Exposing the ideal endogenous-exogenous variables´ combination for a company to foster its eco-innovative action. Results from machine-learning methods.

Ángel Peiró-Signes, anpeisig@omp.upv.es
Management Department, 7D building, Universitat Politecnica de Valencia, Camino de Vera, s/n, 46022 Valencia

Marival Segarra-Oña, maseo@omp.upv.es
Management Department, 7D building, Universitat Politecnica de Valencia, Camino de Vera, s/n, 46022 Valencia

Óscar Trull-Domínguez, otrull@eio.upv.es
Applied Statistics Department, 7A building, Universitat Politecnica de Valencia, Camino de Vera, s/n, 46022 Valencia

Joaquín Sánchez-Planelles, joasanpl@ade.upv.es
PhD School, Management Program, 2E building, Universitat Politecnica de Valencia, Camino de Vera, s/n, 46022 Valencia

Exposing the ideal endogenous-exogenous variables´ combination for a company to foster its eco-innovative action. Results from machine-learning methods.

Abstract

Although many drivers have been commonly considered to impact firms' eco-innovation, our study demonstrates that, truly, few aspects are relevant.
The accuracy of the model and the large and complete spectrum of innovative companies in the sample contributes to the generalizability of the results. This is particularly relevant because the main drivers of firms' eco-innovative orientation depend on firm's innovative behaviour, which indicates that the managerial and policy work has to be directed to raise awareness of the different externalities derived from innovation. On one side, policy regulations should continue to pressure firms with environmental standards. On the other side, managers can stimulate the creation of a corporate innovative culture oriented to improve operational efficiency (reducing unnecessary costs), to improve their workplace environment and to focus on new customer demands which, in essence, will guide the organization to be more environmentally and socially responsible.

Keywords: eco-innovation, drivers, innovative firms, machine learning

1. Introduction

Literature has defined eco-innovation in different ways (Kemp and Pearson, 2007; Eco-Innovation Observatory, 2012; Horbach et al., 2012, Carrillo-Hermosilla et al., 2010). Fussler and James (1996) were one of the firsts authors to research on the topic and defined eco-innovation as the process of developing new products, processes or services which provide customer and business value but significantly decrease environmental impact. Although the topic is still a dynamic concept. Eco-innovation has centered a lot of attention of academics in the last two decades. On one side, because of the synergies that this approach gives to companies to create and/or maintain competitive advantages and, parallelly, allowing firms to be more environmentally friendly and, therefore, more sustainable in the eyes of an increasing number of committed customers (Adams et al., 2012; Bocken et al., 2014). On the other side, because of the social and institutional pressure for a sustainable economy. Eco-innovation has been a major target driving the policy agenda, especially in the European Union (EEA, 2014; OECD, 2011). Indeed, EU policy pointed innovation and particularly, eco-innovation, as a major strategy to ensure environmental, social and economic sustainability within the Europe 2020 framework.
In this topic, research of the drivers that motivate companies to eco-innovate has been very extensive (Horbach, 2008; CarrilloHermosilla et. al., 2009; Triguero, Moreno-Mondéjar, & Davia, 2013; Díaz-García et al. 2015; Horbach, 2016; Hojnik & Ruzzier, 2016). Both, internal and external factors, have been addressed to be significant triggers of companies turning to be more environmentally responsible through innovation (del Río, 2009; Carrillo-Hermosilla et al., 2009; Demirel & Kesidou, 2011; Cainelli et al., 2015; Sáez-Martínez, Díaz-García, & Gonzalez-Moreno, 2016). However, most of the studies have focused on a small number of features to study eco-innovation or they studied the impact of a certain characteristic on the eco-innovation. For example, some studies don't consider the four types of eco-innovations (Marcon et al., 2017). Others restrict their study to a specific industry (Segarra-Oña et al., 2011; Negny et al., 2012; González-Moreno et al., 2013), to a specific group of companies (Horbach et al., 2012;

Triguero et al., 2013; Horbach, 2016; Kiefer et al., 2018) or to few case studies (Del Río et al. 2016). Many are also limited by relatively small samples (Saez-Martínez et al., 2016; Garcia-Granero et al., 2020) and/or on a specific driver of the eco-innovation, like the cooperation with Universities (Arroyave et al., 2020) or the level of engagement on internal R&D, e.g. R&D investment or personnel (Ghisetti & Pontoni, 2015), limiting the generalizability of the results.

Thus, further research is needed for managers and policy makers to understand what characterizes the environmental orientation of the companies while innovating, increasing the knowledge about the drivers of eco-innovation and their relative impact. This knowledge can be useful to orientate the policy measures that the European countries and European Union are undertaking to promote or facilitate the development and diffusion of these practices in the companies, and especially important due to the increasing importance of eco-innovation as a source of competitive advantage for companies. Likewise, it is also important due to the need of a more sustainable economy in terms of their environmental, social and economic contribution.

This study analyzes internal and external factors that have been identified in the literature to affect the environmental orientation of a firm when innovates. We go further than previous studies by using a whole set of relevant firm features and eco-innovation drivers with a new methodological approach. We prove this methodological approach to be more efficient than traditional multivariate statistical techniques. It also helps to focus on all relevant features and how they affect the environmental orientation. Thus, in this paper we propose the following research questions:

1.- What are the key features determining innovative firms to be environmentally oriented when innovating?

2.- Do these features act with the same importance depending on the environmental orientation of the firm?

3.- Are there firms characteristics, such as size, industry or the type of company, that might justify different eco-innovative behaviour and, therefore, different policy measures?

4.- How do the relevant features interact in reaching a specific environmental orientation level?

We apply machine learning techniques to answer these questions in a sample of 4518 Spanish innovative companies from the Spanish Technological Innovation Panel (PITEC) for 2016 (last year with available data). The sample includes firms from all types of industries, including industries with a considerable impact on the environment in terms of the resources they consume or the residues they produce as automotive or chemical companies.

This paper contributes with a novel approach using state of the art machine learning techniques to discover all-relevant features and uncover their role in driving environmental orientation of the firms while innovating. We tested the relative impact of industry (type of industry) and firm specifics (size, income, investment, etc.) on this determination. Additionally, we provide a company profile for each degree of environmental orientation that can be used to promote the eco-innovation more efficiently within companies. To our knowledge, there is no previous research combining this large set of features in the same study to evaluate their relative importance in determining the firms' orientation to eco-innovation.

We structured the paper as follows. In section 2 we present the theoretical background and the hypothesis, including the theoretical support for the variables in the model. In section 3 we describe the sample and the research methodology including the feature selection and model refinement. In section 4, we provide the detailed results of the refined model and we provide a detailed interpretation of the model results. Finally, in section 5 we present the discussion, conclusions and implications, limitations and further research options of the study.

2. Background and hypotheses

Previous research approaches to the motivations that firms have to eco-innovate suggests that they are influenced by internal and external factors. Despite, the bigger attention that external factors have received, the resource based theory and the differences in the environmental approach of firms with similar characteristics suggests that, firms' behaviour and how firms combine multiple internal factors plays a key role in turning innovators into eco-innovators.

For the internal factors, firms' eco-innovation drivers are mainly related to their internal resources and competences (Cainelli et al., 2012; Horbach et al., 2012; Triguero et al., 2013). Resources include assets, equipment and human, financial, information and knowledge resources. Among this type of drivers, literature includes, firms' environmental management (Rennings et al., 2006; Rehfeld, Rennings, & Ziegler, 2007) which relates to organizational innovations (Kesidou & Demirel, 2012; Horbach et al., 2012), the type and availability of financial resources (Keifer et al. 2018) or the innovation activity (Sáez-Martínez et al., 2016; Horbach, 2016; Castellacci & Lie, 2017), among others.

Some firm's knowledge and capabilities to innovate and, particularly, to eco-innovate are built over the time, while others have to be obtained from external agents through cooperation or collaboration (Markard & Worch, 2010). In this sense, technological push (Cleff and Rennings, 1999; Horbach, 2008) can be considered internal, if it has been a result of a technological development based on firm's technological capabilities, or external, when there are insufficient internal resources, capabilities or knowledge to develop the technology in-house.

From this perspective, previous innovation and eco-innovation activity are important for eco-innovations (Horbach, 2008; Horbach, 2016). R&D investment or R&D personnel are among the most important factors contributing to the development of innovations, and particularly, to the development of eco-innovations (Horbach, 2008; Mazzanti & Zoboli, 2009; Horbach et al., 2012; Cainelli et al., 2015). Indeed, some authors (Horbach et al. 2013; Cuerva et al., 2014) found that companies that commit to in-house innovation have greater environmental innovation. Additionally, the internal innovative and eco-innovative activity is also reflected in the formal protection of their innovations (patents) (Segarra-Oña, et al, 2011; De Marchi & Grandinetti, 2013; Sáez-Martínez et al., 2016).

However, in many cases, eco-innovations require knowledge, resources or capabilities that are not available internally. Then, it is the evolution of the technology outside the company which pushes the eco-innovation into the company. The collaboration with external technology developers, such as suppliers, universities, research centers, etc. or the competitors' actions are what it is forcing companies to eco-innovate, especially in SMEs (Saez-Martinez et al. 2014). Indeed, literature includes the firms' ability to interact with stakeholders, incorporating information and knowledge (Horbach 2008; De Marchi, 2012, del Río et al. 2015; Cainelli et al., 2015), as an important factor in the eco-innovation process. These interactions can be configured as more or less formal, and as a direct collaboration or as a net of relations. Indeed, a quite informal relation like the dependence of external information sources, has been systematically related to higher environmental orientation (Segarra-Oña et al., 2016; Segarra-Oña et al. 2017) suggesting that widening the collaboration sources is needed to address the multidimensionality of eco-innovation (Mothe & Nguyen-Thi, 2017). Taking into account the aforementioned, we state the following two hypothesis:

H1: Firms' internal engagement in innovation activities, represented by the innovation results, innovation investment, R&D investment, R&D personnel, etc., is a significant driver of the environmental orientation of the firms while innovating.

H2: Firms' external engagement in innovation activities, represented by the investment in external R&D, the active cooperation or the reliance on information from external agents in the innovation process (customers, suppliers, competitors, Universities, etc.), purchase of equipment, knowledge, etc. is a significant driver of the eco-innovation orientation.

Organizational and managerial antecedents are deemed as eco-innovation drivers (Frondel, Horbach, & Rennings, 2004; Triguero et al., 2013). In fact, Peiro-Signes & Segarra-Oña (2018) and Jove-Llopis & Segarra-Blasco (2018) demonstrated the strategic nature of the eco-innovation orientation, which is aligned with the past trajectory dependence of the eco-innovative activities indicated by other authors (Sáez-Martínez et al., 2016; Horbach, 2016; Castellacci & Lie, 2017). Attending to Mondejar-Jiménez et al. (2015), firms' environmental orientation is the managerial recognition that firm's activities have an impact on the environment. This recognition generates a need to minimize firms' environmental impact and to be socially responsible through Corporate Social Responsibility (CSR) programs (Diaz-García et al. 2015). But firms' investment in eco-innovation and reduction of environmental impact of their activity beyond complying with regulations comes at a cost (Rennings, 2000; Rennings et al., 2006) compared to their polluting competitors (Rennings et al., 2006). Then, it is firm's commitment or self-regulation in this area (Sheehy, 2015) what would be pushing company's behaviour and future prospects. Firms's behaviour relies on corporate culture and firm management (Bossle et al., 2016; Dangelico, 2016; Ortiz-de-Mandojana et al., 2016; Tsai and Liao, 2017). Corporate Social Responsibility culture is just a part of the corporate culture and it has a positive effect on eco-innovation (Rehfeld et al., 2007; Cainelli et al., 2011; Kesidou and Demirel, 2012; Doran and Ryan, 2016). Then, we can expect that thinking further than the regulations and the typical economic objectives in the innovative activities (e.g. improving health and safety workplace conditions) is just an expression of a proactive corporate culture and strategy and it will positively affect a proactive environmental orientation. Therefore, we state our hypothesis as follows:

H3: Eco-innovation orientation is positively affected by firm's orientation to other (non-environmental) CSR objectives.

External factors are related to the interaction of firms with their stakeholders (Del Río González, 2009). Literature highlights environmental regulation and policies (market demand) as crucial external factors to eco-innovation (Reinnings, 2000; Horbach, 2008;

Kesidou and Demirel, 2011; de Marchi, 2012; Horbach et al., 2013, del Río et al., 2015, Doran and Ryan, 2016), also for Spanish firms (Costa-Campi et al, 2015, Jove-Llopis & Segarra-Blasco 2018). Then, we can then state the following hypothesis:

H4: Environmental policies and regulations are a significant driver of the environmental orientation of the firms while innovating.

The external stimulus to eco-innovate is not limited to the regulatory pressure or to the engagement with external actors in collaborations. Policy makers have been promoting using subsidies certain activities to encourage innovation, and particularly, eco-innovations. Although the previous results are contradictory (Johnson and Lybecker, 2012, Jove-Llopis & Segarra-Blasco 2018), the amount of programs oriented to achieve more environmentally friendly solutions at local, regional, national and European level, suggests that companies that applied for these subsidies are more willing to eco-innovate that firms that do not apply for these programs. Then, we state the following hypothesis:

H5: Subsidies are a significant driver of the environmental orientation of the firms while innovating.

The increasing awareness of society, and particularly customers, about the environmental impact of firms' activities and their products and services, has been translated into higher pressures on the companies to introduce environmental improvements in their products or to develop new products for environmentally conscious customers (market demand). Then, firms will be looking to enter new markets or to increase their market share or looking for a wider and updated range of products or services (product orientation) to address customer environmental awareness. The research results pointed to its relevance in many studies (Veugelers, 2012; Horbach et al., 2013, Segarra-Oña et al., 2016; Segarra-Oña et al. 2017, Peiro-Signes & Segarra-Oña, 2018), however, some authors found no evidence of its significance (Del Rio et al., 2015, Jove-Llopis & Segarra-Blasco, 2018).
In addition, customer demand and the competitive pressure are forcing companies to be more efficient and flexible. The search for more efficient and cost-effective ways of producing products and services (process orientation) focusses on reducing the resources consumption (e.g. energy and materials) or on increasing the capacity and flexibility to better serve the customers. This process-oriented innovation has been proven a significant and important driver to environmental innovations (Pereira and Vence, 2012; Horbach et al., 2012; Triguero et al., 2013; Godoy-Durán et al., 2017; Peiro-Signes & Segarra-Oña 2018). We think, in line with previous studies (Segarra-Oña et al., 2016; Segarra-Oña et al. 2017), that these pressures are reflected in the orientation of the firms to introduce process and product innovations. Therefore, we state the following hypothesis:

H6: The orientation to process and to product innovation are significant drivers of the environmental orientation of the firms while innovating.

Although it has attracted less attention, some authors have highlighted the importance of green marketing and the organizational practices on the environmental performance (Marcon et al., 2017; García-Granero et al., 2020). Firms can search for new ways to place, communicate, deliver, price or promote products and services (marketing innovations) and to organize their activities, their decision-making process or their relations with their stakeholders (Organizational innovations). For example, through environmental labeling (marketing innovation) or adopting an Environmental Management System (organizational innovation). Then, we state the following hypothesis:

H7: The orientation to organizational and to marketing innovation are significant drivers of the environmental orientation of the firms while innovating.

Firms' general characteristics (firm size, human and financial resources, their export orientation, type, industry or age, among others) have been reported also as contributors or moderators of the environmental orientation of the firm (Carrillo-Hermosilla et al., 2009; Segarra-Oña et al., 2011, Segarra-Oña et al., 2016; del Río et al., 2016; Hojnik & Ruzzier, 2016). For example, large firms are reported to be more environmentally oriented than small firms (De Marchi, 2012; Costa-Campi et al., 2015; del Río et al., 2015) or manufacturing industries are reported to be more environmentally oriented than service firms (Segarra-Oña et al., 2016). On the contrary, firms belonging to a group do not show significant differences from stand-alone firms in their eco-innovation orientation (Cainelli & Mazzanti, 2013; Doran & Ryan, 2016). Other characteristics such as the export orientation (Segarra-Oña et al., 2011; Horbach et al., 2012; Guoyou, et al, 2013; Galdeano-Gómez et al., 2013; Li, 2014; Horbach, J. & Jabob, J., 2018), the age (Ziegler and Rennings, 2004; Rehfeld et al., 2007; Keskin et al. 2012; Keifer et al 2018), or the region where the company is located (Cooke, 2011; Horbach, 2014) have been also studied. Then, we state the following hypothesis:

H8: Firm's characteristics is a significant driver of the environmental orientation of the firm while innovating.

In table 1, we summarized the different variables included in the study and an extended related literature that gives support for their inclusion in the study. We also added some variables that are present in the database that might be of interest to explore their potential impact. Particularly, we incorporated the type of research that it is done by the company (basic, applied or technological), the frequency of the R&D activity (continuous or occasional), the location in a technological park. Finally, several authors (Könnölä, Unruh, & Carrillo-Hermosilla, 2006; Cainelli et al., 2011; Johnson & Lybecker, 2012; Cuerva, Triguero-Cano, & Córcoles, 2014) reported internal and external factors hindering eco-innovation. Thus, we also considered the perception of barriers to innovation as an element that might affect the innovation behaviour of the company.

Table 1. Variables in the study.

Variable (number of variables in the category)	**Type of Measure**	**Description**	**Theoretical support**
INTERNAL ENGAGEMENT			
PIDT/PIDCA (2) INVT (1) REMUSUP (1)	Cont. (Int)[1] % of R&D personnel % number of employees	Intramural R&D personnel Researchers % of employees with higher education	Horbach, 2008; Mazzanti & Zoboli, 2009; Horbach et al., 2012; Horbach, 2014; Ghisetti, Marzucchi, & Montresor, 2015; Cainelli et al., 2015
TINTID (1) GTINN (1) IDIN (1) GINTID (1) PRODPI (1)	% of innovation expenditure Cont. (Int.)[1] Dummy % of total innovation expenditure	Internal expenditure in R&D Total innovation expenditure Engagement in intramural R&D Expenditure in intramural R&D Expenditure in other Intelectual Property for intramural R&D	Horbach, 2008; Kammerer, 2009; Horbach et al., 2012; Cainelli et al., 2012; Triguero, Moreno-Mondéjar, & Davia, 2013; Horbach, 2014; Cuerva et al. (2014); del Río et al. (2015); Cainelli et al., 2015; Horbach, 2016

	% of total intramural R&D		
INNPRODi (3) INNPROCi (5) NEWEMP/OLD (2) NOVEDAD/NOVEDEMP (2) QUIENPRODi (6)	Dummy Dummy % of the turnover Dummy Dummy	Introduction of product or process innovations in the previous two years (types: goods, services, production, logistics,…) % New products vs. old products Firm introduced new products to the market/firm Good, service or process innovations developed by the company	Rothenberg and Zyglidopoulos, 2007; Horbach, 2008; De Marchi, 2012; Horbach, 2016
FONi (4)	% of internal R&D expenditure	Financial resources origin: own, other companies or foreign founds.	Johnson and Lybecker, 2012; Cuerva, Triguero-Cano, & Córcoles, 2014; Keifer et al 2018
PATi (3)	Cont. (Int.)	Total number of patents, number of European, US,… patents (in the previous 2 years)	Segarra-Oña, et al. 2011; De Marchi & Grandinetti, 2013; Saéz-Martínez et al., 2016
USOi (3)	Dummy	Utility models, trademarks,… in the previous 2 years	
INNOVATION ORIENTATION			
OBJET1-5 (5)	Cat.(0-3)[2]	Importance of product objectives in the innovation activities of the company (product innovation orientation)	Rennings, 2000; Horbach, 2008; Horbach et al. 2012; Triguero et al., 2013; Cuerva et al.; 2014; Segarra-Oña et al. 2016, Segarra-Oña et al. 2017; Godoy-Durán et al., 2017; Peiro-Signes Segarra-Oña 2018
OBJET6-10 (5)	Cat.(0-3)[2]	Importance of process objectives in the innovation activities of the company (process innovation orientation)	Pereira and Vence, 2012; Horbach et al., 2012; Triguero et al., 2013; Godoy-Durán et al., 2017; Segarra-Oña et al. 2016, Segarra-Oña et al. 2017; Peiro-Signes Segarra-Oña 2018
INORGNi (3)	Dummy	Introduction of different types of organizational innovations in the previous two years	Demirel and Kesidou, 2011; Horbach et al., 2012: Kesidou & Demirel, 2012; Cuerva, Triguero-Cano, & Córcoles, 2013; Frondel, Triguero et al., 2013; Marcon et al., 2017; Astuti et al., 2018 Wagner, 2007
INCOMNi (4)	Dummy	Introduction of different types of commercial innovations in the previous two years	Martin et al., 2015; Marcon et al.; 2017; Astuti et al., 2018; Liao, 2018; Garcia-Granero et al., 2020;
EXTERNAL ENGAGEMENT			
FUENTENEWi (12)[1]	Cat.(0-3)[2]	Importance of different information sources (internal, market, institutional and other sources) for the innovation process	Carrillo-Hermosilla et al., 2010; De Marchi, 2012; De Marchi & Grandinetti, 2013; Horbach et al., 2013; Cainelli et al., 2015; Hansen & Coenen, 2015; Segarra-Oña et al. 2016, Segarra-Oña et al. 2017; Peiro-Signes Segarra-Oña 2018
COOPERA (1) COOPNEWi (9)[1]	Dummy	Cooperation arrangements (with agent i: Other companies, suppliers, clients, competitors, university, research centers,….) on innovation activities	Horbach, 2008; Mazzanti & Zoboli, 2009; Carrillo-Hermosilla et al., 2010; Wagner & Llerena, 2011; Belin et al., 2011; Petruzzelli et al., 2011; Caineli, et al., 2012; De Marchi, 2012; Junquera et al., 2012; Hobarch et al., 2013; De Marchi & Grandinetti, 2013; Wong, 2013; Hansen & Coenen, 2015; Cainelli et al., 2015; del Río et al., 2017
CONSULi (4)	Dummy % of R&D personnel % of R&D expenditure	External consultants in intramural R&D	
QUIENPRODi(6)	Dummy	Good, service or process innovations developed in cooperation or by other firms or institutions	
GEXTER (1) IDEX / IDEXi (6) GEXTID / GEXTIDi (6)	% of total innovation expenditure Dummy	External R&D services purchase Engagement in extramural R&D and different types (acquisition of machinery, external knowledge,…) Expenditure in extramural R&D	Markard & Worch, 2010; Demirel & Kesidou, 2011; Horbach et al., 2013; Cai & Zhou, 2014

	% of total innovation expenditure		
CSR ORIENTATION			
OBJET 12 (1)	Cat.(0-3)[2]	Improve health or safety of employees	Rehfeld et al., 2007; Cainelli et al., 2011; Kesidou and Demirel, 2012; Doran and Ryan, 2016
EXTERNAL POLICY AND REGULATIONS			
FINAi (4)	Dummy	Financial support for innovative activities from local, regional, national and EU funds	Horbach, 2008; Hobarch et al 2012; De Marchi, 2012; Johnson and Lybecker, 2012; Río et al., 2015
FONPUBLI (1)	% of internal R&D expenditure	Financial resources origin: public funds	Johnson and Lybecker, 2012
OBJET 13 (1)	Cat.(0-3)[2]	Meet environmental, health or safety regulations	Rennings, 2000; Beise and Rennings, 2005; Horbach, 2008; Wagner, 2008; De Marchi, 2012; Doran & Ryan, 2016; Horbach, et al., 2012; Veugelers 2012, Kesidou and Demirel, 2012; Hobarch et al., 2012 Hobarch et al., 2013; Triguero et al. (2013; Hpbarch, 2016; del Rio et al., 2017
FIRMS' CHARACTERISTICS			
GRUPO (1)	Cat.	Firm part of a group	Cainelli et al.; 2011
RELA (1)	Cat.	Firms' relation with the group	
CIFRA (1)	Cont. (Int.)[1]	Total Income	Revell et al., 2010; Segarra-Oña et al., 2011; Hofer et al., 2012; Kesidou and Demirel, 2012
INVER (1)	Cont. (Int.)[1]	Total investment	
TAMANO (1)	Cont. (Int.)[1]	Number of employees	
MDOLOCAL, MDONAC, MDOUE, OTROPAIS (4)	Dummy	Firm's geographic markets (local/regional, national, UE, others)	Segarra-Oña et al., 2011; Horbach et al., 2012; Guoyou, et al, 2013; Galdeano-Gómez et al., 2013; Li, 2014; Horbach, J., Jabob, J., 2018
EXPORTN (1)	% of total turnover	Export volume	
INTRACOM (1)	% of total turnover	EU and related countries sales	
ACTIN (1)	Cat. (43 groups of industries)	Industry code	Segarra-Oña et al., 2011b; Negny, Belaud, Cortes Robles, Roldan Reyes, & Ferrer, 2012 Sierzchula, Bakker, Maat, & Van Wee, 2012; González-Moreno et al., 2013; Segarra-Oña et al., 2016
CLASEN (1)	Cat.	Type of company: public or private with different percentages of foreign ownership	Cainelli et al., 2011
SEDE (1)	Cat. (Madrid, Catalonia, Andalusia, rest of Spain)	Legal location of the firm	Cooke, 2011; Horbach, 2014
ANIOCREA (1)	Year	Firm's start year	Ziegler and Rennings, 2004; Rehfeld et al., 2007; Keskin et al. 2012; Keifer et al 2018
INFUN, INAPL, DESTEC (3)	% of the type of R&D	Type of research (basic, applied or technological)	
PARQUE (1)	Dummy	Located in a technological park	
IDINTERN (1) TIPOID (1)	Dummy Cat. (Continuous or occasional)	Internal R&D Frequency of the R&D	
BARRIERS			
FACEi(3) FACIi(4) OTROFACi (4)	Cat.(0-3)[2]	Importance of external, internal and other factors humpering innovation	Könnölä, Unruh, & Carrillo-Hermosilla, 2006; Cainelli et al., 2011; Johnson & Lybecker, 2012; Cuerva, Triguero-Cano, & Córcoles, 2014

1 Simulated variable. Based on an anonymization process. 2 Level of importance 3=high, 2=medium, 1=low, 0=not important/not relevant.

3. Sample and methodology

3.1. Sample

In this study we used data from the Spanish Technological Innovation Panel (PITEC) database from 2016. PITEC has been monitoring the innovation activity in Spanish companies since 2004. This harmonized survey based on the EU Community Innovation Survey and it is carried out every year. Besides descriptive variables such as, size, sales or industry, the survey includes other indicators analyzing the innovation process. It is structured in 5 areas including the objectives of the innovative companies, the innovation process structure, the public administration action, conditions that are favoring or hindering the development of innovation, the innovation results and other variables like the cooperation for innovation.

For the study, we disregarded variables that contained disaggregated data regarding the gender of the employees or the region where the company or their activities are located, which had little interest in our study. In the database, cooperation is reported attending to the stakeholder and the region where the stakeholder is located. Following previous research, we aggregated the variables related to cooperation by stakeholder. After this process, we ended with 142 variables of interest. Additionally, as the objective of the study was to evaluate what aspects can be driving the environmental interest of the companies when innovating, we limited the study to companies classified as innovators, 4518 companies from the original sample. Most of the manufacturing firms were in the chemical, pharma and plastics industries (14.4%), electric, electronic and other equipment (13,8%) or in the food industry (7.9%). Regarding services, commerce, transport and storage represented 7.1% of the sample, consulting services 6.7% and other services, the remaining 20,2%. Regarding the size, large, medium and small companies represented 31.2%, 33.8% and 35%, respectively. Table 2, shows class distribution of firms attending to their environmental orientation and firms' characteristics.

Table 2. Descriptive statistics

Clase	0	1	2	3	Total
N	1501	755	1090	1172	4518
%	0,3322	0,1671	0,2413	0,2594	
Size[1]	307,14(1376,2)	382,71(1775,77)	460,51(2514,9)	551,78(2049,27)	420,23(1943,8)
Turnover	7,06(390,2)	86,1(370,2)	129,74(659)	191,67(886)	118,9(619,7)
Investment2	2,16(18,6)	2,82(26,2)	5,23(33,2)	9,99(59,4)	5,05(37,7)
Internal R&D expenditure[2]	43,30(46,93)	59,63(42,9)	65,61(40,4)	66,97(40,18)	57,55(44,41)
External R&D expenditure[2]	4,44 (16,63)	7,88 (20,67)	8,23 (20,48)	6,41 (17,67)	6,44(18,65)
R&D personel[1]	7,44(29,73)	15,03(38,33)	20,04(67,07)	24,73(68,95)	16,23(53,89)

Average values (standard deviations) [1] Number of employees [2] M€

3.2.Methodology

The aim of the study is to identify what variables are determinant in the environmental orientation of a company while innovating. In this study, we used a machine learning approach. Machine learning algorithms are increasingly been used in many research applications (Malefors et a., 2021, Grané et al., 2021). Among the different types of algorithms (linear, non-linear and ensemble), linear ones are simpler, more intuitive and easier to interpret. However, they do not capture complex interrelations between variables that are present in many real applications. Non-linear and ensemble models overcome this issue and outperform linear models in a trade for interpretability. Then, we are looking for a method to construct a model that it is both, accurate and interpretable.

3.2.1. Metric selection

The metric is the measure by which we will evaluate and compare different models. As we want to predict class labels and we don't have an imbalanced sample we chose the accuracy score. The target variable (OBJET11) represents the how important was for the firm the reduction of the environmental impacts in its innovative activity. In other words, it indicates the environmental orientation of the firm towards the reduction of the environmental impact when innovating or the eco-innovation orientation of the firm. It has four classes, high, medium, low and not oriented. Then, we are looking for a model able to predict the class (environmental orientation) of the firm. The distribution of cases within each class and the main descriptive statistics are shown in table 2.

3.2.2. Algorithm selection

To select the machine learning algorithm, we evaluated as set of different types of algorithms (linear, non-linear and ensemble algorithms) with the default options to determine the optimal algorithm for our study. We used the most basic model, the dummy classifier, as the baseline model. We used stratified k-fold cross-validation with 10 splits and 3 repeats to ensure that each fold has the same class distribution as the original dataset and to effectively capture a sample of model performance on the dataset. We summarized the mean and standard deviation of the scores in table 3. Based on the results we decided to use Extreme Gradient Boosting algorithm (XGBoost).

XGBoost (Chen & Guestrin, 2016) has demonstrated to be a powerful technique to build predictive models. XGboost is a stage-wise additive model that uses trees as weak learners. Following a gradient descent procedure, new trees that reduce the log loss are added to the model. Finally, the multiple classification trees in the model are used for the prediction of new values of the outcome.

Table 3. Model development: Algorithm performance

Algorithm	**Mean accuracy (Std. Dev)**
DummyClass	0.332 (0.000)
Multinomial Logistic Regression (LR)	0.333 (0.012)
Linear Discriminant Analysis (LDA)	0.663 (0.010)
QuadraticDiscriminantAnalysis (QDA)	0.379 (0.056)
Gaussian Naïve Bayes (GNB)	0.344 (0.006)
Multinomial Naïve Bayes (MNB)	0.315 (0.012)
Support Vector Machine (SVM)	0.332 (0.000)
K-Nearest Neighbours (KNN)	0.320 (0.010)
Bagged Decision Trees (BAG)	0.752 (0.000)
Random Forest (RF)	0.738 (0.012)
ExtraTreesClassifier (ET)	0.738 (0.013)
Extreme Gradient Boosting (XGBoost)	0.754 (0.011)

3.2.3. Data preparation

XGboost can handle categorical numerical data. As most of the variables are dummy or categorical numerical in the database we needed few data treatment before adjusting the model. Particularly, due to the nature of the answer coding used for some of the variables in the survey, e.g. the target variable (OBJET11), we reverse coded them for the shake of clarity. After this process, a higher number indicates a higher degree of importance of the selected value (see categorical variables from 0 to 3 in table 1). The survey also contained multiclass variables, with each class representing a group of companies in the sample similar regarding a determined characteristic, for example, the industry (ACTIN) or the type of company (CLASE). We applied "one hot encoding" to transform these categorical variables where no relationship exists between categories. One hot encoding transforms each categorical variable with n categories into n dummy variables with a value 1 if the sample case belongs to the suggested category and 0 otherwise. The transformed dataset includes now a total of 200 variables.

3.2.4. Feature selection

We determined the baseline performance of the model with all the features and XGboost default options. We divided the dataset into train and test sets using a stratified split system to maintain sample percentage of cases in each of the 4 groups throughout the training and testing datasets. We used 2/3 of the data to train the model and the rest 1/3 for validation purposes. The model gave an initial accuracy of 73.78%.

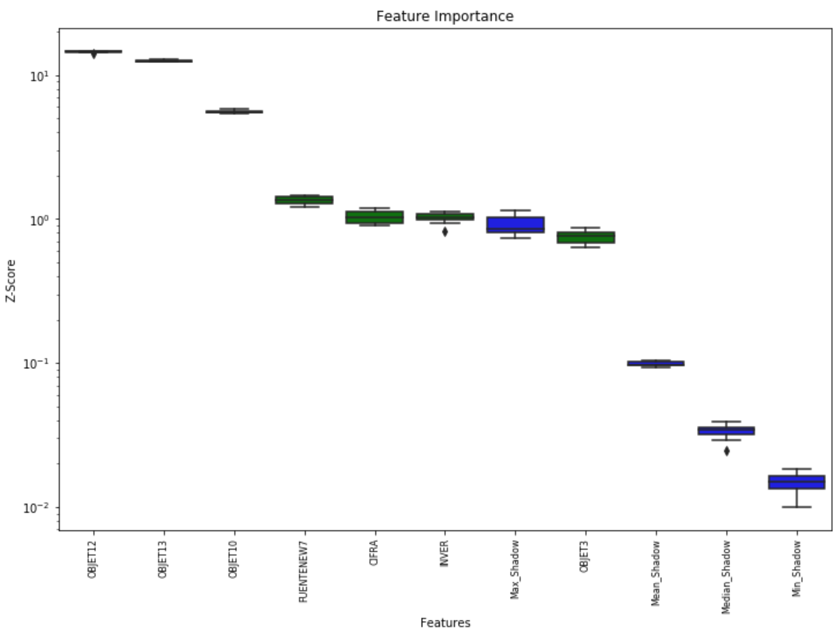
Typically, not all the variables are relevant to the classification problem. Besides the technical drawbacks of an increase of the calculation time and computing resources needed, including a large number of variables increases complexity, adds no or little valuable information and in many cases, degrades performance (Kohavi & John, 1997; Kuhn, M., & Johnson, K., 2013). Then, it is advisable to look for a smaller number of features to give the best possible results to our problem (Nilsson et al., 2007).

There are many algorithms which have been developed to reduce the feature set to a manageable size. XGboost automatically generates scores on the input features as part of fitting process and indicates the usefulness of a feature in the construction of the decision trees in the model. However, as in other classifiers, the feature importance looks for the minimal optimal set of features usable for prediction and depend on the classifier used. That is, it is context (methodology) dependent. Additionally, these built-in feature importance methods rely on accuracy to include or remove features from the feature set. However, a decrease in accuracy when removing a feature is sufficient condition to declare it as important but not sufficient to declare it as unimportant (Kudsa and Rudnicki, 2010). We cannot use correlation to filter features either, as the lack of direct correlation is not proof of the lack of importance of the feature, because it can be important in conjunction with other features (Guyon and Elisseeff 2003). Some features might be useful to predict the environmental orientation of the firm without being causally related, and may therefore be irrelevant. (Nilsson et al., 2007). Thus, we need a feature selection method that is able to provide us with all relevant features and not just the minimal-optimal. We want to understand the phenomenon, the mechanisms related to the environmental orientation by finding all factors that contribute to it, not only the non-redundant ones (Nilsson et al., 2007).

Boruta algorithm removes in an iterative process all the features which are proved, by a statistical test, to be less relevant than random probes (Stoppiglia, Dreyfus, Dubois, and Oussar, 2003). Then, Boruta is a selection method for all relevant features. Boruta was initially developed by Kudsa and Rudnicki, (2010) as an R-package built around a Random Forest classifier and later Homola (2017) implemented it in Python. Boruta classifies features in three groups, features that consistently significantly outperform over the random probes (important features), those that marginally outperform or that outperform at a less than statistically significant value (tentative features) and those that underperform (not selected features). However, although Boruta feature selection algorithm is statistically robust and rigorous, it is difficult for interpretation purposes as it is built around repeated random forest models. As our study is focused on identifying

and interpreting the role of eco-innovation drivers in the environmental orientation of the firms, we used SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2016 and Lundberg et al., 2017).
SHAP is a popular explainable artificial intelligence tool. It helps us interpret machine learning models with Shapley values. Shapley values are measures of the contribution each feature has in a machine learning model. In other words, these values indicate how much impact each feature has on the model output for individuals in the validation dataset. Predictions are explained as the addition of the values attributed to each feature. The global feature importance is the average of the marginal contributions of each feature. SHAP enables all permutations of the Boruta algorithm's to be averaged and globally compared. Then, the combination of the Boruta algorithm for feature importance and Shap, called BorutaShap, developed by Eoghan Keany (2020) leads to an unbiased and consistent classification of important and non-important features that also facilitates interpretability. Additionally, BorutaShap is implemented so we can choose any tree-based machine learning technique, like the efficient and fast XGboost, as the base model in the feature selection process. Thus, we used BorutaShap with a XGboost learner to determine the all the relevant features.
The BorutaShap algorithm confirmed 7 attributes as important (Figure 1), one tentative attribute and 192 unimportant attributes. We used the 7 selected features as a base model for the model development and tuning.

Figure 1. z scores of the selected variables resulting from the BorutaShap algorithm



3.2.5. Tunning the model.

We developed a new model with the 7 relevant attributes identified in the previous analysis and we tuned it. Tunning the model implies determining the optimal learning task parameters. These parameters are used to define the optimization objective the metric to be calculated at each step. We used multi:softprob objective, that returns predicted probability of each data point belonging to each class, and mlogloss evaluation metric.

From the existing trees, XGBoost algorithm adds new trees to reduce residual errors which results in the model to fit rapidly, resulting in overfit. To reduce the learning of the model we can weight the corrections for the errors (learning rate parameter).

Adding trees improves performance to a certain limit where we are overfitting the model to our training dataset and results in a poorer performance in our predictions. The tree depth adds more complication to the model, then we can expect a smaller number of trees to make a better prediction (max_depth parameter).

The algorithm creates each decision tree selecting the split points so they minimize the objective function, which can result on the use of the same features and split points over and over. Allowing the algorithm to use a random subsample of rows (cases) and columns (variables) when choosing the splitting points, creates slightly different trees, adding variance and improving the model performance. This procedure is called stochastic Gradient Boosting and it is controlled with the subsample or colsample by tree parameters.

By default, XGBoost algorithm creates 100 trees, which can be or not the optimal number of trees in a model. Early-stopping parameter will stop building new trees when the evaluation metric has not improved after adding a certain number of new trees.

Based on the later we adjusted the model parameters to improve the performance using default values as the base line accuracy. We used early stopping parameter to automatically determine the number of estimators in each step. Each tunning step consisted on a grid search of parameters to find the optimal solution. The optimal parameter remained fixed for the following step of the process. The steps in the tuning process, the optimal parameter value and the accuracy are showed in the Table 2 in appendix.

3.2.6. Interpreting the model.

The methodology design allows us to interpret several aspects. Firstly, we can evaluate the accuracy of the model. The model accuracy indicates the proportion of cases in the validation dataset that have been correctly predicted by the model.

Secondly, we can evaluate the feature importance of the predictors. With Boruta-Shap we reduced the model variables (features) including only relevant features. Then, using shap for the post-hoc explanation of the adjusted model, we can determine the relative importance of each variable in classifying cases in each of the four classes, its direction and how they interact with other variables in the model.

Finally, XGBoost nature makes interpretation of the logic that the model follows very difficult. XGboost is an ensemble algorithm, therefore, difficult to explain the predictions. It converts the model in a kind of a black box. To uncover the classification decision rules or policies used by the model we can create a tree surrogate. A tree surrogate is a tree model that fits to the actual prediction of our model. In other words, we built a simpler approximation to our model replicating its predicted results so it is more transparent and interpretable. The model will allow us to determine decision boundaries and the combinations of features leading to the different classes. We used a grid search to test different depth values and minimum cases per leaf of a tree classifier. We the tree surrogate parameters (depth and cases per leaf) based on balance between the complexity of the tree and its quality. We evaluated the quality calculating the fidelity score (how well the tree surrogate predicts the predicted results of our model) and the accuracy score (how well the tree surrogate predicts the real values of our outcome).

4. Results

4.1. Model results

After the tunning process, our final model gave an accuracy of 75.25%. The detailed results on the classification performance of the model throughout the tunning process is showed in table 4. The final accuracy metrics and the classification results are reported in table 5.

Table 4 Tunning procedure

Step	Optimal parameter value	Accuracy
Base model	Default parameters	74.78%
Step 1: Tune learning rate	Default (Learning rate=0,1)	74.78%
Step 2: Tune max_depth and min_child_weight	Default (max_depth=3 and min_child_weight=3)	74.85%
Step 3: Tune subsample and colsample_bytree	subsample=0.6 colsample_bytree=0.4	75.05%
Step 4: Tuning Regularization Parameter (reg_alpha)	Default (reg_alpha=0.03)	75.25%
Step 5: Tune gamma	Default (gamma=0)	75.25%

Table 5. Confusion matrix and model performance parameters

True\Predicted	Class 0	Class 1	Class 2	Class 3	Precision	Recall	F1-score	Support
Class 0 (Not)	**454**	15	15	11	0.85	0.92	0.88	495
Class 1 (Low)	25	**152**	51	21	0.71	0.61	0.66	249
Class 2 (Med)	29	34	**220**	77	0.65	0.61	0.63	360
Class 3 (high)	27	13	51	**296**	0.73	0.76	0.75	387
accuracy							0.75	1491
macro avg					0.74	0.73	0.73	1491
weighted avg					0.75	0.75	0.75	1491

The confusion matrix (in Table 5) represents the comparison between the test sample predicted classes and their actual classes. We can see in the diagonal the correct classifications of the cases in the evaluation (test) dataset. Out of the diagonal we find the misclassified firms. Precision evaluates correct prediction in the class over the total cases predicted in that class (e.g. Class 0 - 454/535=0.85) and it is more appropriate when minimizing false positives. Recall evaluates the correct predictions in the class over the total cases that really belong to that class (e.g. Class 0 - 454/495=0.92) and it is more appropriate when minimizing false negatives. F-measure provides a single measure balancing both precision and recall. Focusing on the F1-score, we can see the model works quite well classifying the extreme positions, f1-score equals to 75% detecting highly environmentally oriented innovators and 88% detecting non-environmentally oriented innovators. Overall, we can evaluate the performance of the model in predicting class labels using accuracy. Accuracy represents the total number of correct predictions divided by the total number of predictions (1122/1491=0.7525). According to Hair et al, 1998, a good classification model using multivariate techniques should have an accuracy 25% higher than the proportional chance criterion and higher than the maximum chance criterion. Our XGboost model accuracy is substantially higher than the traditional multivariate thresholds and, therefore, the accuracy that can be achieved with traditional multivariate techniques, such as discriminant analysis (around 56 %).

4.2. Feature importance evaluation

Figures 2 and 3 represents the features in the model sorted by increasing importance. The importance is representing the average shapley values per feature. That is, it is a measure of the global importance of the feature in the model. Figures 2 and 3also reflects the importance of each feature for each class in the data.

Figure 2. Model feature importance.

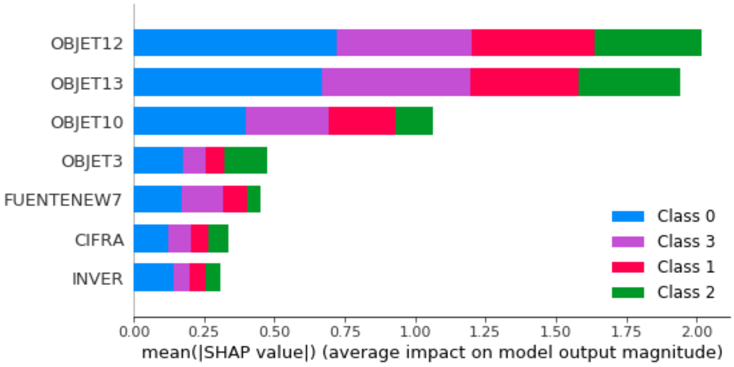
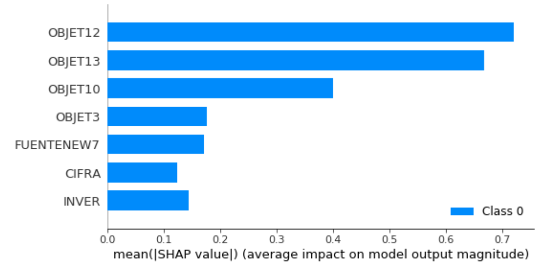
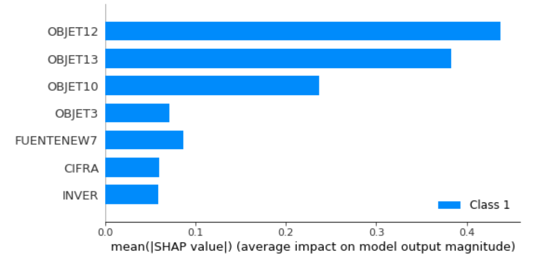


Figure 3. Shap feature importance by class

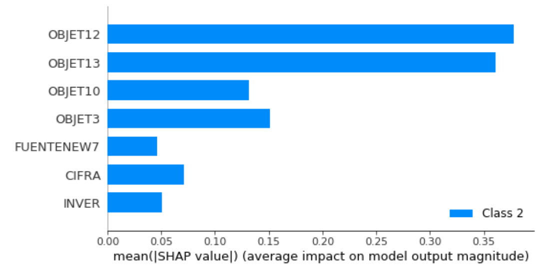
Class 0 (Not oriented)



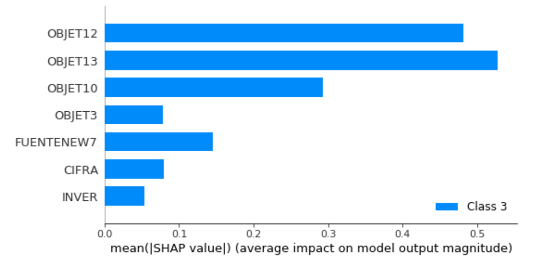
Class 1 (Low oriented)



Class 2 (Medium oriented)



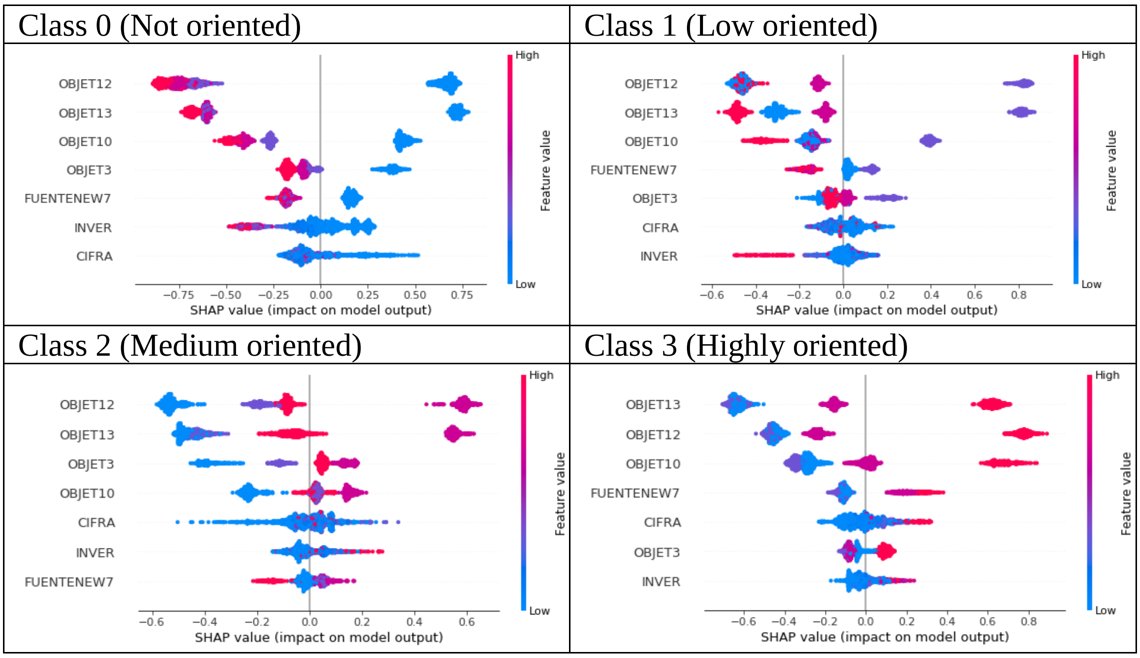
Class 3 (Highly oriented)



Overall, we clearly see clearly that 3 features emerge as the most important in order to determine the environmental orientation of the company when innovating. These features are OBJECT12, OBJECT13 and OBJET10. They represent, respectively, the importance of the improvement of the health or safety of the employees, of meeting the environmental, health or safety regulations and of the reduction of energy costs per unit produced, when innovating. In other words, the eco-innovation orientation is driven mainly by the external regulative pressure (OBJET13), the internal orientation to process innovations related to energy reduction (OBJET 10) and the internal orientation to improve workplace quality (OBJET12). With a lower importance, we find the orientation

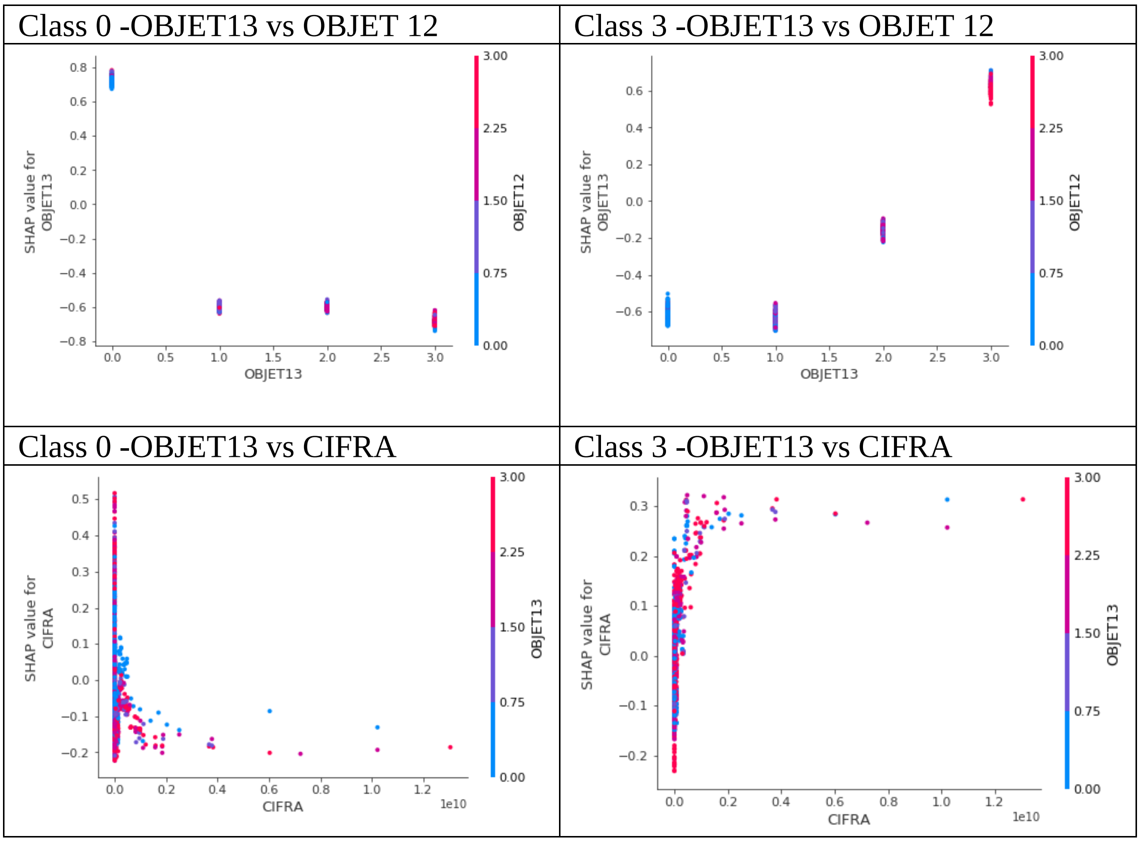
to product innovation (OBJECT3: importance to enter in new markets or increase market share) and the dependence on external information sources (FUENTENEW7: importance of information from Universities and other higher education institution for the innovation). Finally, firms' characteristics related to the total income and the total investment, which are common proxies of the company size, reveal themselves as relevant but with a relatively low importance. Then we can confirm the hypothesis H2, H4, H6 and H7, rejecting the rest of the hypothesis.
We can evaluate particularly, the importance of each feature in each class using a summary plot (see figure 4). The summary plot shows a density scatter plot of shap values for each class and each feature in order of importance, getting a sense of their distribution. The value of the feature is represented using colors.

Figure 4. Shap summary plots



For example, we can see that high values of OBJET13 (red color), that represents a high orientation to innovate looking to meet environmental regulations, push shap values higher in Class 3 (shap values between 0.6 and 0.8), which increases the probability to be classified as a company with high environmental orientation while innovating. On the contrary, these same high values of OBJECT13, push shap values to the negative area in Class 0 (shap values between -0,75 and -1). When we look at intermediate positions (medium and low oriented), we can see the correspondence of the medium and low feature values (purple color) with higher shap values (higher chances to be classified in this group). Ultimately, this is indicating that the higher the orientation to meet the environmental regulations the higher the chances to be environmentally oriented while innovating. Similarly, we can see this patter in almost all of the other features in the model.
Additionally, we can evaluate the interaction between any pair of different features (see figure 5).

Figure 5. Shap two-way dependence plots for class 0 and 3



For example, we can see that companies that have no orientation to meet the environmental regulations (OBJET13=0) and to improve health and safety of their employees (OBJET12=0), obtain mostly high shap values for feature OBJET13 (between 0.6 an 0.8) in class 0. That is, the combination of a lack of orientation in these two features increases a lot the probability to be classified in group 0 (not environmentally oriented), decreasing the probability (negative shap values) for the other classes. Simmilarly, we can evaluate the interactions with other features. Overall, we can see positive interactions between the variables considered in the model. That is, the confluence of higher values of the predictors increases the chances to be environmentally oriented and vice versa. However, for those interactions including the total investment or the total turnover, it is not clear from the analysis of the plots that the interaction alters significantly the shap value.

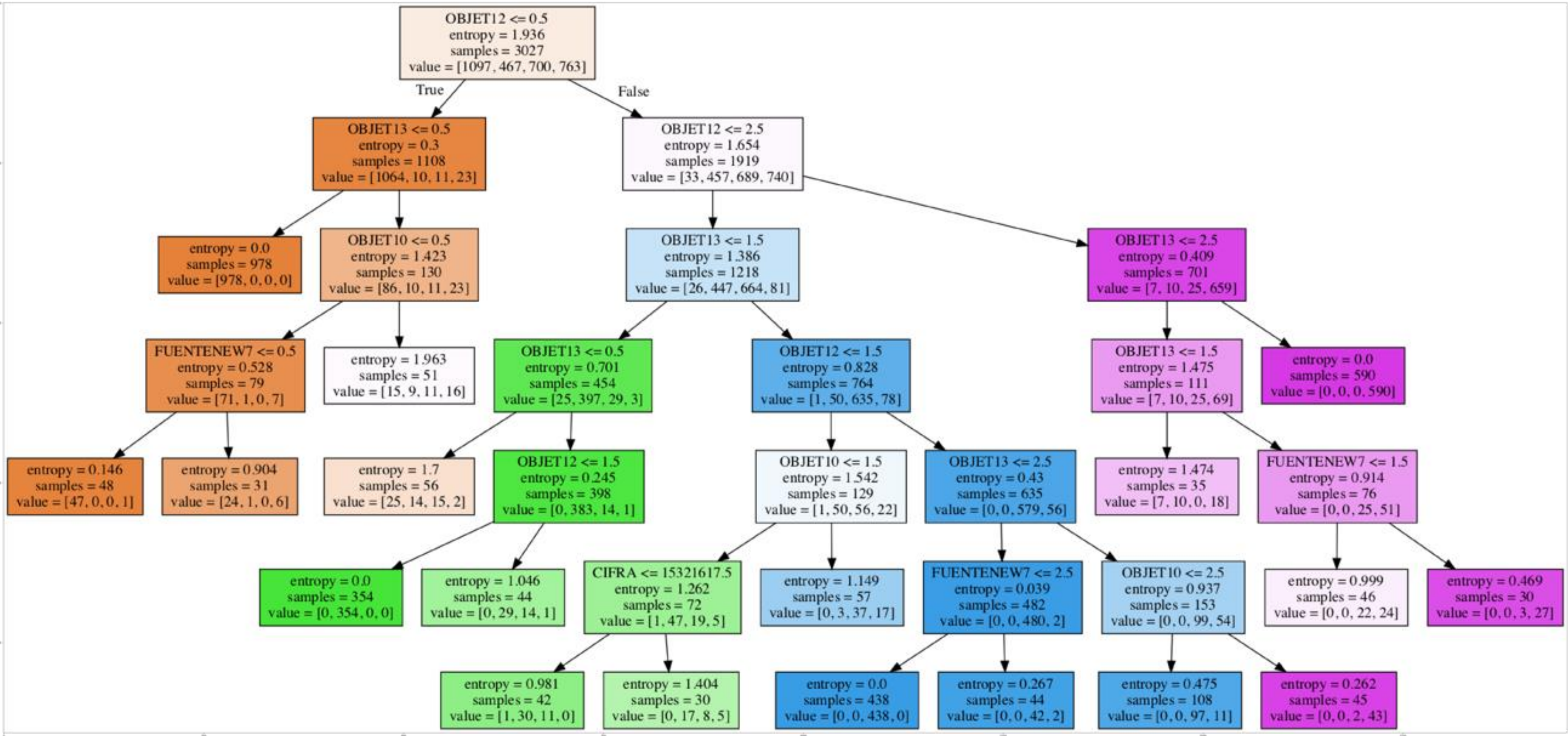
The environmental orientation patterns can be easily detected when we evaluate what features contribute to increase the shap value over the average model output (base value) for a single prediction. For example, for a company classified as not environmentally oriented (e.g. case 46 in figure 6), we will see high positive shap values for class 0 and low or negative shap values for the other classes, and particularly for class 3 (see figure 6). We can see, in red, the features pushing the shap values up and, in blue, the ones pushing them down. In line with the overall impact, being not oriented towards innovations focused on the health and safety of the employees (OBJET12=0), on meeting the environmental regulations (OBJET=13) or to energy reduction (OBJET10=0) are the most important contributors to increase the shap value (for case 46) for not environmentally oriented (Class 0) and decreasing the shap value for highly environmentally oriented (Class 3). Similarly, in a company classified as highly environmentally oriented (e.g. case 199) we will see high shap values for class 3 and low or negative values for the other cases (e.g. Class 0). Thus, we can evaluate for each company how and in what magnitude their feature values contribute to their environmental orientation.

Figure 6. Shap values for single predictions.

Finally, we built a tree surrogate (see figure 7). We performed a grid search changing the maximum depth and the minimum number of cases in a leaf. We evaluated the results in terms of the complexity of the model, the fidelity and the accuracy. The model with a maximum depth of 6 levels and a minimum of 30 cases per leaf resulted in an optimum balance between a reasonable level of complexity and a good performance (fidelity = 93.90% and accuracy=74.45% for the test dataset).

The features in the tree paths that connect the root to the leaf interact with each other (Breiman et al., 1984). These paths show the combination of feature values that lead to a certain classification of the cases (environmental orientation). In a way, this very similar the combination of conditions leading to the desired outcome that crisp and fuzzy sets Qualitative Comparative Analysis (cs and fsQCA) methodology produces (Ragin, 2000). We followed the tree paths to summarize (see table 6) the combinations of attributes leading to each of the four classes. We computed two indicators, consistency, that represents the number of cases in that combination that fall into the desired category, and coverage, that represents the amount of cases covered by that solution.

Figure 7. Decision tree surrogate



Thus, similar to fsQCA, consistency is equivalent to significance and coverage to the variance explained in a regression. In the same line, we considered for the analysis the solutions (combination of conditions leading to the desired outcome/class) with a consistency above 0.8.

Table 6. Tree surrogate summarized paths and their evaluation metrics

Sol.	Combination of condition	Class 0	Class 1	Class 2	Class 3	Consistency	Coverage
1	(Objet12=3)*(Objet13=3)	0	0	0	**590**	**1.000**	**0.773**
2	(Objet12=3)*(Objet13=2)*(FUENTENEW7= [2,3])	0	0	3	**27**	**0.900**	**0.035**
3	(Objet12=3)*(Objet13=2)*(FUENTENEW7= [0,1])	0	0	22	**24**	0.522	0.031
4	(Objet12=3)*(Objet13=[0,1])	7	10	0	**18**	0.514	0.024
5	(Objet12=2)*(Objet13=3)*(Objet10=3)	0	0	2	**43**	**0.956**	**0.056**
6	(Objet12=2)*(Objet13=3)*(Objet10=[0,2])	0	0	**97**	11	**0.898**	**0.139**
7	(Objet12=2)*(Objet13=2)	0	0	**480**	2	**0.996**	**0.686**
8	(Objet12=1)*(Objet13=[2,3])*(Objet10=[2,3])	0	3	**37**	17	0.649	0.053
9	(Objet12=1)*(Objet13=[2,3])*(Objet10=[0,1])*(Cifra>15.32 M€)	0	**17**	8	5	0.567	0.036
10	(Objet12=1)*(Objet13=[2,3])*(Objet10=[0,1])*(Cifra<15.32 M€)	1	**30**	11	0	0.714	0.064
11	(Objet12=2)*(Objet13=1)	0	**29**	14	1	0.659	0.062
12	(Objet12=1)*(Objet13=1)	0	**354**	0	0	**1.000**	**0.758**
13	(Objet12=[1,2])*(Objet13=0)	**25**	14	15	2	0.446	0.023
14	(Objet12=0)*(Objet13>0)*(Objet10>0)	**15**	9	11	16	0.314	0.021
15	(Objet12=0)*(Objet13>0)*(Objet10=0)	**71**	1	0	7	**0.899**	**0.001**
16	(Objet12=0)*(Objet13=0)	**978**	0	0	0	**1.000**	**0.892**
		1097	467	700	763	3027	

Then, for example for class 3, three different paths (solutions 1, 2 and 5 on table 6) lead to high levels of environmental orientation (class 3). The first, and most important path, is the combination of high levels of orientation to meet the regulations (OBJET13=3) and to the improvement of health and safety of the employees (OBJET12=3). This path covers 77.3% of the companies showing high environmental orientation and it is fully consistent (consistency=1). That is, all the companies that showed high levels in both values (OBJET12 and OBJET13) are highly oriented towards eco-innovation. The second, combines 3 conditions, high values of OBJET12 and medium values of OBJET13 and medium or high importance of the information from Universities and other higher education centers for the innovation process of the firm (FUENTENEWW7). Therefore, these three conditions if they appear in a firm, the firm is likely to have a high orientation to eco-innovation. This path covers only 3.5% of high environmentally oriented companies with a consistency of 90%. Then, 90 % of the companies with this combination of conditions showed high environmental orientation. The third path combines medium orientation in OBJET12, and high orientation in OBJET 13 and OBJET10 (orientation to the reduction of energy consumption). This path covers 0.56% of the cases with a consistency of 95.6%. In a similar way, we can analyze all the other relevant paths (rows with bold values for consistency). However, as happens for class 3, there is a dominant path in each of the classes. Indeed, the dominant paths in medium, low and not environmentally oriented companies are companies that show medium, low or no orientation in both OBJET13 and OBJET12, respectively. The consistent paths can be

seen as a combination of conditions that are sufficient for the outcome (the corresponding level of environmental orientation). Then, all the solutions that we have found with consistencies higher than 0,8 are combinations of conditions sufficient for the corresponding level of environmental orientation.
We can also analyze if there is any necessary condition. Analogously to fsQCA, a necessary condition would be a condition which is present in all the paths to a desired outcome (a desired level of environmental orientation). In other words, its absence will indicate that the outcome is not present either. If we consider a threshold of 0.9 consistency for the necessity analysis, we can see that Class 3 has no condition present in the three solutions. However, for class 2, 1 and 0 we find that levels of OBJET12 equals to 2, 1 and 0, respectively, are necessary conditions for the outcome. For example, for class 2, the two solutions (solutions 6 and 7 in table 6) contain OBJET12 equals 2, and the combined consistency is 97.7%. Thus, a medium level of innovative orientation to the improvement of health and safety of the employees is a necessary condition to achieve medium level of eco-innovation orientation.

5. Discussion and conclusions

This study uses machine learning techniques to evaluate the impact of a large number of firms' features in the eco-innovation orientation of the Spanish innovative firms. The study provides evidence of the impact of both internal and external eco-innovation drivers, and some firms characteristics in the environmental orientation of the firms while innovating. We provide empirical evidence about what aspects and how these aspects motivate companies' orientation towards environmental innovation. In addition, we uncovered their relative importance and how they interact between each other.
Firstly, we looked for a machine learning algorithm able to maximize efficiency and effectiveness in classifying the companies attending to their environmental orientation. We used a machine learning algorithm that improves significantly the classification task over traditional multivariate techniques such as, discriminant analysis, which opens new research possibilities in social sciences studies.
Secondly, we used a state-of-the-art machine learning technique to identify the so called "all relevant attributes" instead of those non-redundant, that result from a minimal-optimal approach to feature selection. We extracted a set features which are proved to be statistically more relevant than random probes, and therefore, relevant in determining the environmental orientation of the firm. Although, we included an extensive set of attributes (142 variables) that literature has pointed out as drivers to innovation and, particularly, to eco-innovation, only 7 attributes were deemed to be relevant. Then, unlike several research studies, we took a wide approach including a large sample of companies covering the whole spectrum of industries, sizes, and firm's behaviour related to their innovative activities, which gives strong support to the generalizability of the conclusions.
The relevant features detected reinforce the idea that both internal and external drivers are key to eco-innovation. Particularly, we confirmed the great importance of the regulative pressure as a motivator to eco-innovation, in line with previous research (Horbach, 2008; Demirel & Kesidou, 2011; De Marchi, 2012; Horbach et al., 2013; del Río et al., 2015). We also confirmed that the process orientation is another important driver. Particularly, the results point to the innovations looking to reduce the energy costs as very relevant. This is aligned with the double externality nature of eco-innovation (Rennings 2000). Looking for cost/efficiency innovations (process innovations) can produce environmental externalities as a side effect. Then, even the original trigger for

the innovation might not be the reduction of the environmental impact, the synergies that eco-innovations bring might be also be considered in the innovation process. Then, the process innovation approach might be overcoming the higher costs indicated by Rennings et al. (2006) that act as a disincentive to eco-innovation.

Our study also confirms the relevance of the product innovation orientation. Particularly, companies looking to expand their market or their market share through innovation, which is a representation of the so-called market demand factor (Veugelers, 2012; Horbach et al. 2013), also are more willing to eco-innovate. These results are aligned with several previous studies (Segarra-Oña et al., 2016; Segarra-Oña et al., 2017; Peiro-Signes & Segarra-Oña 2018) that evidenced a higher impact on the environmental orientation of process versus product innovation orientation.

One important contribution of this study is uncovering the relevance of the quality of an organization in their environmental orientation. Normally, companies that go further than the regulations and other economic related business in their activities (i.g. with CSR measures) are considered examples of business excellence or the quality of the organization. The search for an improvement of the health and safety conditions when innovating is nothing less than a representation this excellence. Our findings position this approach as the most relevant feature in medium, low and not environmentally oriented companies and the second most relevant in the high oriented companies. In other words, the lack of interest in this approach while innovating is a significant indicator that the company is not going to be environmentally oriented either, and vice versa.

Cooperation and collaboration have been highlighted as important for eco-innovation in many studies (Horbach, 2008; del Río et al., 2015). Several studies have evaluated several degrees of engagement of different stakeholders in firms' innovation activities, from information dependence to active cooperation. In our study, the only relevant relation with stakeholders that influences in the environmental orientation of the firm is the dependence of the information from Universities and other higher education centers for the innovation process. Particularly in Spain, Universities R&D expenditure represented half of the total R&D performed by Spanish companies in 2016 (INE, 2017). Moreover, there is a high number of SMEs in the Spanish economy and they lack of knowledge to undertake environmental innovations (European Commission 2019). Then, the shortage of in-house knowledge or technology, particularly environmental capabilities, is more likely to be supplemented through the knowledge from Universities. Those companies that do rely on the Universities as a source of information in the innovation process have a greater exposure to environmental innovations and might be more open to implementing environmental practices. These results contrast with the studies that indicate a need of active cooperation with one or more stakeholders.

Among firms' characteristics, size has been many times somehow considered in studying eco-innovation (Carrillo-Hermosilla et al., 2009; del Río et al., 2016; Hojnik & Ruzzier, 2016), either because the studies have been centered in large firms or SMEs, or because it has been considered as a control variable. In our case, we demonstrated that the size of the firm, in terms of the company income and investment, is also relevant. Indeed, both income and investment positively affect environmental orientation. However, its impact on the environmental orientation is rather marginal compared to the aforementioned attributes. Thus, their relatively low impact and the nature of these firm characteristics suggests that managerial and policy interventions should be focusing on the improvement of firms' innovative approach rather than a specific type of companies or industries.

One important conclusion of the study is not related to the presence of these attributes in the model, but for the absence of certain attributes. We can highlight that some extensively studied drivers have no relevance in determining the environmental orientation of the firm, for example, the formal R&D activity (investment in R&D, R&D personnel, patents,…), the organizational and commercial innovation orientation, the subsidies or the active collaboration with stakeholders. The lack of relevance questions the generalizability of the previous research that consider these drivers. Indeed, many previous studies are limited by the industry, the type of companies (e.g. SMEs), the variables included in the study or methodology employed. Additionally, these results point to the exploration of the role of the identified relevant features as mediators in the relation of a priori non relevant features.

Finally, our study provides a set of different combination of conditions that define the environmental orientation of the company. Overall, the confluence of a certain level of orientation to meet the regulations and to improve the health and safety of the employees while innovating is the most important driver to the eco-innovation orientation. These results are indicating that there is a combination of drivers what influences the environmental orientation. This combination includes both external and internal approaches, suggesting that there is a need to maintain the regulatory pressure over the firms and to facilitate more socially responsible approaches that encourage innovations that go further than the typical product, process, organizational and marketing approaches to innovation.

Moreover, we uncover that a determined orientation of the firm's innovation for an improvement in the health and safety of the employees is a necessary condition for the corresponding eco-innovative orientation.

This analysis carries important policy implications. First, regulations are an effective tool driving eco-innovation orientation regardless the industry. However, the regulatory pressure needs to be complemented with a proactive social approach from the inside of the company. As actual subsidies have been proven to be not effective, the policy makers should rethink the activities that should be targeted with this founds. The improvement of firms' efficiency, through the reduction of waste (energy, materials, water, etc.), would take advantage of the double externality nature of eco-innovations, creating a new innovative culture that sees innovation as a multidimensional solution to several firm's problems. The promotion of socially responsible practices would also rise the awareness about the environmental implications of firms' activities. Particularly, this could be reflected in the improvement of the environmental orientation of firms when redefining, redesigning or developing of new processes or new products.

Second, we have evidenced small impact that increasing the market share or developing new markets have triggering eco-innovation within firms. The promotion of the consumption among customers of environmentally responsible goods and services or from environmentally responsible firms can also be addressed through policy instruments. The promotion of environmental labels or tax reductions might be explored to increase customer awareness and to overcome some barriers ballasting the increase of these market demand. Finally, the lack of environmental and innovation knowledge and capabilities in Spanish firms is not being complemented by active collaboration with the different stakeholders available. The fact that an important number of the companies in the survey do not report any innovation activity and, that in those that we can classify as innovators, cooperation is not a main driver of eco-innovation, indicates that there is an urgent need to facilitate the firms' cooperation with different stakeholders and the innovation diffusion though new public policies.

Limitations and further research

This study has some limitations. First, this study relies on Spanish data. Spain is ranked as a moderate innovator in the Eco-Innovation Scoreboard (Eco-IS 2015). Moreover, culturally, other European countries have a greater awareness for environmental related issues. Then, an extended study including high eco-innovative and environmentally conscious countries might help to understand better the similarities and dissimilarities among companies in different European countries regarding their environmental orientation when innovating. Thus, a similar study using CIS data should be explored.
Second, we used cross-sectional data for the study. However, PITEC is a panel study which has been monitoring the innovate activity of a sample of Spanish companies for more than a decade. Then, future research can explore incorporating previous data from the sample companies to the study or use data from previous years to train the model and the target year to test it. In the first case, the strategic character of the environmental orientation has been proven by persistence of eco-innovation drivers over time. Then, the results would likely lead to a more accurate model, due to the inclusion of the previous eco-innovation approach, in a trade-off for a lower comprehension of what other factors are characterizing the environmental orientation of the firms. In the second case, we would be able to evaluate the stability and the predictive power of the model over time.

References

Adams, R., Jeanrenaud, S., Bessant, J., Overy, P., & Denyer, D. (2012). Innovating for sustainability. Network for Business Sustainability, 107. https://doi.org/10.4324/9780203889565

Astuti, M., Prawoto, P., Irawan, Y.S., Sugiono, S., 2018. The eco-innovation variables which influence the performance of creative industries center of natural stone crafts. J. Ecol. Eng. 19 (1), 14–24.

Beise, Marian, Rennings, Klaus, 2005. Lead markets and regulation: a framework for analyzing the international diffusion of environmental innovations. Ecol. Econ. 52 (1), 5–17.

Belin, Jean, Horbach, Jens, Oltra, Vanessa, 2011. Determinants and Specificities of Eco-innovations—An Econometric Analysis for the French and German Industry Based on the Community Innovation Survey. http://ideas.repec.org/p/grt/wpegrt/2011-17.html (26.06.13.).

Bocken, N., Short, S., Rana, P., & Evans, S. (2014). A literature and practice review to develop sustainable business model archetypes. Journal of Cleaner Production, 65, 42–56. https://doi.org/10.1016/j. jclepro.2013.11.039

BorutaShap : A wrapper feature selection method which combines the Boruta feature selection algorithm with Shapley values. Eoghan Keany 2020 DOI: 10.5281/zenodo.4247610.

Bossle, M.B., De Barcellos, M.D., Vieira, L.M., Sauvée, L., 2016. The drivers for adoption of eco-innovation. J. Clean. Prod. 113, 861–872.

Breiman L., Friedman J., Olshen R., Stone C., Classification and regression trees (1984).

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC press.

Cai, W., & Zhou, X. (2014). On the drivers of eco-innovation: Empirical evidence from China. Journal of Cleaner Production, 79, 239–248. https://doi.org/10.1016/j.jclepro.2014.05.035

Cainelli, G., & Mazzanti, M. (2013). Environmental innovations in services: Manufacturing- services integration and policy transmissions. Research Policy, 42, 1595–1604.
Cainelli, G., De Marchi, V., & Grandinetti, R. (2015). Does the development of environmental innovation require different resources? Evidence from Spanish manufacturing firms. Journal of Cleaner Production, 94, 211–220. https://doi.org/10.1016/j.jclepro.2015.02.008
Cainelli, G., Mazzanti, M., & Montresor, S. (2012). Environmental innovations, local networks and internationalization. Industry and Innovation, 19(8), 697–734. https://doi.org/10.1080/13662716.2012.739782
Cainelli, G., Mazzanti, M., Montresor, S., 2011. Environmental innovations, local networks and internationalization (No. 20.2011), sustainable development series, sustainable development series. Padova, Italy.
Carrillo-Hermosilla, Javier, del Río, Pablo, Könnölä, Totti, 2010. Diversity of eco-innovations: reflections from selected case studies. J. Clean. Prod. 18 (10–11), 1073–1083.
Carrillo-Hermosilla, J., del Río, P., & Könnölä, T. (2009). Eco-innovation. When sustainability and competitiveness shake hands. London, UK: Palgrave Macmillan.
Castellacci, F., & Lie, C. M. (2017). A taxonomy of green innovators: Empirical evidence from South Korea. Journal of Cleaner Production, 143, 1036–1047. https://doi.org/10.1093/icc/dtr051
Chen T. & Guestrin C. 2016. Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, ACM, pp. 785–794.
Cleff, T., Rennings, K., 1999. Determinants of environmental product and process innovation. Eur. Environ. 9, 191–201.
Cooke, P. (2011). Transition regions: Regional-national Eco-innovation systems and strategies, Progress in Planning, 76, 105–146.
Costa-Campi, M. T., García-Quevedo, J., & Segarra-Blasco, A. (2015). Energy efficiency determinants: An empirical analysis of Spanish innovative firms. Energy Policy, 83, 229–239. https://doi.org/10.1016/j. enpol.2015.01.037
Cuerva, M. C., Triguero-Cano, Á., & Córcoles, D. (2013). Drivers of green and non-green innovation: Empirical evidence in Low-Tech SMEs. Journal of Cleaner Production, 68, 104–113. https://doi.org/10.1016/j. jclepro.2013.10.049
Cuerva, M. C., Triguero-Cano, Á., & Córcoles, D. (2014). Drivers of green and non-green innovation: Empirical evidence in Low-Tech SMEs. Journal of Cleaner Production, 68, 104–113. https://doi.org/10.1016/j. jclepro.2013.10.049
Dangelico, R.M., 2016. Green product innovation: where we are and where we are going. Bus. Strategy Environ. 25 (8), 560–576.
De Marchi, V. (2012). Environmental innovation and R&D cooperation: Empirical evidence from Spanish manufacturing firms. Research Policy, 41, 614–623. https://doi.org/10.1016/j.respol.2011.10.002
De Marchi, V., & Grandinetti, R. (2013). Knowledge strategies for environmental innovations: The case of Italian manufacturing firms. Journal of Knowledge Management, 17, 569–582. https://doi.org/ 10.1108/JKM-03-2013-0121
Del Río, P. (2009). The empirical analysis of the determinants for environmental technological change: A research agenda. Ecological Economics, 68, 861–878. https://doi.org/10.1016/j.ecolecon. 2008.07.004
Del Río, P., Carrillo-Hermosilla, J., Könnölä, T., & Bleda, M. (2016). Resources, capabilities and competences for eco-innovation. Technological and Economic

Development of Economy, 22, 274–292. https://doi.org/10.3846/20294913.2015.1070301
del Río, P., Peñasco, C., & Romero-Jordán, D. (2015). Distinctive features of environmental innovators: An econometric analysis. Business Strategy and the Environment, 24(6), 361–385. https://doi.org/10.1002/ bse.1822
Del Río, P., Romero-Jordán, D., & Peñasco, C. (2017). Analysing firm-specific and type-specific determinants of eco-innovation. Technological and Economic Development of Economy, 23, 270–295. https://doi.org/ 10.3846/20294913.2015.1072749
Demirel, P., & Kesidou, E. (2011). Stimulating different types of eco-innovation in the UK: Government policies and firm motivations. Ecological Economics, 70, 1546–1557. https://doi.org/10.1016/j.ecolecon.2011.03.019
Díaz-García, C., González-Moreno, Á., & Sáez-Martínez, F. J. (2015). Eco-innovation: Insights from a literature review. Innovation: Management, Policy and Practice, 17, 6–23. https://doi.org/10.1080/ 14479338.2015.1011060
Doran, J., Ryan, G., 2016. The importance of the diverse drivers and types of environmental innovation for firm performance. Bus. Strateg. Environ. 25, 102–119.
EEA (2014). Resource-efficient green economy and EU policies. Luxembourg: Publications Office of the European Union.
EIO (Eco-Innovation Observatory), 2012. Methodological Report. Eco-Innovation Observatory. Funded by the European Commission, DG, Environment, Brussels.
Frondel, M., Horbach, J., Rennings, K., 2004. End-of-pipe or cleaner pro- duction? An empirical comparison of environmental innovation decisions across OECD countries (No. No. 04–82), ZEW discussion papers.
Fussler, C., & James, P. (1996). Driving eco-innovation: a breakthrough discipline for innovation and sustainability. Financial Times/Prentice Hall.
Galdeano-Gómez, E., Aznar-Sánchez, J.A., Pérez-Mesa, J.C., 2013. Sustainability dimen- sions related to agricultural-based development: the experience of 50 years of in-tensive farming in Almería (Spain). Int. J. Agric. Sustain. 11, 125–143.
García-Granero, E. M., Piedra-Muñoz, L., & Galdeano-Gómez, E. (2020). Measuring eco-innovation dimensions: The role of environmental corporate culture and commercial orientation. Research Policy, 49(8), 104028.
Ghisetti, C., Marzucchi, A., & Montresor, S. (2015). The open eco-innova- tion mode. An empirical investigation of eleven European countries. Research Policy, 44, 1080–1093. https://doi.org/10.1016/j. respol.2014.12.001
Godoy-Durán, A., Galdeano-Gómez, E., Perez-Mesa, J.C., Piedra-Muñoz, L., 2017. Assessing eco-efficiency and the determinants of horticultural family-farming in southeast Spain. J. Enviro. Manag. 204 (1), 594–604.
González-Moreno, A., Sáez-Martínez, F.J., Díaz-García, C., 2013. Drivers of Eco-Innovation in the chemical industry. Environ. Eng. Manag. J. 12 (10), 1535e1541.
Grané, A., Salini, S., & Verdolini, E. (2021). Robust multivariate analysis for mixed-type data: Novel algorithm and its practical application in socio-economic research. Socio-Economic Planning Sciences, 73, 100907.
Guoyou, Q., Saixing, Z., Chiming, T., Haitao, Y., & Hailiang, Z. (2013). Stakeholders' influences on corporate green innovation strategy: A case study of manufacturing firms in china. Corporate Social Responsibility and Environmental Management, 20(1), 1–14.
Guyon I, Elisseeff A (2003). "An Introduction to Variable and Feature Selection." Journal of Machine Learning Research, 3, 1157–1182
Hansen, T., & Coenen, L. (2015). The geography of sustainability transi- tions: Review, synthesis and reflections on an emergent research field. Environmental Innovation and Societal Transitions, 17, 92–109. https://doi.org/10.1016/j.eist.2015.07.004

Hofer, Christian, David Cantor, E., Dai, Jing, 2012. The competitive determinants of a firm's environmental management activities: evidence from US manufacturing industries. J. Oper. Manag. 30 (January (1–2)), 69–84.
Hojnik, J., & Ruzzier, M. (2016). The driving forces of process eco- innovation and its impact on performance: Insights from Slovenia. Journal of Cleaner Production, 133, 812–825. https://doi.org/10.1016/ j.jclepro.2016.06.002
Homola, D. (2017). Integration and visualisation of clinical-omics datasets for medical knowledge discovery.
Horbach, J. (2008). Determinants of environmental innovation—New evidence from German panel data sources. Research Policy, 37, 163–173. https://doi.org/10.1016/j.respol.2007.08.006
Horbach, J. (2014). Do Eco-innovations need specific regional characteristics? An econometric analysis for Germany. Review of Regional Research, 34, 23–38.
Horbach, J. (2016). Empirical determinants of eco-innovation in European countries using the community innovation survey. Environmental Innovation and Societal Transitions, 19, 1–14. https://doi.org/10.1016/j. eist.2015.09.005
Horbach, J., Jabob, J., 2018. The relevance of personal characteristics and gender diversity for (eco)-innovation activities at the firm-level. Results from a linked employer-employee database in Germany. Business strat- egy and the environment 1–11. https://doi.org/10.1002/bse.2042
Horbach, J., Oltra, V., Belin, J., 2013. Determinants and specificities of eco- innovations compared to other innovations: an econometric analysis for the French and German industry based on the Community Innovation Survey. Ind. Innov. 20 (6), 523e543.
Horbach, Jens, Rammer, Christian, Rennings, Klaus, 2012. Determinants of eco-innovations by type of environmental impact—the role of regulatory push/pull, technology push and market pull. Ecol. Econ. 78, 112–122.
Johnson, D. K. N., & Lybecker, K. M. (2012). Paying for green: An economics literature review on the constraints to financing environmental innovation. Electronic Green Journal, 1. Retrieved from https://escholarship.org/uc/item/6ks30864
Jové-Llopis, E., & Segarra-Blasco, A. (2018). Eco-innovation strategies: A panel data analysis of Spanish manufacturing firms. Business Strategy and the Environment, 27(8), 1209-1220.
Junquera, B., Del Brío, J. Á., & Fernández, E. (2012). Clients' involvement in environmental issues and organizational performance in businesses: An empirical analysis. Journal of Cleaner Production, 37, 288–298. https:// doi.org/10.1016/j.jclepro.2012.07.029
Kammerer, D. (2009). The effects of customer benefit and regulation on environmental product innovation. Ecological Economics, 68, 2285–2295. https://doi.org/10.1016/j.ecolecon.2009.02.016
Keany E. BorutaShap 1.0.14 2020 [Available from: https://pypi.org/project/BorutaShap/]
Kemp, René, Pearson, Peter, 2007. Final Report MEI Project About Measuring Eco innovation. Maastricht: 1-120.
Kesidou, E., & Demirel, P. (2012). On the drivers of eco-innovations: Empir- ical evidence from the UK. Research Policy, 41, 862–870. https://doi. org/10.1016/j.respol.2012.01.005
Keskin, D., Diehl, J. C., & Molenaar, N. (2012). Innovation process of new ventures driven by sustainability. Journal of Cleaner Production. doi:10.1016/j.jclepro.2012.05.012

Kiefer, C. P., Del Rio Gonzalez, P., & Carrillo-Hermosilla, J. (2019). Drivers and barriers of eco-innovation types for sustainable transitions: A quantitative perspective. Business Strategy and the Environment, 28(1), 155-172.
Kohavi R, John GH (1997). “Wrappers for Feature Subset Selection.” Artificial Intelligence, 97, 273–324.
Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. Artificial intelligence, 97(1-2), 273-324.
Könnölä, T., Unruh, G. C., & Carrillo-Hermosilla, J. (2006). Prospective voluntary agreements for escaping techno-institutional lock-in. Ecological Economics, 57, 239–252. https://doi.org/10.1016/j. ecolecon.2005.04.007
Kuhn, M., & Johnson, K. (2013). Applied predictive modeling (Vol. 26). New York: Springer.
Kursa M., Rudnicki W., "Feature Selection with the Boruta Package" Journal of Statistical Software, Vol. 36, Issue 11, Sep 2010
Li, Yina, 2014. Environmental innovation practices and performance: moderating effect of resource commitment. J. Clean. Prod. 66 (March), 450–458.
Liao, Z., 2018. Market orientation and FIRMS’ environmental innovation: the moderating role of environmental attitude. Bus. Strateg. Environ. 17, 117–127.
Lundberg S, Lee S-I (2017) A Unified Approach to Interpreting Model Predictions. Nips 16, 426–430.
Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee S-I (2020) From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2, 56–67.
Marcon, A., de Medeiros, J.F., Ribeiro, J.L.D., 2017. Innovation and environmentally sustainable economy: identifying the best practices develop by multinationals in Brazil. J. Clean. Prod. 160, 83–97. https://doi.org/10.1016/j.jclepro.2017.02.101.
Markard, J., Worch, H., 2010. Technological innovation systems and the resource based view—Resources at the firm, network and system level.
Martin, C.J., Upham, P., Budd, L., 2015. Commercial orientation in grassroots social innovation: insights from the sharing economy. Ecol. Econ. 118, 240–251.
Mazzanti, M., & Zoboli, R. (2009). Embedding environmental innovation in local production systems: SME strategies, networking and industrial relations: Evidence on innovation drivers in industrial districts. Interna- tional Review of Applied Economics, 23, 169–195. https://doi.org/ 10.1080/02692170802700500
Mondéjar-Jiménez, J., Segarra-Oña, M., Peiró-Signes, Á., Payá-Martínez, A. M., & Sáez-Malefors, C., Secondi, L., Marchetti, S., & Eriksson, M. (2021). Food waste reduction and economic savings in times of crisis: The potential of machine learning methods to plan guest attendance in Swedish public catering during the Covid-19 pandemic. Socio-Economic Planning Sciences, 101041.
Martínez, F. J. (2015). Segmentation of the Spanish automotive industry with respect to the environmental orientation of firms: towards an ad-hoc vertical policy to promote eco-innovation. Journal of Cleaner Production, 86, 238-244.
Mothe, C., & Nguyen-Thi, U. T. (2017). Persistent openness and environmental innovation: An empirical analysis of French manufacturing firms. Journal of Cleaner Production, 162, S59–SS69. https://doi.org/ 10.1016/j.jclepro.2016.12.063
Negny, S., Belaud, J. P., Cortes Robles, G., Roldan Reyes, E., & Ferrer, J. B. (2012). Toward an Eco-innovative method based on a better use of resources: Application to chemical process preliminary design. Journal of Cleaner Production, 32, 101–113.

Nilsson R, Pen˜a J, Bjo¨rkegren J, Tegn´er J (2007). “Consistent Feature Selection for Pattern Recognition in Polynomial Time.” The Journal of Machine Learning Research, 8, 612.
OECD (2011). Better policies to support eco-innovation. OECD Studies on Environmental Innovation. Paris, France: OECD.
Ortiz-de-Mandojana, N., Aguilera-Caracuel, J., Morales-Raya, M., 2016. Corporate governance and environmental sustainability: the moderating role of the national institutional context. Corp. Soc. Respons. Environ. Manag. 23 (3), 150–164
Peiró-Signes, Á., & Segarra-Oña, M. (2018). How past decisions affect future behavior on eco-innovation: An empirical study. Business Strategy and the Environment, 27(8), 1233-124
Pereira, A., & Vence, X. (2012). Key business factors for Eco-innovation: An overview of recent firm-level empirical studies. Cuadernos de Gestion, 12, 73–103.
Petruzzelli, A. M., Dangelico, R. M., Rotolo, D., & Albino, V. (2011). Organizational factors and technological features in the development of green innovations: Evidence from patent analysis. Innovation-Management Policy and Practice, 13, 291–310.
Ragin, C. C. (2000). Fuzzy-set social science. University of Chicago Press.
Rehfeld, K.-M., Rennings, K., Ziegler, A., 2007. Integrated product policy and environmental product innovations—an empirical analysis. Ecol. Econ. 61, 91–100.
Rehfeld, K.-M., Rennings, K., & Ziegler, A. (2007). Integrated product policy and environmental product innovations: An empirical analysis. Ecologi- cal Economics, 61, 91–100. https://doi.org/10.1016/j. ecolecon.2006.02.003
Rennings, K., Ziegler, A., Ankele, K., & Hoffmann, E. (2006). The influence of different characteristics of the EU environmental management and auditing scheme on technical environmental innovations and economic performance. Ecological Economics, 57(1), 45–59. https://doi.org/ 10.1016/j.ecolecon.2005.03.013
Rennings, Klaus, 2000. Redefining innovation—eco-innovation research and the contribution from ecological economics. Ecol. Econ. 32 (2), 319–332.
Revell, A., Stokes, D., & Chen, H. (2010). Small businesses and the environment: Turning over a new leaf? Business Strategy and the Environment, 19, 273–288.
Rothenberg, Sandra, Zyglidopoulos, Stelios C., 2007. Determinants of environmental innovation adoption in the printing industry: the importance of task environment. Bus. Strategy Environ. 16 (July), 39–49.
Sáez-Martínez, F. J., Díaz-García, C., & Gonzalez-Moreno, A. (2016). Firm technological trajectory as a driver of eco-innovation in young small and medium-sized enterprises. Journal of Cleaner Production, 138, 28–37. https://doi.org/10.1016/j.jclepro.2016.04.108
Segarra-Ona, M. D. V., Carrascosa-Lopez, C., Segura-Garcia-del-Rio, B., & Peiro-Signes, A. (2011). Empirical analysis of the integration of environmental proactivity into managerial strategy. Identification of benefits, difficulties and facilitators at the Spanish automotive industry. Environmental Engineering & Management Journal (EEMJ), 10(12).
Segarra-Oña, M., Peiró-Signes, Á., & Mondéjar-Jiménez, J. (2016). Twisting the twist: how manufacturing & knowledge-intensive firms excel over manufacturing & operational and all service sectors in their eco-innovative orientation. Journal of Cleaner Production, 138, 19-27.
Segarra-Oña, M. d. V., Peiró-Signes, A., Albors-Garrigós, J., & Miret-Pastor, P. (2011). Impact of innovative practices in environmentally focused firms: Moderating factors. International Journal of Environmental Research, 5, 425–434.

Segarra-Oña, M., Peiró-Signes, A., Albors-Garrigós, J., & Miguel-Molina, B. D. (2017). Testing the social innovation construct: An empirical approach to align socially oriented objectives, stakeholder engagement, and environmental sustainability. Corporate Social Responsibility and Environmental Management, 24(1), 15-27.

Sheehy, B. (2015). Defining CSR: Problems and solutions. Journal of business ethics, 131(3), 625-648.

Sierzchula, W., Bakker, S., Maat, K., & Van Wee, B. (2012). Technological diversity of emerging Eco-innovations: A case study of the automobile industry. Journal of Cleaner Production, 37, 211–220.

Stoppiglia, H., Dreyfus, G., Dubois, R., & Oussar, Y. (2003). Ranking a random feature for variable and feature selection. The Journal of Machine Learning Research, 3, 1399-1414.

Triguero, A., Moreno-Mondéjar, L., & Davia, M. a. (2013). Drivers of different types of eco-innovation in European SMEs. Ecological Economics, 92, 25–33. https://doi.org/10.1016/j.ecolecon.2013.04.009

Tsai, K., Liao, Y., 2017. Sustainability strategy and eco-innovation: a moderation model. Bus. Strateg. Environ. 26 (4), 426–437.

Veugelers, R. (2012). Which policy instruments to induce clean innovating? Research Policy, 41, 1770–1778.

Wagner, M. (2008). Empirical influence of environmental management on innovation: Evidence from Europe. Ecological Economics, 66, 392–402.

Wagner, M., & Llerena, P. (2011). Eco-innovation through integration, reg- ulation and cooperation: Comparative insights from case studies in three manufacturing sectors. Industry and Innovation, 18, 747–764. https://doi.org/10.1080/13662716.2011.621744

Wagner, Marcus, 2007. On the relationship between environmental management, environmental innovation and patenting: evidence from german manufacturing firms. Res. Policy 36 (December (10)), 1587–1602.

Wong, S. K. S. (2013). Environmental requirements, knowledge sharing and green innovation: Empirical evidence from the electronics industry in China. Business Strategy and the Environ- ment, 22, 321–338.

Ziegler, A. and K. Rennings. 2004. Determinants of environmental innovations in Germany: do organizational measures matter? A Discrete Choice Analysis at the Firm Level. ZEW Discussion Paper, vols. 04–30: 1–41.