their strengths and weaknesses, recommending actions to improve their innovation capabilities and a roadmap (Enterprise Ireland Client Services; Lean Business Ireland, 2020). Also worth mentioning is Tehnopol, a science and business park organisation located in Tallinn, Estonia. Tehnopol offers an "Innovation Health Check Audit," designed to help companies assess their current innovation level and identify improvement areas (Tehnopol, 2022). The audit comprehensively evaluates the company's innovation processes, strategies, and culture. After completing the audit, Tehnopol provides a detailed report with recommendations and action items to help the company increase its innovation potential and impact. The service primarily aims at small and medium-sized businesses looking to grow faster (Tehnopol, 2022a).

The findings of this literature review have significant implications for the development of a comprehensive framework to assess the factors influencing innovation potential and validate the Innovation Ladder Model. By leveraging these findings, firms can effectively identify their strengths, weaknesses, and developmental stages, enabling them to devise strategies to enhance their innovation capabilities. Considering the vital role of innovation in fostering economic growth and prosperity, conducting thorough assessments becomes imperative for firms to identify areas of improvement and implement strategies that fuel their innovation capacities. Such endeavours are crucial for firms to remain competitive and flourish within the dynamic landscape of the modern business environment. Assessing a business's current innovation stage holds paramount importance as it guides the establishment of policies and practices aligned with key objectives and goals, leading to enhanced innovation outcomes. Innovation potential encompasses a multifaceted and multidimensional process that permeates all aspects of business operations, encompassing production, processes, delivery, and services. Evaluating innovation potential offers two pivotal benefits. Firstly, it allows for the identification of factors that either facilitate or impede innovation, enabling the development of policies and strategies that foster a culture of innovation. Secondly, it equips firms with an indispensable tool to improve their innovation capabilities and effectively navigate the rapidly evolving business landscape while maintaining competitiveness.

Despite positive strides made in the innovation landscape of Estonia, local firms encounter various challenges. These obstacles include limited access to financing, inadequate collaboration between academia and industry, and a scarcity of skilled labour. Nonetheless, case studies have consistently demonstrated that firms with higher innovation potential are more inclined to engage in innovative activities and achieve superior performance outcomes. Addressing gaps in the existing literature regarding the assessment of innovation potential in Estonian firms becomes imperative. For instance, the development of comprehensive and accessible methods to assess innovation stages is necessary, particularly for small and medium-sized businesses that may lack the resources for expensive and extensive professional innovation audits. The Innovation Ladder Model was specifically designed to assist such businesses, and this research endeavour will contribute to the validation of its current version using publicly available data sourced from the Community Innovation Survey 2018 in Estonia. This collaborative effort with the Innovation Ladder Working Group 2022 aims to refine the model and enhance its effectiveness in guiding businesses toward improved innovation outcomes. By leveraging these research findings, the aim is to foster a more conducive environment for innovation, driving economic growth, and empowering Estonian firms to thrive in an increasingly competitive global market.

3. Methodology and Data

In the Methodology and Data section, we outline the methodology and data employed in this study to address the research question. First, we briefly mention the research design of the thesis while addressing each step of the thesis and its outputs. Second, we address which and why specific dataset we used to reach the aim of the thesis. Third, we explain the logic behind selecting specific questions from the CIS 2018 survey dataset and the categorization of questions. Fourth, we point out scientific methods, the K-means Clustering algorithm, we employed during ranking companies and validating Innovation Ladder Model. Lastly, we describe two objectives of the thesis, which are ranking companies by applying the K-means clustering algorithm to the data and validating the Innovation Ladder Model developed by the Innovation Ladder Working Group.

3.1. Research Design

This study employs a quantitative research design to decode the innovation ecosystem in Estonia using the CIS 2018 survey dataset. The primary objectives are to (1) rank companies by applying the K-means clustering algorithm to the data and (2) validate the Innovation Ladder Model developed by the Innovation Ladder Working Group by applying the K-means clustering algorithm. We applied the following steps during the research:

1. Selecting questions from the CIS 2018 survey dataset, categorising questions for pre-determined categories, and then ranking companies in each category by applying the K-means clustering algorithm.

2. After ranking each company in different categories, we classified companies with the Innovation Ladder Model by applying the K-means clustering algorithm on the ranked companies dataset we created during the first step.

There are two outputs of these objectives. First, we created ranks for companies in each category that measure different aspects of the innovation capabilities of companies. Second, we calculated how many companies there are in each step of the Innovation Ladder Model, validating the model itself and based on the Innovation Ladder model we provide sector analysis..

3.2. Data and Analysis

The data used in this study is obtained from the Community Innovation Survey (CIS) 2018 survey dataset. To conduct our analysis, we obtained data from Statistikaamet, a public organisation in Estonia that provides country statistics (About Us | Statistikaamet, n.d.). Our data version was finalised during the STI Working Group meeting in November 2017 and considered final as of 20 December 2017.

The dataset gathers Innovation-related data from 1049 businesses in Estonia and their response to 237 questions. CIS is the reference survey on Innovation in enterprises and is conducted every two years among EU member countries, EFTA countries, and candidate countries in the European Union. The current standard of questions, responses, and quality of results of CIS are established by EU regulation 995/2012, which also provides the legal framework for the survey (*Community Innovation Survey - Microdata - Eurostat*, n.d.). From Table 1, we can see the categories of the questions and their brief descriptions.

Table 1
Categories and Descriptions of Questions

Category	Description
Identification	Name, Address, Postal Code, Main Activity and Registration Code.
Strategies	Product offerings, deliverables, strategies, and annual sales. Determine the overall strategy of the business
Knowledge	Copyrights, patenting, licensing and how the business generates/acquires the knowledge.
Product Innovation	New/improved product/service offerings. Differentiation from competitors and contribution to annual turnover?
Innovation	Degree of new products introduced any predefined innovative products/services—employees' contributions and expectations from these innovations.
Research & Development	Assess the company's expenditure on research & development, funding opportunities availed, and partnerships formed within the country, EU and EFTA.
Obstacles	Different obstacles the business faced and the degree to which it could handle them quickly or not?
Enterprise	The last section concerns the questions related to the business and its nature.

Source: Community Innovation Survey (CIS) 2018

The CIS is one of the most relevant datasets for our study as CIS is designed to gather information about the innovative competence of an economy, the most innovative sectors, the significant issues businesses face in being innovative, and the use of local/EU funding programs to drive innovation (Community Innovation Survey - Microdata - Eurostat, n.d.) which are the most critical factors of the Innovation Ladder Model. The questionnaire consists of both mandatory and voluntary questions. The mandatory questions include the nature of the business, strategies for being innovative, products and services, expenditure on innovation, patent and trademark applications and other essential factors. As the CIS is self-reported survey has almost all the variables from our knowledge that affects the innovation of a business and our model is also a self analysis tool which business can use to analyse their innovation stage, this harmony between the CIS data and our model makes the CIS data suitable for our validation analysis. The voluntary questions cover topics such as environmental impact and workforce management which is another part of the Innovation Ladder Model to analyse the work innovation culture and how it is recognized within the business. The dataset is preprocessed and cleaned to ensure data quality and relevance. Data preprocessing involves removing incomplete or inaccurate records, dealing with missing values, and standardising data formats. From Table 2, we can see which treatment method was used for missing values in the dataset.

Table 2
Responds for Questions and Missing Value Treatment Methods

Response	Range	Example	Missing Value Treatment Method
Nominal	0,1	Yes (1): Organization was part of CIS 2016	Mode
Ordinal	1,2,3,4	Least Important (1) - Most Important (4)	Mode
Interval	0-100	% of expenditure in R&D	Mean
Ratio	0-∞	Annual Turnovers was 600k EUR	Mean

Source: authors' calculations

Data analysis, including data preprocessing, clustering, and validation, is conducted using statistical software packages R programming, with relevant libraries and packages for data manipulation, clustering, and validation.

3.3. Categorization of Questions

Innovation, though lacking a universally accepted definition, is widely acknowledged as a crucial determinant of business success in today's rapidly evolving markets (Crossan & Apaydin, 2010). Firms that demonstrate adaptability and innovative tendencies often exhibit superior growth and longevity (Tidd & Bessant, 2020). Essential prerequisites for fostering innovation within organisations include strategic clarity, a supportive organisational culture, unique value propositions, efficient operational processes, and adequate financial resources (Chesbrough, 2006; Teece, 2007). Additionally, openness to collaboration and partnership is seen as a means to stimulate innovation and foster organisational growth (Chesbrough, 2003). To evaluate the levels of innovation within an organisation, it is crucial to understand the multidimensional nature of innovation (Adams, Bessant & Phelps, 2006). In the context of the Community Innovation Survey (CIS) 2018, analysing the data can be facilitated by categorising the questions based on the aspect of innovation they are designed to measure. For instance, some questions might delve into the company's investments in innovation activities. In contrast, others might explore the novelty of the innovations produced, and still, others might probe into the company's collaborative relationships for innovation (Oslo Manual, 2018). These categorizations offer distinct perspectives on the firm's innovation performance, providing a comprehensive understanding of their innovative capabilities." We chose questions from the CIS 2018 survey dataset that related to each category to rank companies in each category that measure different aspects of companies' innovation capabilities. In total, we have six different categories for questions, and each group represents a critical dimension of innovation, as presented in Table 3.

Table 3
Question Categories and Number of Questions

Categories	**Definition**	**No. of Question**
Strategy, Management and Ambition	Identifying whether the management is reactive, tactical, systematically focused, strategically long-term, or globally ambitious helps tailor an innovation plan that aligns with the company's vision and provides a clear roadmap for future growth.	30
Culture and People	A company's culture and people significantly impact innovation. Understanding if innovation stems from individual interests, leaders, or team involvement is vital. Fostering an innovative environment encourages creativity and welcomes fresh ideas.	4
Novelty of Value Proposition	Assessing the novelty of a company's value proposition enables it to outpace competition and create distinctive, competitive products or services.	11
Process and Practice	Innovation involves generating exceptional ideas and implementing effective processes to realize them. Assessing a business's understanding of innovation drivers and existing management measures, such as diversified models or ISO standards, ensures sustainable, repeatable, and scalable innovation efforts.	16
Finance	Innovation necessitates investment, and identifying funding sources is critical for a company's innovation activities. Well-defined financial plans ensure a company has the resources to realize its innovative ideas.	34

Partnership and Open Innovation	Collaboration is another crucial driver of innovation, and it is crucial to determine whether a company is open to collaboration with other businesses for specific use cases or has partnered with the research institute, global projects, and universities.	82

Source: Innovation Ladder Model (2022)

3.4. K-means Clustering Approach

The K-means clustering algorithm is applied to each question category to rank the companies and group companies based on the Innovation Ladder Model by validation model during the second step. First, we will explain the K-means clustering algorithm with reasons suitable for our thesis and discuss why other clustering methods, factor analysis, and regression models are not suitable for the study. K-means clustering is an unsupervised machine learning algorithm widely used for partitioning data into distinct, non-overlapping groups (Jain, 2010). The algorithm iteratively assigns data points to clusters based on similarity, minimising the within-cluster variance and maximising the between-cluster variance (MacQueen, 1967). The K-means clustering algorithm is chosen for this study due to several key attributes that make it suitable for the research objectives. As an iterative algorithm, K-means is computationally efficient, scalable, and capable of handling large datasets, such as the CIS 2018 survey data (Jain, 2010). It is a centroid-based method, which groups data points based on the Euclidean distance between them, ensuring that each point is closer to its cluster centre than other clusters (Lloyd, 1982). K-means is well-suited for the type of data in this study As it operates on the principle of minimising variance, it is effective for continuous numerical data, such as the scores derived from the survey responses (Hartigan & Wong, 1979). Moreover, K-means is ideal for studies where the number of clusters is predetermined or guided by theoretical considerations. In this case, the number of clusters aligns with the categories in the Innovation Ladder Model. Despite its strengths, K-means has limitations. It is sensitive to the initial placement of centroids and may converge to local optima. However, strategies such as the K-means++ initialization can help mitigate these issues (Arthur & Vassilvitskii, 2007). While other clustering methods like hierarchical clustering and DBSCAN were considered, they were not chosen for several reasons.

Hierarchical clustering, while providing a rich structure, is more computationally intensive, especially for large datasets, and does not allow for the reassignment of points once a step in the hierarchy is made (Murtagh & Legendre, 2014). DBSCAN, on the other hand, does not require specifying the number of clusters but is sensitive to the parameters for neighbourhood size and minimum points, making it less straightforward to apply in this context (Ester et al., 1996). While Factor Analysis and Regression analysis are powerful statistical techniques, they are not the optimal choice for this study due to their inherent properties and the specific objectives of this research. *Factor Analysis* is a data reduction technique used to identify a data set's underlying structure (Fabrigar et al., 1999). It is generally used when the goal is to identify underlying dimensions, or factors, that explain the correlations among a set of variables. However, this study aims not to identify underlying dimensions but to cluster companies based on their responses to a predefined set of categories. The research is not trying to simplify or reduce the data structure but to find influential groups within the data. Therefore, there would need to be a more appropriate application of Factor Analysis. Regression analysis, on the other hand, is a predictive modelling technique, and it is used to estimate the relationships among variables (Freedman, 2009). It examines the influence of one or more independent variables on a dependent variable. In this study, the goal is not to predict a specific outcome based on a set of predictors but to explore the data's multi-dimensional structure and classify companies into groups. Hence, regression analysis would not fulfil the study's objectives.

Working with the K-means clustering algorithm on the CIS 2018 survey dataset requires attention to two crucial points. The first important point is determining the optimal number of clusters for each category of questions. To determine the optimal number of clusters for each category, five validation metrics were employed: average silhouette score (Rousseeuw, 1987), Davies-Bouldin index (Davies & Bouldin, 1979), Calinski-Harabasz index (Calinski & Harabasz, 1974), elbow method (Kodinariya & Makwana, 2013) and the within-cluster sum of squares (WSS) method. These validation metrics were selected based on their ability to assess cluster validity and cohesion and their widespread use in the literature. Each of these validation techniques has its strengths and weaknesses. The silhouette score evaluates cluster cohesion and separation, with higher values indicating better clustering quality (Rousseeuw, 1987). The Davies-Bouldin index measures the average similarity between clusters, with lower values indicating better separation (Davies & Bouldin, 1979). The Calinski-Harabasz index evaluates the ratio between cluster to within-cluster variance, with higher values indicating better clustering (Calinski & Harabasz, 1974). Lastly, the elbow method involves plotting the within-cluster sum of squares against the number of clusters and selecting the "elbow point" as the optimal number (Kodinariya & Makwana, 2013). The second important point is the visualisation of clustering. In this study, we used Principal Component Analysis (PCA) to visualise the clusters formed by the K-means algorithm. Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into values of linearly uncorrelated variables called principal components (Jolliffe, 2002). This transformation is defined in such a way that the first principal component has the most considerable possible variance, and each succeeding component has the highest variance possible under the constraint that it is orthogonal to the primary components. The high-dimensional data in this study, resulting from the multiple survey question categories, can be challenging to visualise. However, visualisation is crucial for exploring and understanding the data structure, including cluster formation (Kandogan, 2001). PCA comes in handy as it reduces the dimensionality of the data, making it possible to visualise the data in a two-dimensional or three-dimensional space (Wold, Esbensen, & Geladi, 1987). By projecting the data onto the first few principal components, PCA can reveal the structure of the data in a way that best explains their variance. In this study, PCA can help visualise the clusters formed by the K-means algorithm, providing a way to examine the separation between clusters and assess the clustering quality (Ringnér, 2008). Therefore, PCA is an essential tool for this study because it allows for the visualisation and interpretability of the high-dimensional data, facilitating an understanding of the clustering results.

In conclusion, K-means clustering, an unsupervised learning algorithm, is more suitable for the study. It is designed to identify clusters or groups in the data based on the similarity of responses across multiple dimensions (Jain, 2010). This aligns with the research objective of categorising and ranking companies based on their responses to multiple survey questions across various innovation dimensions because of the K-means clustering algorithm's efficiency, applicability to the data type and structure, and alignment with the research objectives.

3.5. Ranking Clusters and Validation of Innovation Ladder Model

As we mentioned, the primary objectives are to (1) rank companies by applying the K-means clustering algorithm to the data and (2) validate the Innovation Ladder Model developed by the Innovation Ladder Working Group by applying the K-means clustering algorithm. After selecting relevant questions from the CIS 2018 survey dataset, we applied the K-means clustering algorithm for the first objective while choosing how many clusters would be in each question category with five validation metrics. In each category of questions, we have a different number of clusters. We analyse the responses of companies to multiple survey questions in each cluster, and we decide on which cluster represents better performance in each related category, which are Strategy, Management, and Ambition; Culture & People; Novelty of Value Proposition; Process and Practice; Finance; Partnership & Open Innovation. After applying the K-means clustering algorithm, the average response score for each cluster in each question category is calculated by summing the companies' responses and taking the mean value. Higher average response scores indicate better performance in the respective category. The average response scores for all clusters in each category are compared to determine the best-performing cluster in each category. The cluster with the highest average response score performs best in that specific category. Then, in the final step of the first objective, companies are ranked based on their performance in each category.

For the second objective, validating the Innovation Ladder Model by calculating how many companies will be at each step of the ladder, we applied K-means clustering on data we created during the first objective. The Innovation Ladder Model is still in development and not public under any licensee, thus does not have any mentions in research papers. In this model, businesses are classified into five different categories, each reflecting their innovation level, starting with those with Inspiration, Basic Knowledge and Motivation, Processes and Organisational Integration, R&D and Investment Intensity, and Global Impact and Scope. Here is a brief description of each category, answering what it represents:

A. Interested in Innovation: A business that is interested in innovation but needs to be driven internally within the organisation. In this category, the business may need a formal innovation strategy, and innovation is separate from the work culture promoted at all levels of work. The business may need a clearer view of its value proposition, and no specific funding may be allocated for R&D projects.
B. Experimenter with Innovation: Businesses are open to pursuing some specific ideas. However, it needs the structure of people/teams to take these ideas forward. Innovation is restricted to quite a low number of product and process categories. The business has basic knowledge and motivation to innovate and can get funding. However, the business may need a formal innovation strategy or a holistic view of its value proposition.
C. Driven process Innovator: The business has processes and organisational involvement with a medium-term view of its strategy, and the general work atmosphere promotes innovation for its team/employees. It has a holistic view of its value proposition, has basic know-how to get funds, and, most importantly, is open to partnerships and

cooperation with other businesses for specific use cases. In this category, the business has a formal innovation strategy and has implemented an innovation management model across the organisation.

D. Strategic Innovator: Strategic Innovators are businesses that have an R&D and investment intensity focus. A formal innovation strategy and innovation are part of the work culture promoted at all levels of work. The business has a strong value proposition and a competitive advantage for its product and services. The innovation management model is applied across the organisation, and the business can raise external funds for its growth. The business is open to collaboration with other research organisations and global institutes. In this category, the business has a high level of investment in R&D and Innovation.

E. Ecosystem Designer: Business with a global impact, a role model for the industry. The business has a strategy with a vision to change the world. Continuous in-house innovation projects are carried out, and everyone at all levels is involved in innovation projects. The value proposition is globally distinguished, and the business has spearheaded a competitive advantage. The business has a certified ISO Innovation management system. Businesses can publicly raise funds from specific deep tech funds and continuously collaborate with organisations, research institutes, and peer businesses. The business has created a common shared platform for innovation. In this category, the business has a high level of investment in R&D and Innovation and has a significant global impact and scope.

As the model has five categories, we have added a dimension - "Not interested in Innovation." This additional category recognizes that not all companies are engaged in the innovation process or have innovation as a priority. By adding this dimension, we acknowledge the reality of the business landscape, where not all companies are on the innovation ladder, and provide a more comprehensive view of the innovation ecosystem. While the traditional Innovation Ladder Model provides a framework to assess and classify companies based on their innovative activities, it inherently assumes that all companies are actively engaged in some level of innovation. This assumption, however, might only hold for some companies, especially when considering a diverse and broad dataset like the CIS 2018. Companies may not prioritise innovation for various reasons. For example, resource constraints can be one of the reasons. Some companies, especially smaller ones or those in highly competitive markets, might prioritise immediate survival and profitability over longer-term innovation (Freel, 2005). They might need more financial resources, time, or human capital to invest in innovation. From another perspective, focusing on operational efficiency can be another reason. Companies may focus on refining and optimising their existing processes and offerings rather than investing in new, untested, potentially risky innovations (Porter, 1996).

In summary, this study employs a quantitative research design using the K-means clustering algorithm to analyse and rank companies in Estonia's innovation ecosystem based on the CIS 2018 dataset. The results will validate the Innovation Ladder Model, providing insights into the performance of companies across various innovation dimensions and offering a comprehensive understanding of Estonia's innovation ecosystem. The methodology detailed above ensures a rigorous and systematic approach to data analysis, allowing for meaningful interpretation of the findings and their implications for policy and practice.

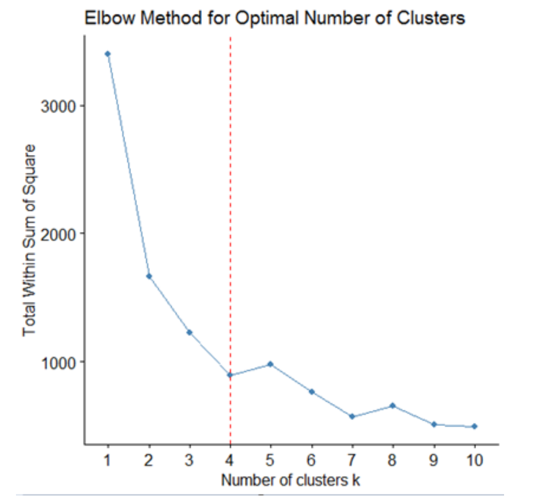
4. Empirical Results

This chapter presents the empirical results derived from the analysis of the survey dataset, as outlined in the methodology section. The primary objectives are to (1) rank companies by applying the K-means clustering algorithm to the data and (2) validate the Innovation Ladder Model developed by the Innovation Ladder Working Group by applying the K-means clustering algorithm. For the first objective, we identify patterns and relationships of companies based on their response to the questions we categorise: Strategy, Management, and Ambition; Culture & People; Novelty of Value Proposition; Process and Practice; Finance; Partnership & Open Innovation. Questions for each category measure different aspects of the companies' innovation capabilities. We applied the K-means clustering approach separately for each category. The validation metrics—average silhouette score, Davies-Bouldin index, Calinski-Harabasz index, and the elbow method—were employed to determine the optimal number of clusters for each category. First, we present the clustering analysis results for each category, including the optimal number of clusters identified and a description of the clusters' characteristics in the empirical results section.

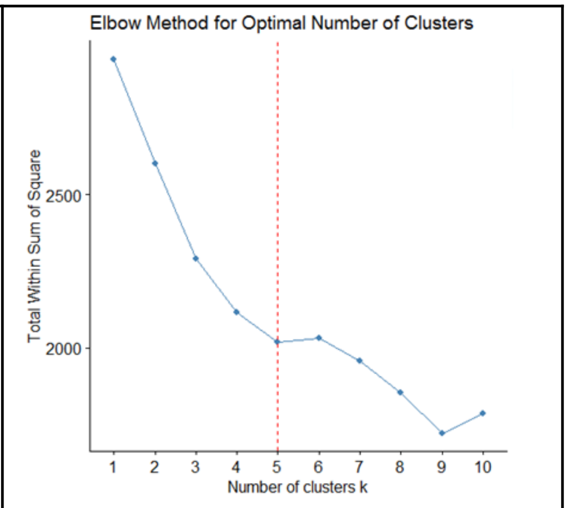
4.1 Ranking Companies

In the first step, we start with clustering companies based on their responses to each group of questions. The primary objective is to identify patterns and relationships in the data that reflect the surveyed companies. By analysing these patterns, we aim to understand better the different strategic behaviours and preferences among the companies in the dataset. This can ultimately contribute to a more nuanced understanding of the factors influencing business success and organisational performance. By clustering the companies based on their responses to questions in this category, we can identify distinct groups of companies that share similar behaviour within each category. We used four techniques to decide the number of clusters in each category. From Appendix B, we can see the four cluster validation method results for the clustering during the first objective of the thesis.

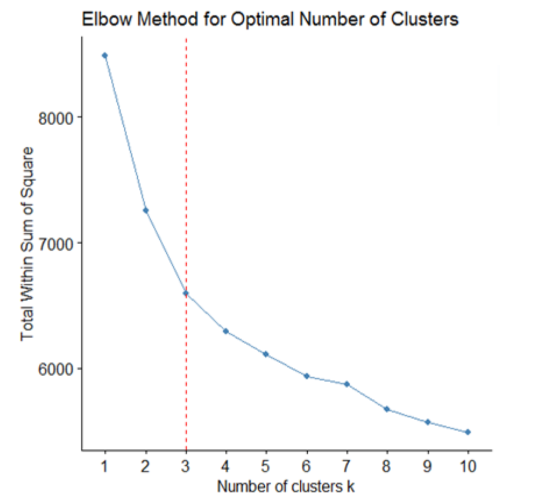
After these four validation methods, we used the elbow method to visualise the cluster-choosing process. It provides us with the ideal number of clusters for each category of questions. Figure 1 shows the optimal number of clusters for each group of questions to rank companies using the elbow method. We combine the elbow method with four techniques: average silhouette score, Davies-Bouldin index, Calinski-Harabasz index, elbow method and the within-cluster sum of squares (WSS) method.



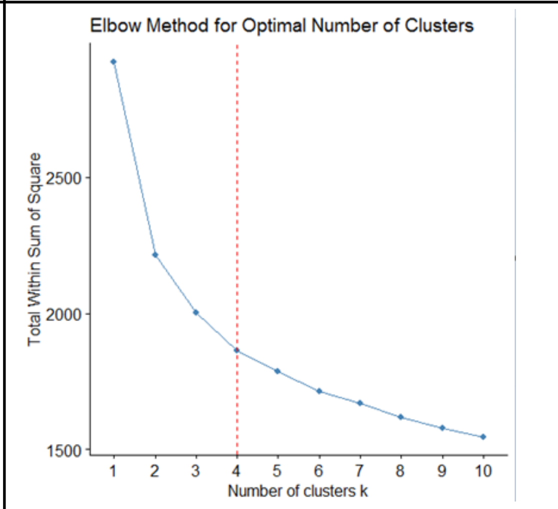
Culture and People



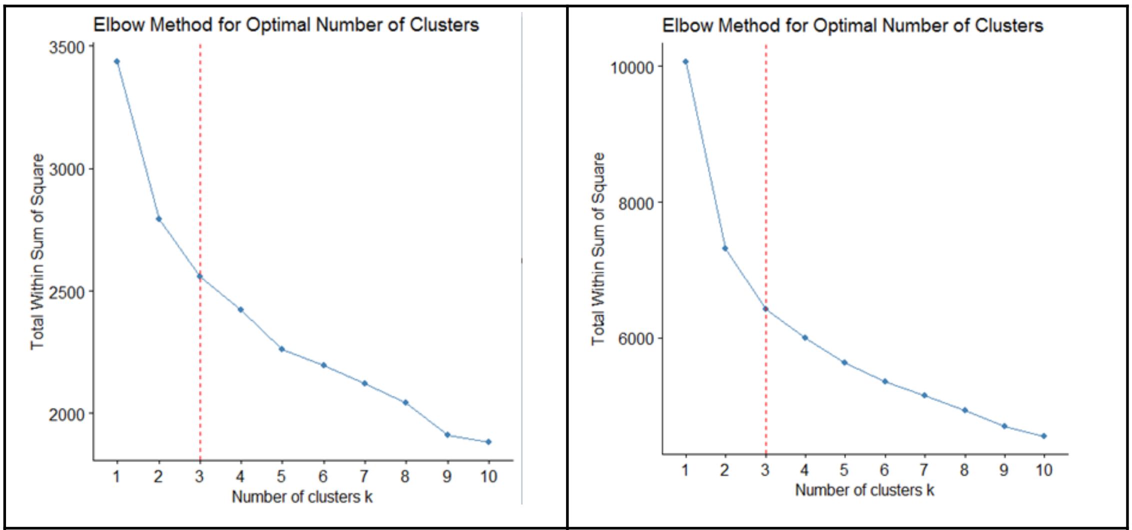
Finance



Partnership & Open Innovation



Process and Practice



Strategy, Management and Ambition	Novelty of Value Proposition

Figure 1: Elbow Methods for Optimal Number of Clusters
Source: Authors' calculations

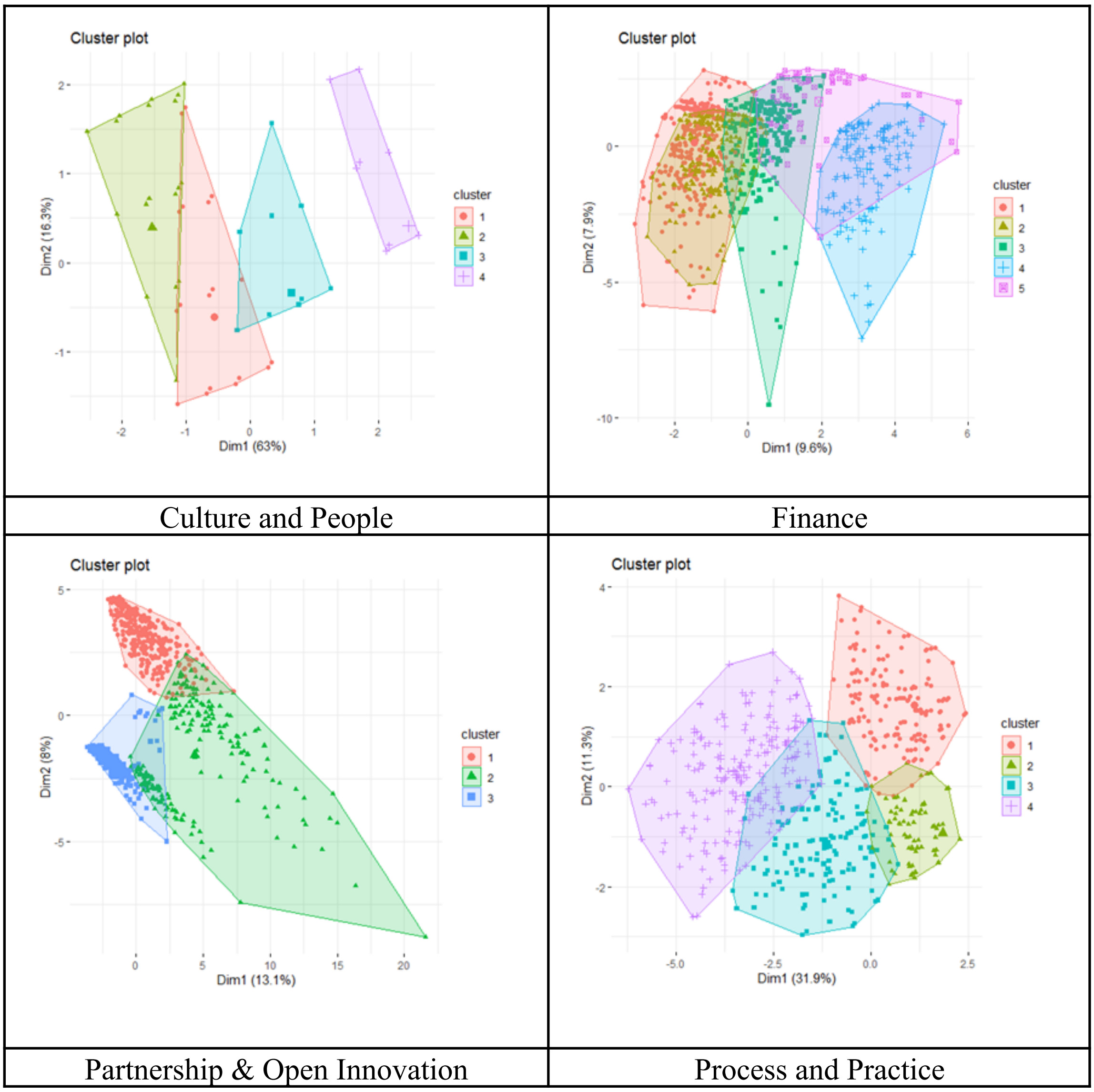
We combine the elbow method with four techniques: average silhouette score, Davies-Bouldin index, Calinski-Harabasz index, elbow method and the within-cluster sum of squares (WSS) method and choose the optimal value for k, which represents the number of clusters. Table 5 summarises the determined number of clusters for each category.

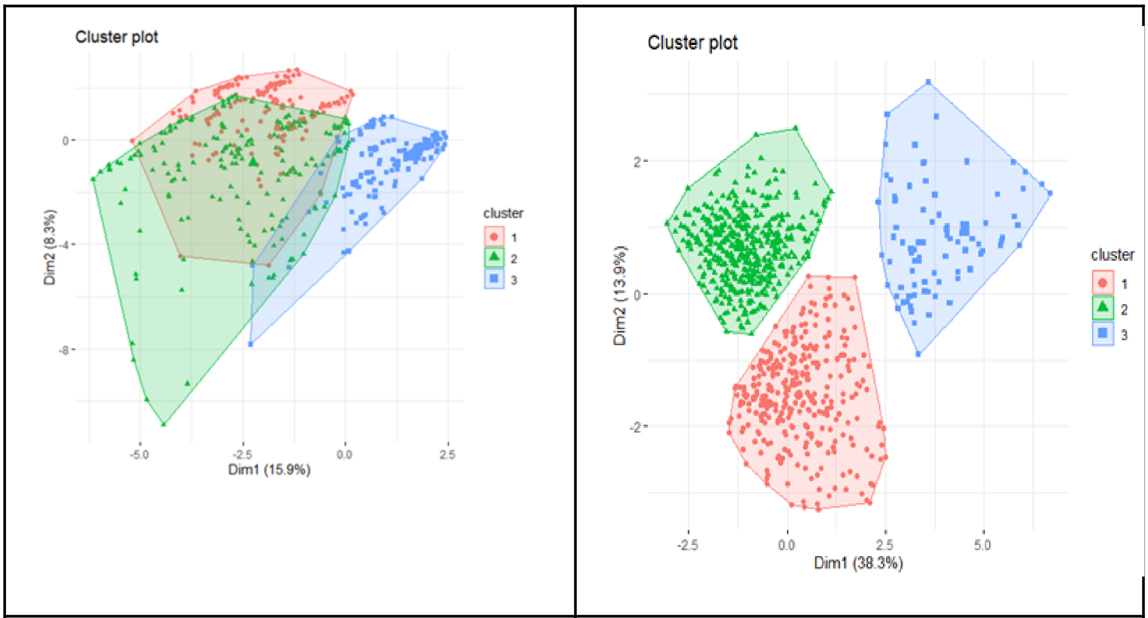
Table 5
Number of Clusters for Each Category

Category	Number of Clusters
Strategy, Management and Ambition	3
Culture & People	4
Novelty of Value Proposition	3
Process and Practice	4
Finance	5
Partnership & Open Innovation	3

Source: authors' calculations

After identifying the optimal clusters for each category, we applied the K-means clustering approach separately. From Figure 2, we can see the distribution of points. However, Principal Component Analysis has been applied to the data for visualisation. The main focus is on the distribution of points and the separation between clusters. In this context, the axes represent the data's first principal components (PC1 and PC2) after applying Principal Component Analysis (PCA). PCA is a dimensionality reduction technique that transforms the original, high-dimensional data into a lower-dimensional space while preserving as much variance as possible. The first principal component (PC1) captures the most significant variation in the data and is represented on the x-axis of the plot. The second principal component (PC2) captures the second-largest amount of variation, orthogonal to PC1, and is represented on the y-axis. By visualising the data this way, we can better understand the structure and relationships between clusters. As we can see from Figure 2, there are overlaps among the clusters, which is one of the expected results. The main reason is that we evaluated and clustered the companies for specific categories.





Strategy, Management and Ambition	Novelty of Value Proposition

Figure 2: Clusters for Each Category of Questions
Source: authors' calculations

Each category and each cluster represents the levels for companies. For instance, we can assume there are 5 clusters for the finance category for companies and businesses in the overlapped zones that could be in a transitional state, moving from one innovation pattern to another. After applying the K-means clustering algorithm, we need to determine which cluster represents the better level for companies. From Table 6, we can see the average response of companies for each group of questions grouped by clusters.

Table 6
Average Response Score of Clustered Companies to Specific Categories

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Strategy, Management and Ambition	6.72	6.98	8.60		
Partnership & Open Innovation	4.76	15	22.83		
Novelty of Value Proposition	6.28	17.64	23.27		
Culture & People	0.33	3.27	3.85	6.85	
Process and Practice	1.82	3.82	6.24	9.97	
Finance	6.45	6.81	7.38	7.81	7.95

Source: authors' calculations

We calculated the average response score as follows: We have ordinal and binary variables, primarily for companies, and we grouped companies by their cluster group for each question category. Their answers are summed and taken as mean values. If the company gave a higher response for the questions in each category, it was labelled as the highest cluster. We ranked the companies based on the average response score for each cluster and each category.

For instance, Cluster 3 has the highest average response score in Strategy, Management and Ambition category. That is why companies that are included in this group have the best performance in this strategy category.

4.2 Validation Innovation Ladder Model

After deciding which cluster represents the best performance in the relevant category, we continue validating the Innovation Ladder Model, its results and sector analysis. As the model has five categories, we have added a dimension - "Not interested in Innovation". With this additional dimension, the Innovation Ladder Model has six dimensions. We applied the K-means clustering algorithm on the ranked companies' data and clustered companies into six as we have six dimensions in the Innovation Ladder Model. Figure 4 shows the distribution of points, elbow method visualisation, and the separation between clusters for the clustering algorithm we applied to determine Innovation Ladder categories for Estonian companies.

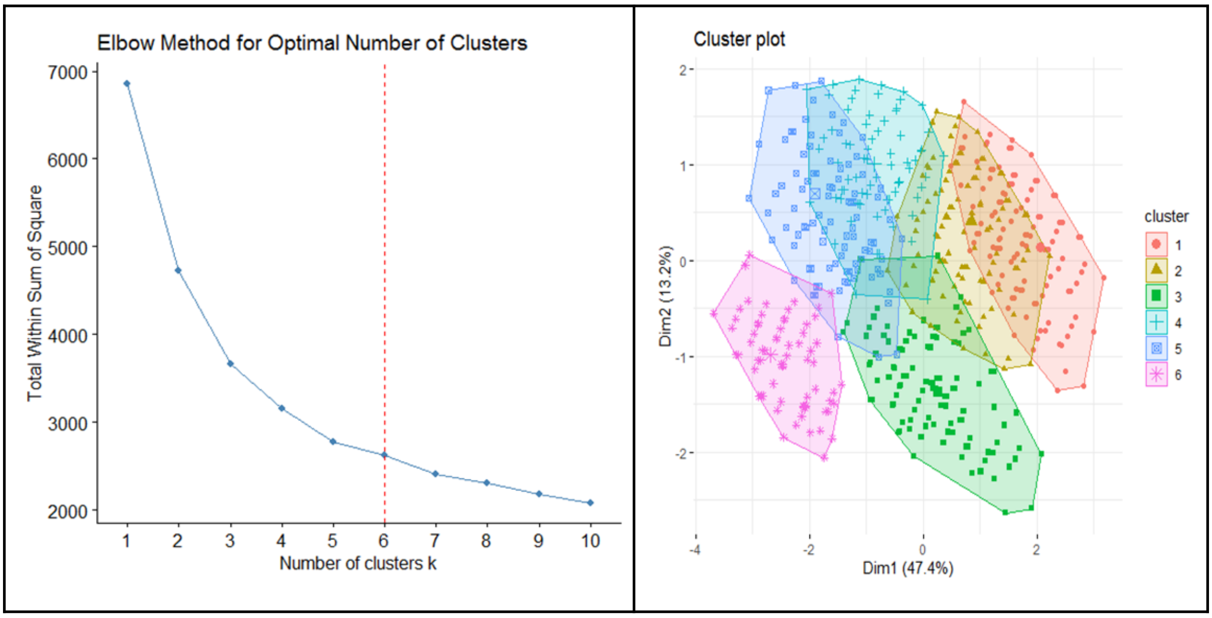


Figure 4: Elbow Method and Cluster Plot for Innovation Ladder Model
Source: authors' calculations

As we stated the k value before applying the K-means clustering algorithm, we used the Elbow method visualisation to verify that the optimal k value was selected for the clustering. From Table 7, we can see the average rank score of companies. We calculated the average rank score for clusters as follows: ranks (clusters label) of companies for each category obtained during the first objective for each cluster summed and taken mean value. For instance, cluster 6 represents the highest average rank score. As we can see, the companies in cluster 6 have the best performance or high rank for each category: Strategy, Management, Ambition; Culture & People; Novelty of Value Proposition; Process and Practice; Finance; Partnership & Open Innovation.

Table 7
Average Rank Score of Companies

Clusters	Average Rank Score
Cluster 6	19.47
Cluster 5	16.55
Cluster 4	14.42
Cluster 3	13.93
Cluster 2	11.76
Cluster 1	9.02

Source: authors' calculations

Based on this outcome and assumption, we assign each cluster to one of the Innovation Ladder categories. While companies in Cluster 6 were assigned to the Ecosystem Designer step, companies in Cluster 1 were assigned to the Not Interested Innovation step. We can see the descriptive statistics for Innovation Ladder based on our outcome from Table 8.

Table 8
Innovation Ladder for Estonian Companies

Innovation Ladder	**Number of Businesses**	**Percentage**
Not Interested in Innovation	267	26%
Interested in Innovation	209	21%
Experimenter with Innovation	146	14%
Driven Process Innovator	94	9%
Strategic Innovator	169	17%
Ecosystem Designer	134	13%
Total	1019	-

Source: authors' calculations

4.3 Sector Analysis

We used the NACE code from the CIS 2018 dataset for sector analysis. From Figure 5, we can see the frequency of companies with the Innovation Ladder Model dimension based on the main activities performed by the company.

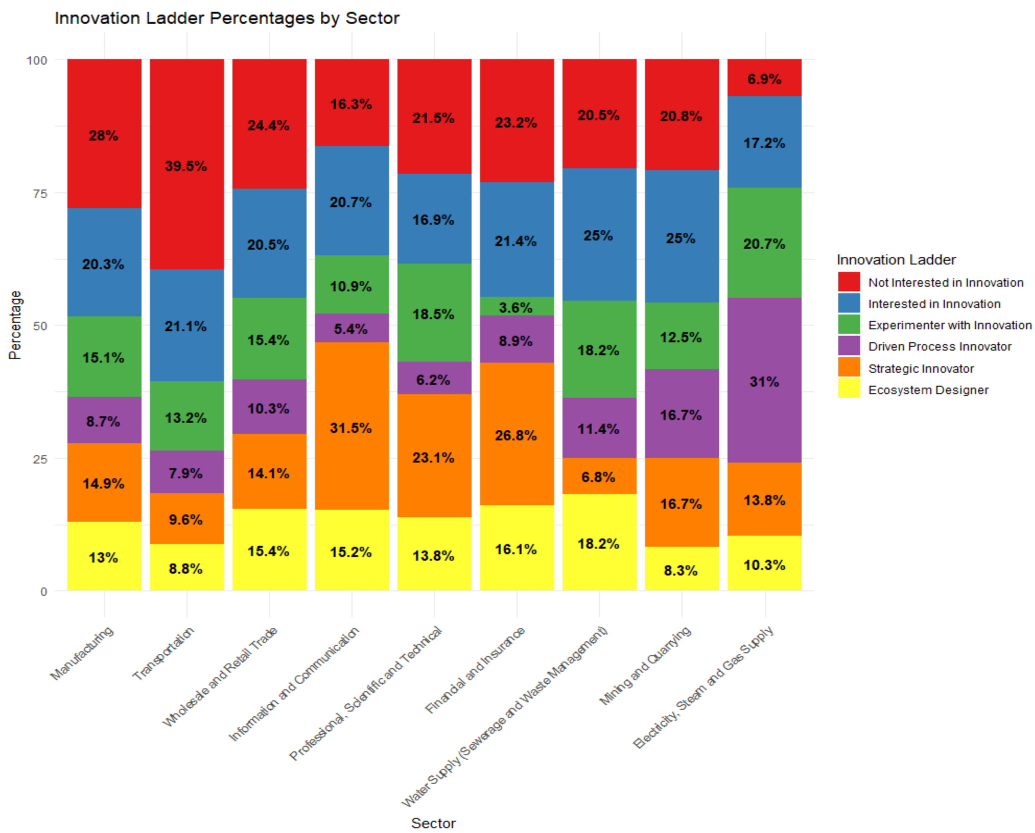


Figure 5: Innovation Ladder Percentages by Sector
Source: authors' calculations

NACE, an acronym for the French term "Nomenclature statistique des activités économiques dans la Communauté européenne," represents the Statistical Classification of Economic Activities in the European Community. NACE is a hierarchical system, organised on four levels, developed to categorise and standardise the representation of economic activities across the European Union (Eurostat, 2008).

As shown in Figure 5, the distribution of interest and engagement with innovation across different sectors reveals valuable insights into industry behaviour. Manufacturing exhibits the highest level of total involvement across all categories of the innovation ladder (n=517), followed by Transportation (n=114) and Wholesale and Retail Trade (n=78). This dominance of Manufacturing may be attributed to the sector's inherent reliance on technological advancements and process efficiencies, often driven by innovation (Porter & Heppelmann, 2014). However, examining the data in more detail reveals nuanced differences in innovation engagement across sectors. For instance, despite having lower overall participation (n=65), the Information and Communication sector shows the highest percentage of Strategic Innovators (31.5%). This could suggest that organisations in this sector are more inclined to focus on strategic, impactful innovations (Chesbrough, 2003). On the other hand, despite its overall high participation, the Manufacturing sector has a comparatively lower percentage of Strategic Innovators (14.9%) and a higher proportion of entities Not Interested in Innovation (28.0%). This discrepancy might suggest a potential innovation gap in this sector, with many entities needing to fully leverage the benefits of innovative practices (Schumpeter, 1942). Interestingly, sectors such as Water Supply and Electricity, Steam, and Gas Supply, which have lower total participation (n=24 and n=29, respectively), show a balanced distribution across the innovation ladder. Such patterns indicate a more evenly distributed innovation culture within these sectors, potentially due to regulatory or environmental pressures fostering innovative approaches (Ostrom, 1990). In conclusion, our analysis provides a comprehensive view of the innovation landscape across different sectors. It highlights the need for tailored strategies and policies to foster innovation, considering each sector's unique characteristics and challenges (Rogers, 2010).

Figure 6 shows our analysis results based on the Innovation Ladder Model; as a business becomes more innovative, it moves from market dependent to competitive to the market creator.

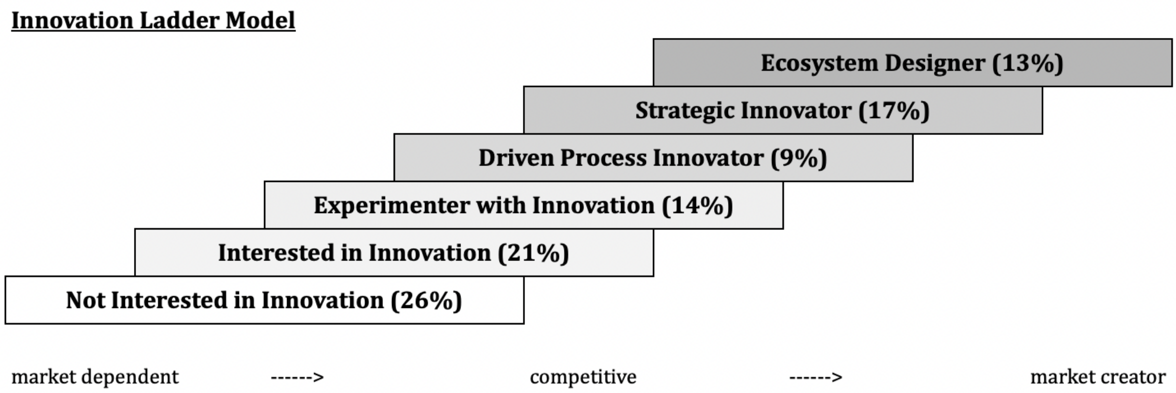


Figure 6: Innovation Ladder Model for CIS 2018 participant businesses in Estonia
Source: authors' calculations

Our analysis of businesses participating in the Community Innovation Survey 2018 revealed some interesting findings. The study showed that 26% of the businesses were "Not Interested in Innovation" and carried out their activities without introducing new products and services. This lack of innovation could put these businesses at risk of being outperformed by competitors or technological advancements. While 21% of the businesses were "Interested in Innovation" and had introduced at least one new product or service in the past two years. However, these businesses still needed more resources and motivation to implement innovative strategies fully.

The study categorised 14% of the businesses as "Experimenters with Innovation." These businesses had the basic knowledge of innovation and carried motivation but could not leverage partnership networks with other businesses and organisations. Next, 9% of the businesses in the analysis were classified as "Driven process Innovator." These businesses had successful business models and were considered knowledgeable efficiency seekers. Additionally, they were able to use partnership networks, though still in the initial stage of networking, and these networks can be driven by requirement rather than willingness.

17% of the businesses were identified as "Strategic Innovators." These businesses had a leading-edge business model, knowledge-intensive product development, and access to enough funding sources. They were also able to utilise government-sponsored programs for Innovation Management. Finally, 13% of the businesses were considered "Ecosystem Developers." These businesses met almost all of the parameters described in the study's methodology, from distinguished product positioning to ongoing R&D projects and access to internal and external finances. What set these businesses apart from the Strategic Innovators was their network of partners and emphasis on knowledge sharing with other companies.

As a result, the clustering analysis provided valuable insights into the different levels of innovation among businesses. The study showed that some businesses needed more resources and motivation for innovation. In contrast, others had successful models and access to funding sources, which enabled them to implement innovative strategies fully.

5. Conclusion

This study aims to validate the Innovation Ladder Model developed by the Innovation Ladder Model Working Group 2022. The model assists in identifying the various stages that businesses go through when implementing innovation policies, using data from the Community Innovation Survey 2018. Our analysis focused on data from Estonian businesses, as the current version of the model is intended for their use. Our objective was to assess the applicability of the current version of the model to real-world data, identify any gaps in the model, and provide recommendations to the Innovation Ladder Model Working Group.

To accomplish this, we categorised the questions from the Community Innovation Survey into the relevant categories, discarding generic questions. We then employed the Elbow method to determine the optimal number of clusters for each group and applied K-means clustering to classify businesses into different stages of innovation. The analysis revealed a precise classification of companies across the various stages of the Innovation

Ladder Model. However, only a few companies displayed a strong inclination towards innovation, primarily due to the costs and efforts associated with fostering an innovative culture within their organisations.

While the current version of the Innovation Ladder Model mandates the execution of all factors within a specific group, we have identified certain limitations. For example, some businesses may fulfil the Novelty of Value Proposition parameters, which are essential for being classified as Ecosystem Designers. However, they may still need to meet additional requirements under Partnership & Openness. Through our analysis and a comparison with other models used by industry professionals, we have identified several ways to enhance the model's alignment with small and medium businesses. These proposed changes aim to provide flexibility to the targeted audience expected to utilise the model while maintaining its classification framework.

A. Proposed Changes for Improved Classification:

   The proposed changes in the model aim to enhance how companies are classified, providing greater flexibility for the targeted audience expected to use the model.

   a. Relaxation on Rigid Boundaries:

      The model should consider businesses for a higher stage even if they fulfil most parameters but miss a couple. This approach acknowledges that although they may still need to reach the higher stage, they are on the path.

   b. Weighted Score Method:

      Assigning weights to different parameters and sub-parameters would ensure a fairer classification, as certain factors hold more importance than others. However, implementing this weighted score method adds complexity as the model must be redesigned with an index scale.

B. Minor Alterations for Enhanced Adaptability:

   The following proposed changes do not require significant alterations to the core methodology of the model but make the current version more adaptable.

   a. Not Interested in Innovation:

      The current version of the model needs more criteria for companies not interested in innovation. When analysing the general population of businesses, non-innovative companies will require a feature or policy to embark on the path of innovation. A category titled "Not Interested in Innovation" should be included to accommodate such companies. Our analysis confirms this need, as

26% of companies do not fit any existing categories and fall into this "Not Interested in Innovation" classification.

b. Raising money from the stock exchange through public listing of shares:
The criterion of raising funds from the public market through stock exchange listing should be removed from the category of Ecosystem Designers. A business can possess a dedicated research and development facility, partnerships with other organisations, and government program participation while still privately held and not raising funds from the public market. For instance, a small pharma research company can exhibit these characteristics. Removing this criterion allows for a more inclusive classification.

c. Innovation management ISO applied:
While the Innovation Management ISO provides a structured approach to enhancing innovation capabilities and competitiveness, complying with its rigid guidelines can be challenging for small organisations and businesses. Even for genuinely innovative organisations, this can pose a significant obstacle. Therefore, considering this as a good feature would be more appropriate instead of an essential requirement.

The revised version of the model is more adaptable as shown in Appendix C, by reconsidering these rigid boundaries and introducing flexibility for businesses to move across lateral categories based on their current practices, the model will provide better classification, enabling companies to understand their current stage, identify areas for improvement, and prioritise their following innovation parameters accordingly. This will support companies in designing their policies to foster innovation effectively.

6. List of References

1. About Us; Lean Business Ireland. (2022, September 23). Lean Business Ireland. Retrieved April 19, 2023, from https://www.leanbusinessireland.ie/about-us-2/
2. About us | Statistikaamet. (n.d.). Retrieved May 17, 2023, from https://www.stat.ee/en/statistics-estonia/about-us
3. Adams, R., Bessant, J., & Phelps, R. (2006). Innovation management measurement: A review. International Journal of Management Reviews, 8(1), 21-47.
4. Alhusen, H., Bennat, T., Bizer, K., Cantner, U., Horstmann, E., Kalthaus, M., Proeger, T., Sternberg, R., & Töpfer, S. (2021). A New Measurement Conception for the 'Doing-Using-Interacting' Mode of Innovation. *Research Policy*, *50*(4), 104214. https://doi.org/10.1016/j.respol.2021.104214
5. Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, 1027-1035.
6. Boly, V., Morel, L., Assielou, N. G., & Camargo, M. (2014). Evaluating innovative processes in french firms: Methodological proposition for firm innovation capacity evaluation. *Research Policy*, *43*(3), 608–622. https://doi.org/10.1016/j.respol.2013.09.005
7. Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. Communications in Statistics, 3(1), 1-27.
8. Cavdar, S. C., & Aydin, A. D. (2015). An Empirical Analysis about Technological Development and Innovation Indicators. *Procedia - Social and Behavioral Sciences*, *195*, 1486–1495. https://doi.org/10.1016/j.sbspro.2015.06.449
9. Chesbrough, H. (2003). Open Innovation: The New Paradigm for Understanding Industrial Innovation. In Chesbrough, H., Vanhaverbeke, W. & West, J. (Eds.), Open Innovation: Researching a New Paradigm. Oxford University Press.
10. Chesbrough, H. (2006). Open innovation: The new imperative for creating and profiting from technology. Harvard Business Press.
11. Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly, 35(1), 128-152.
12. *Community innovation survey - Microdata - Eurostat*. (n.d.). Eurostat. Retrieved April 19, 2023, from https://ec.europa.eu/eurostat/web/microdata/community-innovation-survey

13. Crossan, M.M. & Apaydin, M. (2010). A Multi-Dimensional Framework of Organizational Innovation: A Systematic Review of the Literature. Journal of Management Studies, 47(6), 1154-1191.
14. Damanpour, F. and Aravind, D. (2012) Managerial Innovation: Conceptions, Processes, and Antecedents. Management and Organization Review, 8, 423-454.
15. https://doi.org/10.1111/j.1740-8784.2011.00233.x
16. Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, (2), 224-227.
17. Eamets, R., Philips, K., & Randma-Liiv, T. (2014). Innovation policy in small transition economies: the case of Estonia. Journal of Baltic Studies, 45(4), 473-489.
18. Eesti ettevõtetest ligi kolmveerand on innovaatilised | Statistikaamet. (2020, May 25). Retrieved March 25, 2023, from https://www.stat.ee/et/uudised/2020/05/25/eesti-ettevotetest-ligi-kolmveerand-on-innovaatilised
19. Ehin, P., Solvak, M., Willemson, J., & Vinkel, P. (2022). Internet voting in Estonia 2005–2019: Evidence from eleven elections. Government Information Quarterly, 39(4), 101718. https://doi.org/10.1016/j.giq.2022.101718
20. Enterprise Ireland Client Services – Lean Business Ireland. (2020, September 3). Lean Business Ireland. Retrieved April 19, 2023, from https://www.leanbusinessireland.ie/funding-supports-overview/are-you-an-enterprise
21. European Innovation Scoreboard. (2021). European Commission.
22. Eurostat. (2018). Community Innovation Survey 2018 (CIS 2018) [Data set]. European Commission.
23. Eurostat (2008). NACE Rev. 2 - Statistical classification of economic activities in the European Community. Eurostat Methodologies and Working Papers.
24. Ester, M., Kriegel, H.P., Sander, J., Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231.
25. Freel, M. S. (2005). Patterns of innovation and skills in small firms. Technovation, 25(2), 123-134.

26. Gunday, G., Ulusoy, G., Kilic, K., & Alpkan, L. (2011). Effects of innovation types on firm performance. International Journal of Production Economics, 133(2), 662-676.
27. Hartigan, J. A., & Wong, M. A. (1979). A k-means clustering algorithm. Applied statistics, 28(1), 100-108.
28. Hogeforster, M. (2014). Future Challenges for Innovations in SMEs in the Baltic Sea Region. Procedia - Social and Behavioral Sciences, 110, 241–250. https://doi.org/10.1016/j.sbspro.2013.12.867
29. Innovation360 Group. (2021, October 16). Innovation Assessment | Innovation Capability Assessment Tools. Innovation360. Retrieved April 19, 2023, from https://innovation360.com/offers/innovation-assessment/
30. Innovation360 Group. (2023, March 15). Innovation Assessment with InnoSurvey® - Free test and report. Innovation360. Retrieved April 19, 2023, from https://innovation360.com/innovation-assessment/
31. Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651-666.
32. Jolliffe, I. (2002). Principal component analysis. Springer.
33. Kandogan, E. (2001). Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 107-116.
34. Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. International Journal, 1(6), 90-95.
35. Lööf, H., & Heshmati, A. (2006). On the relationship between innovation and performance: a sensitivity analysis. Economics of Innovation and New Technology, 15(4-5), 317-344.
36. Lloyd, S. P. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129-137.
37. MacQueen, J. (1967, June). Classification and analysis of multivariate observations. In 5th Berkeley Symp. Math. Statist. Probability (pp. 281-297). Los Angeles LA USA: University of California.
38. Ministry of Economic Affairs and Communications (2014). Estonian Entrepreneurship Growth Strategy 2014-2020.
39. Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?. Journal of Classification, 31(3), 274-295.
40. National Innovation Scoreboard. (2021). European Commission.

41. OpenAI. (2022). ChatGPT (Dec 20 version), large language model, https://chat.openai.com/
42. Oslo Manual (2018). Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition. The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris/Eurostat, Luxembourg.
43. Ostrom, E. (1990). Governing the commons: The evolution of institutions for collective action. Cambridge University Press.
44. Porter, M. E. (1996). What is strategy? Harvard Business Review, 74(6), 61-78.
45. Porter, M. E., & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. Harvard Business Review, 92(11), 64-88.
46. Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology. *Administrative Science Quarterly*, *41*(1), 116. https://doi.org/10.2307/2393988
47. Ringnér, M. (2008). What is principal component analysis?. Nature biotechnology, 26(3), 303-304.
48. Rogers, E. M. (2010). Diffusion of innovations. Simon and Schuster.
49. Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy*, *94*(5), 1002–1037. https://doi.org/10.1086/261420
50. Reila, B. (2021, November 1). International reputation survey: Estonia seen as the global leader in digitalisation - Estonia. estonia.ee. https://estonia.ee/estonia-as-the-global-leader-in-digitalisation/
51. Results of the community innovation survey (CIS2-CIS7). (n.d.). Eurostat. https://ec.europa.eu/eurostat/cache/metadata/en/inn_cis2_esms.htm
52. Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53-65.
53. Schumpeter, J. A. (1934). The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle. Harvard University Press.
54. Schumpeter, J. A. (1942). Capitalism, socialism and democracy. Routledge.
55. Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, *15*(6), 285–305. https://doi.org/10.1016/0048-7333(86)90027-2
56. Teece, D. J. (1998). Capturing value from knowledge assets: The new economy, markets for know-how, and intangible assets. California Management Review, 40(3), 55-79.

57. Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. Strategic Management Journal, 28(13), 1319–1350.
58. Tehnopol. (2022a, January 14). *Innovation Health Check audit | Sihtasutus Tallinna Teaduspark Tehnopol*. Sihtasutus Tallinna Teaduspark Tehnopol. Retrieved April 19, 2023, from https://www.tehnopol.ee/innovation-health-check-audit/
59. Tehnopol. (2022, September 8). *About us | Tallinn Science Park Tehnopol*. Tallinn Science Park Tehnopol. Retrieved April 19, 2023, from https://www.tehnopol.ee/en/about/
60. Tidd, J., & Bessant, J. R. (2020). Managing innovation: integrating technological, market and organizational change. John Wiley & Sons
61. Tiits, M., Kalvet, T., & Mürk, I. (2015). Smart Specialisation in Cohesion Economies. *Journal of the Knowledge Economy*, *6*(2), 296–319. https://doi.org/10.1007/s13132-015-0239-6
62. Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), 37-52.
63. *World Bank Open Data*. (n.d.). World Bank Open Data. Retrieved April 19, 2023, from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=EE

7. Appendices

Appendix A

Innovation Ladder Model Methodology

Category	Interested in Innovation	Experimenter with Innovation	Driven Process Innovator	Strategic Innovator	Ecosystem Designer
Partnership & Open Innovation		*Existing cooperation partners*	*Cooperation partners selected for the purpose of innovation*	*Research institutions and global cooperation projects*	*Open innovation, partners contribute to "the center of the ecosystem"*
			Open innovation	Innovation partnership	A platform for innovative partners and customers
		Cooperation capabilities	Primary cooperation partners focused on innovations	Continuous innovation-oriented cooperation	Collaboration platform for innovation, initiator of cooperation and commercializer of results
Finance		*Self-financing*	*Local subsidies*	*International finance*	*Public offering / listing*
	Investors	Personal contacts (FFF)	Business angels of early stage ideas	VC Funds	Raising money from the stock exchange through public listing of shares
	Local subsidies / International funds	Primary support (inno-vouchers)	Local process and product development subsidies	International grants for applied research and product development	Global Deep-TECH Funds
	Self-finance / Loans	Random spending (reactive)		In addition to digitization and automation, the R&D budget is >2% of turnover	R&D >10% of turnover
Process & Practice		*Basic knowledge*	*First agreements of innovation process management*	*Innovation management model applied*	*Open innovation*
			Innovation management ISO	Created and implemented the innovation management model of your company	Innovation management ISO applied
		Management of innovation portfolio	Consolidated view on innovation projects	Projects divided into 3 horizons of innovation, managed as a whole	Shared / segmented innovation portfolios with separate budgets and management
		Management of ideas	Idea collection in minimum once a year and methodology of market testing of ideas	Regular (2-4 times a year) publicly visible idea collection, validation and market testing	Continuous idea collection, transparent testing
	Product development	Basic experience with product development	Client-centered and systemic product development	Knowledge intensive product development	R&D activities based on unique basic technologies
Novelty of Value Proposition	*Search / noticing inspiration*	*Innovations in some/single categories*	*Holistic view on value proposition*	*A strongly differentiated value proposition and a protected competitive advantage*	*Globally distinctive competitive advantage / uniqueness*
	Business model / Sales channel	Formulated business model	Successful business model	A leading edge business model	Unique business model
	Technology	Experimenting with existing technologies	Some novelty in technology	Strong technological innovation	DeepTech
	Product / IO	Improvement	Own product	Protected own product, recognized brand	A product is synonymous with a product category
	Process	Reactive adaptation of work processes	Knowledgeable efficiency seeker about automation and digitalization	Proactive process innovator based on constant product/service innovations	Benchmark in updating processes
Culture & People	*Individual interested persons*	*Single/some leaders*	*Broader team Invovlvement*	*Innovation culture*	*An example of innovation*
		Culture	General supportive attitude	Innovation is part of the company's culturet	An entrenched culture of innovation
	People	Single/some leaders	Innovation related positions	A dedicated innovation team	Everyone is involved in the innovation process
Strategy, Management & Ambition	*Interested owners and managers*	*Initiative to pursue a specific idea*	*Readiness for a stystemic view*	*Formalised innovation strategy*	*Platform / Consolidator Ambition*
	Reactive Management	Tactical Management (1-2yrs)	Strategic Management (2-5yrs)	Innovation strategy is part of business strategy, time horizon 5-10 years	A development strategy based on a world-changing vision

Source: Innovation Ladder Model Working Group 2022

Appendix B

Clustering Validation Methods Result

	WSS	Average Silhouette Score	Davies Bouldin Index	Calinski Harabasz Index
Strategy, Management and Ambition				
k = 3	2556	0.17	0.24	174.4
k = 4	2379	0.18	0.24	150
k = 5	2250	0.18	0.24	133.3
Culture and People				
k = 3	1221	0.39	0.22	903.8
k = 4	891	0.43	0.23	950.1
k = 5	731	0.47	0.23	923.87
Process and Practice				
k = 3	2002	0.20	0.27	233.2
k = 4	1860	0.19	0.28	193.2
k = 5	1770	0.19	0.28	164.9
Novelty of Value Proposition				
k = 3	6422	0.19	0.12	287.6
k = 4	6002	0.18	0.13	228.7
k = 5	5624	0.12	0.13	199.8
Finance				
k = 3	2242	0.15	0.26	157.6
k = 4	2114	0.14	0.26	131.9
k = 5	1996	0.14	0.04	119.6
Partnership & Open Innovation				
k = 3	6598	0.17	0.13	145.2
k = 4	6289	0.13	0.13	118.1
k = 5	6090	0.14	0.14	99.7

Source: authors' calculations

Appendix C

Revised Version of Innovation Ladder Model

Category	Not Interested in Innovation	Interested in Innovation	Experimenter with Innovation	Driven Process Innovator	Strategic Innovator	Ecosystem Designer
Partnership & Open Innovation			*Existing cooperation partners*	*Cooperation partners selected for the purpose of innovation*	*Research institutions and global cooperation projects*	*Open innovation, partners contribute to "the center of the ecosystem"*
				Open innovation	Innovation partnership	A platform for innovative partners and customers
	Partnerships & Collaborations are seen as threats		**Cooperation capabilities**	Primary cooperation partners focused on innovations	Continuous innovation-oriented cooperation	Collaboration platform for innovation, initiator of cooperation and commercializer of results
Finance			*Self-financing*	*Local subsidies*	*International finance*	~~***Public offering / listing***~~
		Investors	Personal contacts (FFF)	Business angels of early stage ideas	**VC Funds**	
		Local subsidies / International funds	Primary support (inno-vouchers)	Local process and product development subsidies	International grants for applied research and product development	Global Deep-TECH Funds
	Avoiding R&D Expenses	**Self-finance / Loans**	Random spending (reactive)		In addition to digitization and automation, the R&D budget is >2% of turnover	R&D >10% of turnover
Process & Practice			*Basic knowledge*	*First agreements of innovation process management*	*Innovation management model applied*	*Open innovation*
				~~**Innovation management ISO**~~	**Created and implemented the innovation management model of your company**	
			Management of innovation portfolio	Consolidated view on innovation projects	Projects divided into 3 horizons of innovation, managed as a whole	Shared / segmented innovation portfolios with separate budgets and management
			Management of ideas	Idea collection in minimum once a year and methodology of market testing of ideas	Regular (2-4 times a year) publicly visible idea collection, validation and market testing	Continuous idea collection, transparent testing
	Negative Approach for New Process & Practies	**Product development**	Basic experience with product development	Client-centered and systemic product development	Knowledge intensive product development	R&D activities based on unique basic technologies
Novelty of Value Proposition		*Search / noticing inspiration*	*Innovations in some/single categories*	*Holistic view on value proposition*	*A strongly differentiated value proposition and a protected competitive advantage*	*Globally distinctive competitive advantage / uniqueness*
	Lack of differentiated USP	**Business model / Sales channel**	Formulated business model	Successful business model	A leading edge business model	Unique business model
		Technology	Experimenting with existing technologies	Some novelty in technology	Strong technological innovation	DeepTech
		Product / IO	Improvement	Own product	Protected own product, recognized brand	A product is synonymous with a product category
		Process	Reactive adaptation of work processes	Knowledgeable efficiency seeker about automation and digitalization	Proactive process innovator based on constant product/service innovations	Benchmark in updating processes
Culture & People		*Individual interested persons*	*Single/some leaders*	*Broader team Involvement*	*Innovation culture*	*An example of innovation*
	Avoid Innovative Ideas		**Culture**	General supportive attitude	Innovation is part of the company's culturet	An entrenched culture of innovation
		People	Single/some leaders	Innovation related positions	A dedicated innovation team	Everyone is involved in the innovation process
Strategy, Management & Ambition	***Non Interested owners***	*Interested owners and managers*	*Initiative to pursue a specific idea*	*Readiness for a stystemic view*	*Formalised innovation strategy*	*Platform / Consolidator Ambition*
	Non Reactive To Market Changes	Reactive Management	Tactical Management (1-2yrs)	Strategic Management (2-5yrs)	Innovation strategy is part of business strategy, time horizon 5-10 years	A development strategy based on a world-changing vision

Source: Author's revision to Innovation Ladder Model

Resümee

Decoding the Innovation Ecosystem in Estonia:

A Clustering Approach to the Innovation Ladder Model

Abhishek Giri Emre Kaan Sarikaya

Eesti akadeemikutest ja professionaalidest koosnev innovatsiooniredeli töörühm 2022 töötab aktiivselt välja innovatsiooniredeli mudelit (ILM) kui avatud lähtekoodiga metoodikat ettevõtte innovatsiooniseisundi hindamiseks, mille eesmärk on aidata ettevõtetel välja selgitada vajalikud poliitikad, protsessid ja koostöö partnerid, et võimaldada nende loomingulist väljundit.

Käesolev magistritöö aitab valida innovatsiooniredeli mudelit 2018. aasta CIS (ingl *Communinty Innovation Survey*) Eesti andmete põhjal. Andmestikust valiti asjakohased andmed ja need rühmitati vastavalt innovatsiooniredeli mudeli mõõtmetele. K-keskmiste algoritmide abil koondati ettevõtted nende vastuste järgi rühmadesse, mille põhjal oli võimalik hinnata nende innovaatilisust igas redeli mõõtmes. Seejärel kasutati taas klasteranalüüsi jagamaks Eesti ettevõtted erinevatesse innovatsiooni etappidesse, et valideerida nnovatsiooniredeli mudelit. Uuringus esitati ka sektorianalüüs, mis põhines mudeli raamistikul. Kokkuvõttes koondati ettevõtted innovatsiooni eri etappidesse CIS 2018 andrete kohaselt. Selle järgi anti ILM töörühmale nõuandeid, kuidas praegune mudeli võiks olla reaalsete ettevõtete jaoks paremini kohandatav.

Non-exclusive licence to reproduce thesis and make thesis public

Abhishek Giri
18.May.2023

Non-exclusive licence to reproduce thesis and make thesis public

I Emre Kaan Sarikaya herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis
Decoding the Innovation Ecosystem in Estonia: A Clustering Approach to the Innovation Ladder Model supervised by Kadri Ukrainski

I grant the University of Tartu a permit to make the work specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

I am aware of the fact that the author retains the rights specified in points 1 and 2.

I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Emre Kaan Sarikaya
18.May.2023