Why is your company not robotic? The technology and human capital needed by firms to become robotic

María Teresa Ballestar [a,b,*], Aida García-Lazaro [c], Jorge Sainz [b,c], Ismael Sanz [b]

[a] *ESIC Business & Marketing School, Spain*
[b] *Universidad Rey Juan Carlos, Madrid, Spain*
[c] *Institute of Policy Research, University of Bath, United Kingdom*

ABSTRACT

The impact of companies' adoption of robotics is increasingly interesting. This study aims to elucidate how the adoption of these technologies will affect companies and society. Companies that use these technologies expect to gain a competitive advantage, but robotization implies risks that must be managed by companies and governments. This research focuses on one of the most sensitive elements of this transformation process—the workforce.

First, we analyze the characteristics of the workforce and the degree of adoption of robotics using a sample of 4,640 firms with 26 years of observation. We develop a predictive model using a supervised artificial neural network multilayer perceptron (ANN-MLP) to evaluate a company's readiness to make this transformation according to its workforce's characteristics.

Second, we focus on the characterization and segmentation of the companies for which the ANN-MLP is unable to correctly predict the degree of adoption of robotics. This classification failure means that there are unidentified factors that determine why a company has a workforce composition and structure that do not correspond to its expected degree of robotization. For this analysis, we investigate the main business indicators of these companies and cluster them using an unsupervised artificial neural network, specifically the Kohonen self-organizing map.

Our findings will enable companies to understand the importance of transforming to robotics at the right moment, considering factors such as the optimum structure and composition of the workforce. The combination of technology and human capital is the key to boosting the efficiency of the transformation process toward robotics. At this point, a recommendation model to determine whether the company has sufficient maturity to make the transition is crucial for decision makers.

1. Introduction

The automation and robotization of firms are modifying industry and society as they are transforming workers' employment conditions, the skills required to find a "good job," and the future of the workforce's contribution to productivity. There is a heated controversy in academic research and the media concerning how these technologies and changes are affecting employment in the US, Europe, and developing countries. Acemoglu and Restrepo (2019, 2020) differentiate between the effects in the short run, including the sector changes of labor, skills, output, and relative prices in the US. Doraszelski and Jaumandreu (2018) and

Gregory et al. (2019) obtain similar findings for different European countries. Faber (2020) and Kugler et al. (2020) report on the effects of reshoring that robotization is having in developing countries, thus changing trade and industrialization patterns.

However, there is no explanation on how the changes in productivity and the complementarities between some of these technologies occur. There are several plausible explanations: erroneous measurement of creative destruction, which suggests that new products offer greater quality and utility to consumers (Aghion et al., 2019), miscalculations in e-commerce goods' price indices, especially when considering free products in online markets (Brynjolfsson & Oh, 2012; Goolsbee &

Klenow, 2018), or the fact that unemployment grows in developed economies as a result of the reversal in corporate productivity growth following the initial gains of the information and communication technology (ICT) revolution (Syverson, 2011).

The productivity growth in the economies of Western countries that has been achieved this century is due to ICT firms. Its impact has been both direct, through innovative firms, and indirect, through the effects that technological innovation has had on the traditional sector, and has been greater when the digital transformation of firms has occurred (Kraus et al., 2019). This transformation has required complementarities related to workers' knowledge and skills, in both new and established companies, to create what Ribeiro et al. (2018) refer to as a "virtuous circle" within a firm.

To bridge this gap and characterize the technologies and human capital that allow firms to become robotic, we utilize data from the Survey on Enterprise Strategies (Encuesta sobre Estrategias Empresariales or ESEE). The ESEE provides a representative data panel of Spanish industrial companies, collected through an annual survey undertaken by the Spanish Ministry of Finance and Public Administration. It covers a lengthy period, from 1990 to 2016, including the economic recession of 2008 and the subsequent recovery period (Torrent-Sellens, 2018).

Unlike much of the previous research that has focused mainly on the effect of robotization at an aggregate level, this study is unique as it focuses on the individual analysis of each company (Seamans & Raj, 2018). The second new feature of this research is its empirical strategy. To provide new evidence on the effects of robotization, digitization, and innovation on productivity and employment in companies, we investigate why some firms demonstrate anomalous behaviors in their adoption of robotics by using artificial intelligence (AI) strategies, as suggested by Athey and Imbens (2019).

The empirical strategy combines, for the first time to our knowledge, two different types of neural networks to understand the complex decisions of robotic adoption. First, it focuses on explaining, through a supervised artificial neural network (ANN), why firms robotize (Ballestar, Doncel, Sainz, & Ortigosa-Blanch, 2019). The model explains 71.06% of firms' decisions regarding the adoption of robotics and their main drivers. These results are in line with the findings of Acemoglu and Restrepo (2019), Beane and Brynjolfsson (2020), and Gregory et al. (2019). These authors, as well all the previous literature review and only focus on firms that have adopted robotics, thus overlooking the question of why some firms do not automate. Our second contribution is to investigate the approximately 30% of firms that do not demonstrate the expected behavior regarding robotization. We research the causes of that choice through an unsupervised ANN, as defined by Kohonen (1995). The results describe the components of those choices in detail and indicate why firms decide not to automate at the expected time or earlier than expected.

The remainder of the document is structured as follows: Section 2 presents the review of the literature and the theoretical framework; Section 3 describes the data and the empirical strategies for their analysis; Section 4 discusses the results; and Section 5 presents the conclusions of this research and offers proposals for future research.

2. Theoretical framework

Robotics, AI, machine learning (ML), and large volumes of data are inseparably linked in the innovation economy in general but especially in the industrial sector. As they have evolved, companies have focused on the need to offer better products of higher quality, improve their technology, and promote new developments (Dickson & Hadjimanolis, 1998).

The effects of this complementarity are diverse and not yet fully understood. Some issues have aroused great social interest, such as the effect on employment, because this not only entails the replacement of workers but also the shift in the manufacturing process as work performed by machines substitutes human labor to increase the quality and reduce the unit cost (Muro et al., 2019). The impact will be striking for companies: the OECD estimated in 2019 that 14% of current jobs may disappear due to automation and that 35% may be seriously affected by the same issue.

The effect on the labor market will be considerable, either directly or indirectly, owing to the successive displacement of employees expelled from their previous jobs to other jobs and often needing other qualifications. Chui et al. (2016) estimate that, in predictable physical activities, such as production lines and packaging, and depending on the complementarities, substitution could reach 78% in less predictable activities, such as forestry or agriculture, whereas the level of substitution is likely to be 25%, while substitution in services will depend on their characteristics.

Graetz and Michaels (2018) focus on the shift in productivity engendered by this change by analyzing the role of robots using data from the International Federation of Robotics (IFR). Their estimates show no change in the number of hours worked following the increase in robot density but a change in employees' composition, with technological bias in favor of high- and medium-skilled employees. Faber (2020) and Kugler et al. (2020) obtain a similar result but for the impact on outsourcing: the increase in automation density in the US is exerting a negative effect on the delocalization of production from Latin America, thereby creating an effect of asymmetry in skills and knowledge.

Acemoglu and Restrepo (2019), also using data from the IFR, estimate the effect of competition between robots and human labor on employment and wages in the US. They report a net reduction in replacement employment between the two, which may be related to differences in use between sectors. Using similar data, Frey and Osborne (2017) point to a reduction in employment in sectors that are already traditional robot users, such as the automotive sector, as well as in non-tradable sectors, such as services.

Seamans and Raj (2018) believe that aggregate data do not offer a full image of the technological shift and suggest the use of company data. Building on this intuition, Ballestar et al. (2021) and Doraszelski and Jaumandreu (2018) identify a positive bias in technological change toward employment. With the same data, through ML methods, Ballestar et al. (2020) confirm that there is a significant (5%) and growing factor in productivity gains linked to a greater presence of quality human capital in the company, but they do not provide details of how these gains are achieved.

Blanas et al. (2019) use the company data included in the EU KLEMS data panel. Thus, using a relatively simple model of labor demand, they find that workers performing routine and low-skilled tasks, typically women and young people, will be the most affected by the introduction of robots because of productivity gains, while they establish that the entry of robots determines the size of firms' workforces.

These analyses represent an advance in, but do not detail, the characteristics of this change corresponding to the training of employees, the way in which enterprises have changed the knowledge management of their human capital, and the effects of these changes linked to training. Thus, it is important to identify the complementarities of change in employment and productivity to design the training policies needed to avoid unemployment and focus on the most disadvantaged groups.

Autor and Salomons (2018), Ballestar et al. (2021), and Doraszelski and Jaumandreu (2018) show that productivity growth is one of the main benefits of expanding competition, with the results that industries can benefit from labor-increasing productivity through complementarities while employment appears to fall within an industry as industry-specific productivity increases. Positive spillover effects for some sectors more than offset the negative effects in others.

As suggested by Seamans and Raj (2018), empirical work using mainly aggregate statistics by industry or country fails to ascertain how complementarities are hidden under market structures. However, as we have seen, all the research has focused on firms that have embraced the change. Our research concentrates, for the first time to our knowledge, on why some firms behave unexpectedly in their adoption or lack

Table 1

Variables used in the multilayer perceptron artificial neural network (ANN-MLP).

Input Variables	Description
Variables that capture the type of professional relationship between the employee and the company	
PERFTC_PERC	Numerical variable that captures the percentage of full-time workers with permanent employment in the company.
PERFTP_PERC	Numerical variable that captures the percentage of part-time workers with permanent employment in the company.
PERE_PERC	Numerical continuous variable that captures the percentage of temporary workers in the company.
Variables that capture the employees' level of education	
PNT_PERC	Numerical variable that captures the percentage of non-graduate workers in the company.
PTIM_PERC	Numerical variable that captures the percentage of workers in the company who hold a bachelor's degree (4-year university degree).
PIL_PERC	Numerical variable that captures the percentage of workers in the company who hold a master's degree (4 + 2-year university degree).
Variables that capture the economic activity of the company related to employees	
PERTOT	Numerical variable. Number of employees. Following the ESEE survey criteria, small or medium enterprises have between 10 and 200 workers and large companies have more than 200 workers.
CP_PER_PERSON	Numerical variable that captures the costs per employee (thousands of euros). These costs are directly related to a company's investment in employees in terms of internal training, improvement in employees' skills, and wages.
NACE	Categorical variable capturing the economic activity of the company. There are 20 economic activities in the manufacturing sector (Appendix A, Table A1).
Output Variable	**Description**
RBN	Dichotomous variable that is the target value or desired output and indicates whether the company has adopted robotics within the period of observation. Value 1: robotic company; value 0: non-robotic company.

thereof of robotics.

As Currie et al. (2020) and Gómez (2019) show, the new availability of data necessitates new methods in the field of economics and management. Empirical models, as well as any theory that they sustain, are related to a specific situation, but if "the heterogeneity is limited enough, the well-understood past can be informative about the future," as corroborated by Angrist and Pischke (2010). Therefore, our first hypothesis (H1) concerns whether an empirical predictive ML model can evince if a given company, a large company or a small and medium-sized enterprise (SME), should be robotic depending on its structure, composition, workforce characteristics, size, and activity sector.

Our second hypothesis (H2) derives from the previous one. We investigate whether the ML results are as similar to econometric models as those obtained by Acemoglu and Restrepo (2019) or Autor and Salomons (2017) or the AI models by Ballestar et al. (2021). That is, H2 indicates that the robotization of companies involves an increase in the average cost per employee, with two underlying sub-hypotheses: this cost increase is directly related to a company's investment in employees with a higher level of education (H2a); and this cost increase is directly related to the improvement in labor conditions and the type of professional relationship between the employee and the company (H2b).

One of the main issues relates to how automation will affect labor in companies (Autor & Salomons, 2017, 2018; Goos et al., 2009; Gregory et al., 2019). Following the previous literature, we test our third hypothesis (H3) that entails whether a company's workforce itself represents a competitive advantage among companies with the same level of adoption of robotics.

Gaimon (1985) focuses on the optimal mix of automation and labor employed by an organization when introducing new technologies, finding that the optimal mix increases the productivity and output and improves the structure of the firm. Our model also aims to show how the composition, structure, characteristics, and size of the workforce play an important role in the degree of success of the adoption of robotics and, consequently, in its expected benefits in terms of productivity and profitability (H4).

We test the relevance of facing the transformation at the optimum time for the adoption of robotics, following Cotteleer and Bendoly (2006) study on the adoption of information systems. Early or postponed adoption of robotics, before a firm is organizationally prepared, has a negative impact on the degree of success of the adoption of robotics and its expected benefits in terms of productivity and profitability (H5).

3. Data collection and empirical analysis

We analyze data from the ESEE, which is a panel that contains data gathered from Spanish manufacturing businesses from 1990 to 2016, the last available year, conducted by the Spanish Ministry of Finance and Public Administration. This is a very comprehensive database and offers two major benefits. It covers 26 years, including the economic recession and the subsequent recovery. The data also allow us to perform a comprehensive analysis of the microeconomics of productivity and an evaluation of the impact of specific economic events on manufacturing firms (Torrent-Sellens, 2018). To maintain the same size of the panel over the period, it is necessary to add some new companies to replace some that have dropped out. The sample of companies participating in the panel is provided by the Spanish Ministry of Finance, following the EUROSTAT guidelines.

The survey differentiates various company sizes: large companies, which have more than 200 workers, and SMEs, with 10 to 200 workers. The sample comprises all the large companies, stratified proportionally and systematically sampled by industry (CNAE: national economic activity two-digit classification code), and SMEs, sampled by the size of the firm. The sample contains companies from 20 different activity sectors (Appendix A Table A1, Table 1, and Table 2).

Our working sample contains the 4,640 manufacturing firms surveyed in the ESEE, whereof 3,678 (79.27%) are SMEs and 962 (20.73%) are large companies, and they cover two main areas. The first area is strategic decision-making, namely, decision-making on prices, costs, markets, and investment. The second area is the value creation process, which involves human capital, organization, innovation, R&D, and ICT use. In addition, the most relevant indicators and ratios from balance sheets and profit and loss accounts are included.

This research is carried out in two stages. First, we analyze the relationships between human capital, workforce structure, composition, and characteristics, comparing the degree of adoption of robotics by companies. Second, based on these results, we analyze the main business indicators of these companies and group them into clusters to clarify why some companies adopt robotics at a different pace from others.

In the survey, business indicators and the composition, structure, and characteristics of the workforce are available for each year, but the information on the use of robotics is collected only every 4 years. Hence, our analysis focuses on a longitudinal sample of 7 years comprising 12,774 records from the 4,640 companies, from 1991 to 2014, without losing representativeness. This method is equivalent to performing systematic sampling of all the companies in the ESEE survey every 4 years (Maravall & del Río, 2007).

Data mining is performed using Python, creating a single table with 12,774 records, 9,883 for SMEs (77.37%) and 2,891 for large firms (22.63%), which capture the longitudinal information of business indicators and the adoption of robotics over time in the form of 96

Table 2
Total working sample: Descriptive analysis of the input and output variables in the multilayer perceptron artificial neural network from all the companies in the working sample between 1991 and 2014.

ANN Input Variables								
Average	**PERFTC_PERC**	**PERFTP_PERC**	**PERE_PERC**	**PNT_PERC**	**PTIM_PERC**	**PIL_PERC**	**PERTOT**	**CP_PER_PERSON**
Large Firms	86.12%	1.01%	12.86%	83.74%	8.43%	7.81%	850.87	€33.62 K
SMEs	79.13%	2.45%	18.42%	89.74%	5.73%	4.52%	69.44	€24.27 K
Total Portfolio	**80.71%**	**2.12%**	**17.16%**	**88.41%**	**6.33%**	**5.25%**	**246.30**	**€26.39 K**
Std Dev.	**PERFTC_PERC**	**PERFTP_PERC**	**PERE_PERC**	**PNT_PERC**	**PTIM_PERC**	**PIL_PERC**	**PERTOT**	**CP_PER_PERSON**
Large Firms	15.34%	3.28%	15.17%	15.57%	9.61%	9.22%	1,490.32	€13.68 K
SMEs	23.89%	7.05%	23.71%	13.16%	9.10%	7.39%	99.47	€14.92 K
Total Portfolio	**22.43%**	**6.42%**	**22.19%**	**13.95%**	**9.28%**	**7.95%**	**785.57**	**€15.16 K**

Economic Activity (NACE)	1	2	3	4	5	6	7	8	9	10
Portfolio	149	460	99	442	169	163	146	244	296	244
Portfolio (%)	3.21%	9.91%	2.13%	9.53%	3.64%	3.51%	3.15%	5.26%	6.38%	5.26%
ow Large Firms (%)	20.13%	22.39%	34.34%	13.12%	1.78%	9.82%	23.97%	17.62%	40.54%	16.80%
ow SMEs (%)	79.87%	77.61%	65.66%	86.88%	98.22%	90.18%	76.03%	82.38%	59.46%	83.20%

Economic Activity (NACE)	11	12	13	14	15	16	17	18	19	20	Total
Portfolio	333	131	516	268	139	207	190	86	238	120	4,640
Portfolio (%)	7.18%	2.82%	11.12%	5.78%	3.00%	4.46%	4.09%	1.85%	5.13%	2.59%	100.00%
ow Large Firms (%)	18.92%	40.46%	11.24%	20.52%	35.97%	37.20%	38.95%	31.40%	4.62%	9.17%	20.73%
ow SMEs (%)	81.08%	59.54%	88.76%	79.48%	64.03%	62.80%	61.05%	68.60%	95.38%	90.83%	79.27%

ANN Output Variable	
Average	**Robotic (RBN=1)**
Large Firms	50.95%
SMEs	19.73%
Total Portfolio	26.80%

variables. After conducting iterative descriptive and causal analyses, 20 of these variables are found to be statistically significant and relevant for the empirical analysis in this research. These variables are calculated by using the same method for both SMEs and large firms. The descriptive and causal analyses of these variables are detailed in the following sections (Sections 3.1 and 3.2). The causal analysis is developed using both supervised and non-supervised neural networks in two different stages. Neural networks are able to capture causal and interactive relationships, and they have the capability of general function approximation, being able to capture both linear and non-linear relationships (Buckler, 2001; Garbe & Richter, 2009).

3.1. Empirical analysis: First stage

The contribution of this research is dual. In the first stage, we focus on the development of an empirical ML predictive model based on an ANN, as designed by Ballestar, Doncel, Sainz, and Ortigosa-Blanch (2019), Ballestar, Grau-Carles, and Sainz (2019). This model determines whether a given company should be robotic within the period of observation. This will depend on the structure, composition, and characteristics of the workforce, as well as the size of the company and its activity sector. This predictive model is valuable for decision makers in companies to reduce uncertainty when evaluating the readiness of the workforce to adopt robotics, thereby allowing them to avoid unnecessary risks derived from adopting the technology prematurely or the costs of opportunity arising from a postponed decision and its late adoption (Ballestar et al., 2021).

3.1.1. First stage: Method

Artificial neural networks are mathematical models that can be implemented easily. They can manage and analyze large datasets even when complex relationships, both linear and non-linear, exist between the input and output variables. These characteristics make ANNs an attractive alternative to the traditional methods that are widely used for developing predictive models when the dependent variable is dichotomous, such as logistic regression (Paliwal & Kumar, 2009; Tu, 1996).

This research employs a feed-forward multilayer perceptron (MLP) ANN model, which is one of the most popular types of ANN (Ballestar, Doncel, Sainz, & Ortigosa-Blanch, 2019; Ballestar, Grau-Carles, & Sainz, 2019; Hu & Weng, 2009; Kavzoglu & Mather, 2003). This is a supervised method that supports both qualitative and quantitative variables as input variables but requires a working sample with target and non-target real cases in the output variable to perform the learning process (Aad et al., 2012). Consequently, the accuracy of the resulting predictive model will depend on the size and variety of cases available in the working sample, which must be representative of all the possible relationships between the independent and the dependent variables (Li & Eastman, 2006).

In this model, the input variables correspond to the independent variables, which capture the type of professional relationship between the employees and the company, the employees' level of education, the cost per employee, the number of employees, and the activity sector of the company. The output variable corresponds to the dependent variable, which is the target value and ascertains whether the company, with the specified characteristics defined by the input variables, should be robotic within the period of observation. These variables are described in Table 1, and their descriptive analysis is detailed in Table 2.

The MLP-ANN model is trained using 12,774 records from 4,640 companies within the period of observation between 1991 and 2014. We use the training–testing–validation method, which randomly splits the working sample into three groups as follows: the training sample is used to create the model and train it, and it comprises 69.5% of the working sample, with 8,882 records. The testing sample is an independent dataset that is used to track the MLP-ANN errors within the training process and prevent overtraining. It consists of 20.5% of the working sample, with 2,617 records. Finally, the validation or holdout sample is also an independent dataset that is used to validate the predictive power and stability of the model and accounts for 10% of the working sample, with 1,275 records (Flexer, 1996).

Because the companies that have adopted robotics only represent 38.28% of the sample (26.80% of the total records), the dataset is unbalanced. This means that the group of robotic companies in the target

variable is less represented in the sample within the period of observation than the group of non-robotic ones.

This type of bias, which can negatively affect the classification performance of the MLP-ANN, is managed by applying an oversampling method to the minority group in the training sample to balance the weights between target (robotic) and non-target (non-robotic) companies (Ballestar, Doncel, Sainz, & Ortigosa-Blanch, 2019; Ballestar, Grau-Carles, & Sainz, 2019; Ganganwar, 2012; Guo & Viktor, 2004; Sun et al., 2009). The testing and validation or holdout samples remain unbalanced. In that way, the MLP-ANN shows its predictive power and stability in a real-world scenario after being trained with the balanced sample. We apply a *marginalization* method to handle 183 records (1.43%) with missing values that are randomly distributed in any of the input variables without biasing the working sample (Wagstaff, 2004; Yu et al., 2014).

Thus, the MLP-ANN architecture is structured in three layers. The connection starts in the input layer, with 28 units that receive values from 9 independent unit variables. These 28 units in the input layer are the sum of 8 covariates (quantitative variables) and 20 categories on the factor level (categorical variable). Thereafter, the connection passes through the hidden layer, which has six units. Finally, the connection ends in the output layer, which has two units that correspond to the dependent output variable. The first unit refers to the target value (companies that have adopted robotics within the period of evaluation), and the second unit refers to the non-target value (companies that have not adopted robotics within the period of evaluation) (Fig. 1).

In our MLP-ANN, the hyperbolic tangent is the activation function for all the units in the hidden layer, and the softmax function is the activation function for the two units in the output layer. The training process is conducted using a backpropagation algorithm to calculate the weights in the MLP-ANN that minimize the cross-entropy error between the target and the predicted outcome.

3.1.2. First stage: Empirical analysis and results

In the following sections, we focus on evaluating the performance of the MLP-ANN using the classification accuracy, its sensitivity, its specificity, the area under the receiver operating characteristic (ROC) curve (AUC), and the Gini coefficient. We also analyze the normalized importance of each input variable in the estimating model, as well as the

MLP-ANN results and implications.

3.1.2.1. Evaluation criteria of the MLP-ANN model. The overall classification of the model is 71.06% (an error rate of 28.94%). As the output variable in this MLP-ANN is dichotomous, the accuracy represents the percentage of success when the model predicts whether a company should be robotic according to its size and activity sector, as well as the structure, composition, and characteristics of its workforce.

The classification accuracy and confusion matrix of the MLP-ANN model by sample (training, testing, and holdout) is shown in Table 3. The percentage of correctly classified cases is similar across the three samples, indicating that the model is not overtrained.

The *sensitivity* or percentage of true positives is 73.27%. This value refers to the percentage of companies that are correctly classified as robotic based on their observed characteristics in a specific year. The *specificity* or percentage of true negatives is 68.90%. This value refers to the percentage of companies that are correctly classified as non-robotic based on their observed characteristics in a specific year.

The complementary value is the percentage of false positives, which is 31.10% and refers to the percentage of non-robotic companies that are incorrectly classified as robotic by the model based on their observed characteristics in a specific year. Finally, false negatives account for 26.73% of robotic companies that are incorrectly classified as non-robotic by the model based on their observed characteristics in a specific year.

In this research, we use the AUC as the main classification performance indicator of the MLP-ANN as it is more robust than the classification accuracy indicator when the sample is imbalanced (Ballestar, Doncel, Sainz, & Ortigosa-Blanch, 2019; Ballestar, Grau-Carles, & Sainz, 2019; Chen et al., 2008; Dželihodžić & Jonko, 2016; Jensen, 1992). The AUC values range from 0.5, meaning that the model makes a random classification, to 1, meaning that the model makes a perfect classification.

We evaluate the testing sample (AUC equals 0.768) and the holdout sample (AUC equals 0.774) but not the training sample, which could be misleading. Both AUCs are higher than 0.7, indicating that the model is good (Hosmer, Lemeshow, & Sturdivant, 2013).

In addition to the AUC, we calculate the Gini coefficient, which is related to the AUC as it represents twice the area between the ROC and
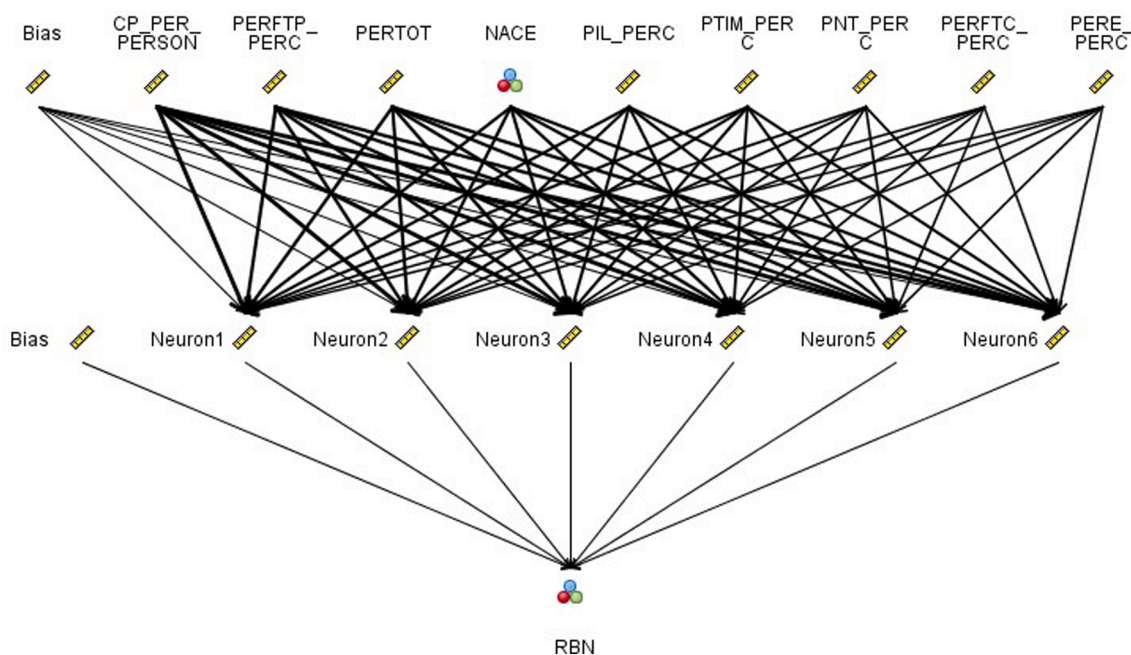

Fig. 1. Multilayer perceptron artificial neural network (MLP-ANN) architecture.

Table 3
Model accuracy and confusion matrix.

Model Accuracy		Confusion Matrix						
				Predicted				
				Sample Size		**Percentage**		
Sample	**Percentage Correct**	**Sample**	**Observed**	**0**	**1**	**0**	**1**	
Training	71.06%	**Training***	0	4,437	2,003	68.90%	31.10%	
			1	1,681	4,608	26.73%	73.27%	
Test	69.49%	**Test**	0	1,281	593	68.36%	31.64%	
			1	191	505	27.44%	72.56%	
Holdout	70.71%	**Holdout**	0	640	285	69.19%	30.81%	
			1	84	251	25.07%	74.93%	

* Oversampled to balance the sample.

the diagonal. The Gini coefficients are 0.53 for the testing sample and 0.54 for the holdout sample.

We also calculate the relative importance of the input variables in estimating the MLP-ANN model. The variables' importance has no relationship with the model's accuracy as their only purpose is to provide information about the importance of each input variable when making the overall prediction. As they are relative values, their total sum is 1 (Ballestar, Doncel, Sainz, & Ortigosa-Blanch, 2019; Ballestar, Grau-Carles, & Sainz, 2019). Three variables accumulate 49% of the relative importance: the cost per employee (0.21), the percentage of part-time workers with permanent employment in the company (0.14), and the number of employees in the company (0.14). The following six variables account for the other 51% as follows: the economic activity sector (0.12), the percentage of workers with a master's degree (0.11), the percentage of workers with a bachelor's degree (0.10), the percentage of non-graduate workers (0.07), the percentage of full-time workers (0.05), and the percentage of temporary workers (0.05).

3.1.2.2. MLP-ANN results. Table 4 shows the descriptive statistics for 71.06% of the cases from companies in which the adoption of robotics is correctly predicted by the MLP-ANN model. These findings allow decision makers in companies to determine whether their workforce is ready to face the transition to robotization or whether structural changes still need to be made to achieve success in their transformation. Following these benchmarks, as shown in Table 4, companies can redirect their efforts to the most relevant aspects when defining the human capital strategies in their robotics adoption life cycle. These results confirm H1 and are in line with those of Acemoglu and Restrepo (2019), Autor and Salomons (2017), and Ballestar et al. (2021), thus reinforcing the idea that the new availability of data calls for new methods in the field of economics and management (Currie et al., 2020; Gómez, 2019).

The costs per employee are higher in large firms than in SMEs, and they increase when companies make the transition to robotics. The minimum average costs per employee are €20.71 K for non-robotic SMEs, and the maximum average costs per employee are €35.27 K for large robotic companies. These findings corroborate H2 and are in line

Table 4
Descriptive statistics from the 71.06% of cases correctly predicted by the model (MLP-ANN).

Average	PERFTC_PERC	PERFTP_PERC	PERE_PERC	PNT_PERC	PTIM_PERC	PIL_PERC	PERTOT	CP_PER_PERSON
Non-robotics	**76.50%**	**2.80%**	**20.69%**	**91.51%**	**4.64%**	**3.83%**	**46.64**	**€20.86 K**
Large	82.00%	1.39%	16.60%	83.14%	5.12%	11.72%	364.99	€26.14 K
SMEs	76.34%	2.84%	20.81%	91.76%	4.63%	3.59%	37.21	€20.71 K
Robotics	**85.81%**	**1.31%**	**12.86%**	**84.40%**	**8.80%**	**6.78%**	**637.63**	**€34.12 K**
Large	86.74%	1.10%	12.15%	83.36%	9.26%	7.37%	1015.83	€35.27 K
SMEs	84.65%	1.58%	13.75%	85.71%	8.23%	6.04%	162.69	€32.67 K
Combined	79.09%	2.38%	18.51%	89.53%	5.80%	4.65%	211.13	€24.55 K

Std Dev.	PERFTC_PERC	PERFTP_PERC	PERE_PERC	PNT_PERC	PTIM_PERC	PIL_PERC	PERTOT	CP_PER_PERSON
Non-robotics	**26.11%**	**8.11%**	**25.96%**	**13.39%**	**9.25%**	**7.90%**	**78.87**	**€15.06 K**
Large	22.66%	7.93%	21.57%	19.54%	9.46%	15.38%	212.21	€14.54 K
SMEs	26.19%	8.11%	26.07%	13.09%	9.25%	7.44%	44.55	€15.05 K
Robotics	**14.61%**	**3.01%**	**14.56%**	**13.36%**	**9.13%**	**6.81%**	**1407.54**	**€12.11 K**
Large	13.05%	2.73%	12.90%	14.97%	9.92%	7.87%	1794.11	€12.77 K
SMEs	16.30%	3.30%	16.37%	10.89%	8.00%	5.11%	151.34	€11.06 K
Combined	23.85%	7.10%	23.61%	13.76%	9.41%	7.72%	791.13	€15.49 K

Economic Activity (NACE)	1	2	3	4	5	6	7	8	9	10
Non-robotics	**4.05%**	**9.87%**	**1.41%**	**13.61%**	**5.64%**	**4.62%**	**3.17%**	**7.08%**	**7.79%**	**3.38%**
Large	8.70%	2.50%	0.00%	4.15%	1.25%	2.29%	1.11%	7.96%	25.34%	0.00%
SMEs	91.30%	97.50%	100.00%	95.85%	98.75%	97.71%	98.89%	92.04%	74.66%	100.00%
Robotics	**2.87%**	**10.81%**	**3.55%**	**3.38%**	**0.17%**	**1.69%**	**3.21%**	**2.36%**	**7.60%**	**5.74%**
Large	60.71%	52.03%	52.50%	80.00%	100.00%	50.00%	59.38%	93.33%	86.54%	32.38%
SMEs	39.29%	47.97%	47.50%	20.00%	0.00%	50.00%	40.63%	6.67%	13.46%	67.62%

Economic Activity (NACE)	11	12	13	14	15	16	17	18	19	20	Total
Non-robotics	**5.20%**	**1.20%**	**10.30%**	**6.50%**	**1.40%**	**2.10%**	**1.30%**	**1.40%**	**6.40%**	**3.50%**	**100.0%**
Large	0.70%	8.80%	0.00%	1.10%	0.00%	0.00%	2.60%	0.00%	0.00%	3.00%	4.30%
SMEs	99.30%	91.20%	100.00%	98.90%	100.00%	100.00%	97.40%	100.00%	100.0%	97.00%	95.70%
Robotics	**7.60%**	**4.60%**	**6.30%**	**7.10%**	**6.10%**	**9.30%**	**11.00%**	**3.40%**	**1.70%**	**1.70%**	**100.0%**
Large	37.50%	50.90%	27.40%	68.90%	60.00%	61.80%	50.80%	54.10%	27.00%	62.50%	50.30%
SMEs	62.50%	49.10%	72.60%	31.10%	40.00%	38.20%	49.20%	45.90%	73.00%	37.50%	49.70%

with Eisfeldt et al. (2021), who find that high-skilled workers bring substantial returns to their human capital, with strong complementarities between high-skilled labor and physical capital.

In SMEs, robotization involves a 57.77% increase in the average costs per employee (from €20.71 K in non-robotic SMEs to €32.67 K in robotic companies). This increase is linked to the increase in the employees' level of education, with a decrease of 6.59% in the average percentage of non-graduate employees in favor of employees with bachelor's and master's degrees, and to an improvement in the labor conditions, with an increase of 10.89% in the average percentage of full-time permanent workers.

In large companies, robotization involves a 34.95% increase in the average costs per employee (from €26.14 K in non-robotic to €35.27 K in robotic companies). This increase is not linked to the increase in the average level of education of their employees but to the improvement in labor conditions, with an increase of 5.78% in the average percentage of full-time permanent workers. Hence, robotization, for both large companies and SMEs, is linked to an improvement in labor conditions regarding both wages and stability. This result confirms H2b.

The robotization of large companies increases the average percentage of employees with a bachelor's degree by 80.74% to 9.26%. In the case of SMEs, the average percentage of employees with bachelor's degrees increases by 77.67% to 8.23%. Meanwhile, in SMEs, this has the effect of reducing the average percentage of non-graduates to 85.71% (6.60% lower than the average for non-robotic SMEs); in large companies, the reduction occurs among the average percentage of employees with a master's degree by 7.37% (37.12% lower than the average for large non-robotic companies). These results partially confirm H2a, as it is confirmed for SMEs but not for large companies.

Even though these conclusions are equally applicable to all the sectors, there are significant differences in the degree of adoption of robotics between them. The least robotic sector is "textiles and clothing manufacturing" (NACE code 4), which consists of 13.61% of non-robotic companies, and the most robotic is "motor vehicles" (NACE code 17), which accounts for 11.0% of the robotic companies.

3.2. Empirical analysis: Second stage

In the first stage of this research, we identify companies from the working sample that have demonstrated unexpected behavior regarding their adoption of robotics. This is based on observations relating to the structure, composition, and characteristics of their workforce in a specific year of observation. This unexpected behavior causes the MLP-ANN to misclassify a company as robotic or not in 28.94% (3,781 records) of the evaluated cases.

In the second stage, we focus on the characterization and segmentation of these misclassified cases of companies into groups that allow us to understand the reasons for such unexpected behavior. The failure of classification meant that there are unidentified factors that determine why a company, at a specific moment in time, is applying a workforce design strategy that does not correspond to its expected degree of adoption of robotics. This classification failure occurs in two different ways. On the one hand, the ANN-MLP determines that a company should be robotic, but it has not made the transition yet. This situation represents 23.24% of the misclassified cases, with 2,881 records from 1,531 companies. On the other hand, the ANN-MLP sometimes indicates that a company should not be robotic but has already adopted robotics. This represents 5.70% of the misclassified cases, with 900 records from 633 companies.

For this analysis, we consider the main business indicators from the whole working sample of companies and focus on the two groups of misclassified cases of companies. We use a neural network analysis as an unsupervised clustering method, specifically the Kohonen self-organizing map (SOM).

3.2.1. Second stage: Method

The Kohonen SOM is a type of unsupervised ANN that is commonly used for performing clustering or dimension reduction without modifying the topology of the original dataset (Kohonen, 1995). In this stage, we develop two clustering models using Kohonen's SOMs as our aim is to group the two types of misclassified cases of companies by using a collection of their main business indicators. These business indicators belong to several categories, such as productivity and labor, firm

Table 5
List of variables used by the Kohonen self-organizing maps (SOMs).

Variables	Description
Productivity and Labor	
PTN	Numerical variable that captures the productivity per employee (thousands of euros).
PHN	Numerical variable that captures the productivity per hour worked (euros).
HETN_PER_PERSON	Numerical variable that captures the total effective hours worked per employee (thousands of hours).
PURE_SME	Dichotomous variable that indicates whether the company has been an SME during its participation in the panel, taking the value 1 if the company has always been an SME and the value 0 otherwise. According to the ESEE survey criteria, small or medium enterprises have between 10 and 200 workers and large companies have more than 200 workers.
Firm Performance	
ADDED_VALUE_PER_PERSON	Numerical variable that captures the added value generated by each employee in the company (thousands of euros). It is calculated as the total gross added value generated by the company divided by the number of workers.
FIN_ASSETS_PER_PERSON	Numerical variable that captures the financial assets of the company that can be converted into cash in the short term per employee. It represents essential information for characterizing the financial profile of the company (thousands of euros). It is calculated as the total assets divided by the number of workers.
SALES_PER_PERSON	Numerical variable that captures the volume of sales of the company per employee (thousands of euros). It is calculated as the total sales generated by the company divided by the number of workers.
VEXPORT_PER_PERSON	Numerical variable that captures the value of exports made by the company per employee (thousands of euros). It is used to track the company's internationalization. It is calculated as the total volume of exports generated by the company divided by the number of workers.
Attitude towards Robotics	
ROBOT_PROFILE_STR	Categorical variable that describes the relationship of the company to the adoption of robotics. Characterization of the company by its journey toward robotization within the period of observation. • **Always Robotic**: These companies have always been robotic within the period of observation. • **Migration to Robotics – Churned**: These companies transitioned from non-robotic to robotic but abandoned robotics later for some reason. • **Migration to Robotics – Loyal**: These companies transitioned from non-robotic to robotic and remained loyal to this adoption of the technology. • **Never Robotic**: These companies never adopted robotics within the period of observation. • **Robotics – Churned**: These companies started to participate in the panel, having adopted robotics, but they ceased robotization within the period of observation for some reason.

Table 6
Descriptive analysis of the variables used by the Kohonen self-organizing maps (SOMs) for all the companies in the working sample between 1991 and 2014.

Average	PTN	PHN	HETN_PER_PERSON	ADDED_VALUE_PER_PERSON	FIN_ASSETS_PER_PERSON	SALES_PER_PERSON	VEXPORT_PER_PERSON
Large Firms	€57.30 K	€32.89	1.76 K	€56.94 K	€217.38 K	€218.25 K	€71.75 K
SMEs	€35.88 K	€20.36	1.80 K	€36.14 K	€119.58 K	€126.04 K	€27.75 K
Total Portfolio	€40.71 K	€23.15	1.79 K	€40.84 K	€142.01 K	€146.89 K	€37.68 K

Std Dev.	PTN	PHN	HETN_PER_PERSON	ADDED_VALUE_PER_PERSON	FIN_ASSETS_PER_PERSON	SALES_PER_PERSON	VEXPORT_PER_PERSON
Large Firms	€90.70 K	€53.42	0.17 K	€88.64 K	€449.83 K	€331.62 K	€162.90 K
SMEs	€30.47 K	€17.39	0.47 K	€33.35 K	€198.75 K	€175.31 K	€86.44 K
Total Portfolio	€51.50 K	€29.96	0.42 K	€52.08 K	€280.24 K	€223.90 K	€110.04 K

ROBOT_PROFILE_STR	Always Robotic	Migration to Robotics – Churned	Migration to Robotics – Loyal	Never Robotic	Robotics – Churned	Total
Large Firms	271	149	160	315	67	962
Large Firms (%)	28.17%	15.49%	16.63%	32.74%	6.96%	100.00%
SMEs	347	283	294	2549	205	3678
SMEs (%)	9.43%	7.69%	7.99%	69.30%	5.57%	100.00%
Total Portfolio	618	432	454	2,864	272	4,640
Total Firms (%)	13.32%	9.31%	9.78%	61.72%	5.86%	100.00%

PURE_SME	Total
Large Firms	1,302
Large Firms (%)	28.06%
SMEs	3,338
SMEs (%)	71.94%
Total Portfolio	4,640
Total Firms (%)	100.00%

performance, and their relationship with the adoption of robotics.

There are no target variables in the models as they are unsupervised, and the input variables, corresponding to 15 neurons, are the same for the two clustering models. The input variables are listed in Table 5, and the descriptive analysis is presented in Table 6.

The architecture of the two neural networks consists of two layers, the input and the output layer, in which the basic units, the neurons, are organized. The function of these neurons is simply to capture the information from the input variables. Each neuron in the input layer is connected to all the output neurons in the output layer; this is also referred to as the output map. These connections can have different weights or strengths, which are initially established in a random way, and they are managed via an n-dimensional weight vector (Richardson et al., 2003).

The output map is a two-dimensional array of neurons that have lateral connections to other adjacent neurons. This means that they have a neighborhood relationship as each output neuron's activity can influence its neighbors. In our two models, the dimensions of the maps are 3×3, as small output layers are faster to train and generalize better (Choudhary & Bhattacharyya, 2002; Goswami et al., 2011). In fact, one of the main advantages of Kohonen's SOM is its capacity to find the right balance between an efficient data compression and a sufficiently faithful preservation of the original data topology to support good generalization (Ritter, 1999).

The unsupervised ANN model learns through the training process, which consists of an iterative and sequential process in which the multidimensional input space, in the form of input data vectors, is projected onto the output two-dimensional space, and then the selection of the winner output neuron is made.

The winner output neuron, or best-matching unit, is the one that presents the weight vector with the smallest Euclidean distance between itself and the input data vector. The winner neuron becomes the center of the neighborhood, and its weighted connection to the input data vector is updated. The weighted connection of its neighborhood to the input data vector is updated accordingly too. The training process of the ANN is repeated until just one output neuron in the neighborhood is activated and there are no changes in the weights between the neurons from the input and output layers (Tian et al., 2014). This competition between output neurons to be activated for the classification of the input patterns is possible because of the existence of lateral connections between them and because of the usage of a learning competitive algorithm, the neighborhood function (Afolabi & Olude, 2007; Williams et al., 2014).

The resultant Kohonen SOM map consists of a grid containing many strong units, which are the cluster centers and represent the patterns from the input data. There are also some weak units that do not correspond to any specific pattern, and they are removed from the network (Ha, 2007).

In our research, the two models are trained in two phases. The first phase is used to capture the gross patterns of the data, which run 20 training cycles. The second is a tuning phase to make adjustments to finer features of the data, which run 150 training cycles. The specific cluster to which each input record is assigned by the model corresponds to the combinations of the X and Y coordinates of the winning neuron.

3.2.2. Second stage: Empirical analysis and results

First, we construct a Kohonen SOM for **Group 1 of records from companies that do not adopt robotics** at the expected time. This group represents 23.24% of the working sample, with 2,881 records and 5 generated clusters. This segmentation is based on the business indicators of productivity and labor, firm performance, and the robotics adoption life cycle of the company, as described in Table 5.

The percentage of the sample in each cluster is as follows (Fig. 2): Cluster 1 (X = 0; Y = 0): 27.49%; Cluster 2 (X = 0; Y = 2): 33.29%; Cluster 3 (X = 1; Y = 2): 5.24%; Cluster 4 (X = 2; Y = 0): 16.52%; and Cluster 5 (X = 2; Y = 2): 17.46%. The taxonomy of the cluster profiles for
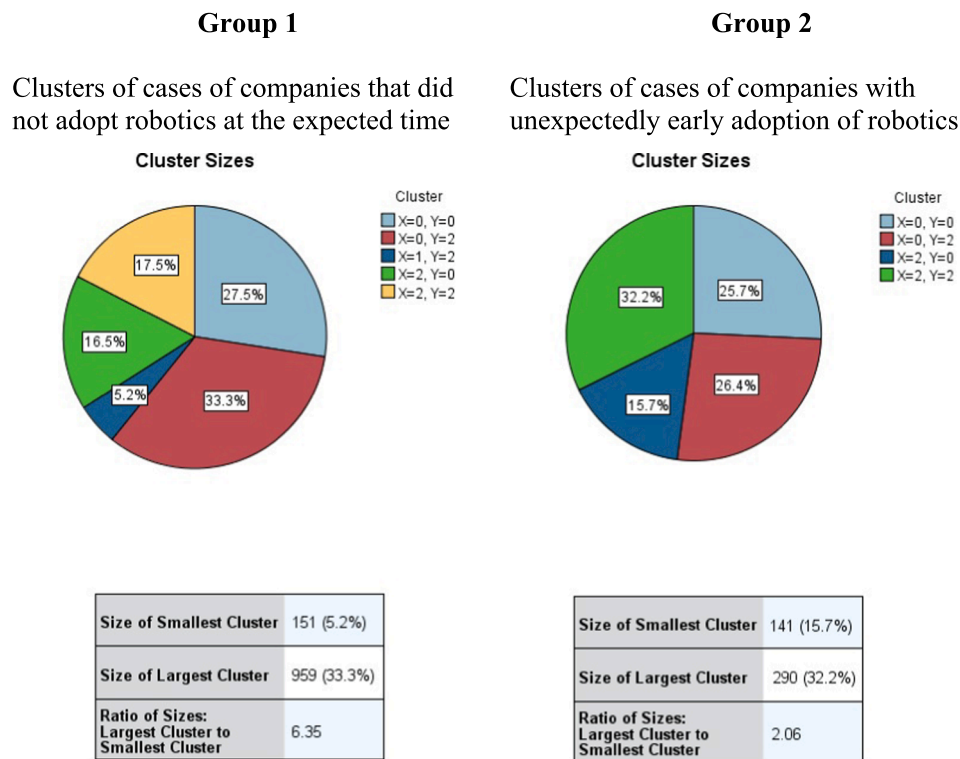
Group 1

Clusters of cases of companies that did not adopt robotics at the expected time

Group 2

Clusters of cases of companies with unexpectedly early adoption of robotics

Fig. 2. Cluster distribution chart for the two misclassified groups of companies.

Group 1 is presented in Fig. 2 and Table 7.

Second, we create a Kohonen SOM for **Group 2 of records from companies with unexpectedly early adoption of robotics**. This group represents 5.70% of the working sample, with 900 records and 5 generated clusters. This segmentation is also based on the same business indicators of productivity and labor, firm performance, and robotics adoption life cycle of the company described in Table 5.

The percentage of the sample in each cluster is as follows (Fig. 2): Cluster 1 (X = 0; Y = 0): 25.67%; Cluster 2 (X = 0; Y = 2): 26.44%; Cluster 3 (X = 2; Y = 0): 15.67%; and Cluster 4 (X = 2; Y = 2): 32.22%. The taxonomy of the cluster profiles for Group 2 is presented in Fig. 2 and Table 8.

Additionally, we calculate the silhouette coefficient statistical measure to assess the goodness of fit of the two Kohonen SOM models (Rousseeuw, 1987). This measure evaluates the consistency of the clustering structure by assessing the cohesion between the records within a cluster and the separation between clusters. The silhouette coefficient values range from −1 to 1: −1 means that the model is very poor, values higher than 0.5 mean that the model is good, and 1 represents when the model is optimal (Aad et al., 2012; Ballestar et al., 2018). The silhouette coefficient is 0.8 for the first group (Group 1) of companies with a lack of robotics adoption at the expected time and 0.7 for the second group (Group 2) of companies with unexpectedly early adoption of robotics, so the two models are robust.

4. Results and discussion

In this section, we describe the clustering structures of both groups of anomalous companies: those that do not adopt robotics at the expected time (Group 1) and those with unexpectedly early adoption of robotics (Group 2).

4.1. Group 1: Cases of companies that do not adopt robotics at the expected time

This group is comprises companies that do not adopt robotics at the expected time. The MLP-ANN predicts that these companies should have been robotic in the period of observation (2,881 records, belonging to 1,531 companies). This group is classified into five clusters based on nine business indicators from categories such as productivity and labor, firm performance, and the robotic adoption life cycle (Table 5). These clusters are described in the following sections using the results from the MLP-ANN and the Kohonen SOM model in Tables 4 and 7.

4.1.1. Clusters of companies that are reluctant to adopt robotics

Cluster 1 encompasses large companies (27.49%, 792 records), none of which have adopted robotics. Cluster 2 consists of SMEs (33.29%, 959 records), most of which (78.92%) do not adopt robotics within the period of observation. The two clusters together represent 60.78% of the cases examined. The MLP-ANN predicts that they would be robotic because their workforce composition, characteristics, and structure correspond to a robotic company. However, their size is far from the average robotic company (492,87 employees for large firms and 65.73 employees for SMEs), being 51.48% and 59.60% smaller than the average robotic company, respectively.

The reason behind this misclassification is that the companies in both clusters outperform the average non-robotic company in all the business indicators analyzed in the Kohonen SOM. These outperformers have a competitive advantage over the average non-robotic company, such as greater workforce productivity and higher sales per employee, which makes them very profitable and less prone to adopt robotics (€49.55 K, which is 17.45% higher productivity per employee for large companies, and €47.60 K, which is 62.70% higher productivity per employee for SMEs; and €201.20 K, which is 42.43% higher sales per employee for large companies, and €163.77 K, which is 66.42% higher sales per employee for SMEs).

This means that the workforce specifically represents a competitive

Table 7
Cluster profiles: Centroids of numerical variables and frequencies of categorical variables for **Group 1 (companies that did not adopt robotics).**

Average by Cluster	PTN	PHN	HETN_PER_PERSON	ADDED_VALUE_PER_PERSON	FIN_ASSETS_PER_PERSON	SALES_PER_PERSON	VEXPORT_PER_PERSON
Cluster 1 (X = 0; Y = 0)	€49.55 K	€28.45	1.76 K	€49.47 K	€173.51 K	€201.20 K	€55.36 K
Cluster 2 (X = 0; Y = 2)	€47.60 K	€27.06	1.79 K	€47.58 K	€157.34 K	€163.77 K	€38.75 K
Cluster 3 (X = 2; Y = 0)	€51.33 K	€29.27	1.78 K	€51.73 K	€205.41 K	€188.00 K	€50.94 K
Cluster 4 (X = 2; Y = 2)	€64.44 K	€37.44	1.74 K	€63.03 K	€252.11 K	€236.34 K	€87.74 K
Cluster 5 (X = 1; Y = 2)	€45.35 K	€25.56	1.87 K	€47.11 K	€175.20 K	€168.69 K	€30.10 K
Combined	**€51.52 K**	**€29.49**	**1.78 K**	**€51.43 K**	**€187.08 K**	**€190.97 K**	**€53.42 K**

Std Dev. by Cluster	PTN	PHN	HETN_PER_PERSON	ADDED_VALUE_PER_PERSON	FIN_ASSETS_PER_PERSON	SALES_PER_PERSON	VEXPORT_PER_PERSON
Cluster 1 (X = 0; Y = 0)	€36.63 K	€21.24	0.17 K	€36.71 K	€217.31 K	€479.04 K	€101.90 K
Cluster 2 (X = 0; Y = 2)	€41.68 K	€23.49	0.40 K	€41.06 K	€194.51 K	€188.07 K	€91.48 K
Cluster 3 (X = 2; Y = 0)	€46.76 K	€26.91	0.25 K	€47.20 K	€496.09 K	€186.52 K	€131.72 K
Cluster 4 (X = 2; Y = 2)	€190.55 K	€112.32	0.19 K	€183.18 K	€456.30 K	€342.45 K	€260.58 K
Cluster 5 (X = 1; Y = 2)	€26.35 K	€14.97	0.79 K	€30.96 K	€205.17 K	€184.03 K	€51.44 K
Combined	**€87.26 K**	**€51.06**	**0.34 K**	**€84.79 K**	**€325.01 K**	**€321.55 K**	**€144.25 K**

ROBOT_PROFILE_STR	Migration to Robotics – Churned		Migration to Robotics – Loyal		Never Robotic		Robotics – Churned	
Cluster	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Cluster 1 (X=0; Y=0)	0	0.0%	0	0.0%	351	42.09%	0	0.00%
Cluster 2 (X=0; Y=2)	0	0.0%	61	24.90%	483	57.91%	68	45.03%
Cluster 3 (X=2; Y=0)	0	0.0%	184	75.10%	0	0.00%	83	55.97%
Cluster 4 (X=2; Y=2)	212	70.43%	0	0.00%	0	0.00%	0	0.00%
Cluster 5 (X=1; Y=2)	89	29.57%	0	0.00%	0	0.00%	0	0.00%
Combined	**301**	**100.00%**	**245**	**100.00%**	**834**	**100.00%**	**151**	**100.00%**

PURE_SME	Value 0 (No)			Value 1 (Yes)	
Cluster	Frequency		Percentage	Frequency	Percentage
Cluster 1 (X=0; Y=0)	351		42.29%	0	0.00%
Cluster 2 (X=0; Y=2)	0		0.00%	612	87.30%
Cluster 3 (X=2; Y=0)	267		32.17%	0	0.00%
Cluster 4 (X=2; Y=2)	212		25.54%	0	0.00%
Cluster 5 (X=1; Y=2)	0		0.00%	89	12.70%
Combined	**830**		**100.00%**	**701**	**100.00%**

Table 8
Cluster profiles: Centroids of numerical variables and frequencies of categorical variables for **Group 2 (companies with unexpectedly early adoption of robotics).**

Average by Cluster	PTN	PHN	HETN_PER_PERSON	ADDED_VALUE_PER_PERSON	FIN_ASSETS_PER_PERSON	SALES_PER_PERSON	VEXPORT_PER_PERSON
Cluster 1 (X = 0; Y = 0)	€40.02 K	€22.89	1.78 K	€40.36 K	€136.77 K	€143.28 K	€38.11 K
Cluster 2 (X = 0; Y = 2)	€35.55 K	€19.98	1.77 K	€35.52 K	€118.52 K	€120.67 K	€23.21 K
Cluster 3 (X = 2; Y = 0)	€52.42 K	€30.34	1.82 K	€52.91 K	€151.69 K	€187.18 K	€60.12 K
Cluster 4 (X = 2; Y = 2)	€36.89 K	€21.02	1.76 K	€36.68 K	€118.15 K	€117.09 K	€22.64 K
Combined	**€39.75 K**	**€22.65**	**1.78 K**	**€39.87 K**	**€128.23 K**	**€135.74 K**	**€32.64 K**

Std Dev. by Cluster	PTN	PHN	HETN_PER_PERSON	ADDED_VALUE_PER_PERSON	FIN_ASSETS_PER_PERSON	SALES_PER_PERSON	VEXPORT_PER_PERSON
Cluster 1 (X = 0; Y = 0)	€29.98 K	€17.51	0.31 K	€30.73 K	€171.80 K	€248.35 K	€182.19 K
Cluster 2 (X = 0; Y = 2)	€27.49 K	€15.28	0.15 K	€27.68 K	€158.15 K	€99.36 K	€57.87 K
Cluster 3 (X = 2; Y = 0)	€122.61 K	€72.51	0.39 K	€128.39 K	€228.97 K	€273.90 K	€212.47 K
Cluster 4 (X = 2; Y = 2)	€26.40 K	€15.43	0.18 K	€26.38 K	€141.06 K	€114.73 K	€73.13 K
Combined	**€54.91 K**	**€32.10**	**0.26 K**	**€57.28 K**	**€169.67 K**	**€186.81 K**	**€135.35 K**

ROBOT_PROFILE_STR	Always Robotic		Migration to Robotics – Churned		Migration to Robotics – Loyal		Robotics – Churned	
Cluster	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Cluster 1 (X=0; Y=0)	155	83.78%	0	0.00%	0	0.00%	0	0.00%
Cluster 2 (X=0; Y=2)	0	0.00%	150	81.97%	0	0.00%	0	0.00%
Cluster 3 (X=2; Y=0)	30	16.22%	33	18.03%	28	16.47%	13	13.68%
Cluster 4 (X=2; Y=2)	0	0.00%	0	0.00%	142	83.53%	82	86.32%
Combined	**185**	**100.00%**	**183**	**100.00%**	**170**	**100.00%**	**95**	**100.00%**

PURE_SME	Value 0 (No)		Value 1 (Yes)	
Cluster	Frequency	Percentage	Frequency	Percentage
Cluster 1 (X=0; Y=0)	0	0.00%	155	29.30%
Cluster 2 (X=0; Y=2)	0	0.00%	150	28.36%
Cluster 3 (X=2; Y=0)	104	100.00%	0	0.00%
Cluster 4 (X=2; Y=2)	0	0.00%	224	42.34%
Combined	**104**	**100.00%**	**529**	**100.00%**

advantage for both large firms and SMEs among companies with the same level of adoption of robotics. This supports H3 and is fully compatible with the results of Dickson and Hadjimanolis (1998) for the IT revolution or those of Díaz-Chao, Sainz-González, and Torrent-Sellens (2016) concerning innovation. Koch et al. (2021) find robust evidence that ex ante better-performing firms are more likely to adopt robots, conditional on size, and an ex ante cost share of 5–7% points leads to net job creation at a rate of 10%.

4.1.2. Clusters of companies with late adoption of robotics followed by early cessation of robotization

Cluster 4 is made up of large companies (16.52%, 476 records), and Cluster 5 consists of SMEs (17.46%, 503 records). Companies in these clusters (33.98% of the records) adopt robotics late but abandon the process and become non-robotic thereafter.

These clusters differ from the average robotic company in two ways. They are much smaller, and the level of education of their employees is much lower. The average numbers of employees are 739 for large firms and 77.42 for SMEs, which is 27.25% and 52.41% lower than the figure of the average robotic company. The level of education of their employees, albeit higher than that of the average non-robotic company, is still below the average for a robotic one. In the case of the large companies, the average percentage of BA workers is 7.76%, which is 16.20% lower than that of the average robotic company. In the case of SMEs, this effect is even greater, as the average percentage of workers with a BA is 6.45%, which is 21.62% lower than that of the average robotic company.

Regarding SMEs, this effect is reflected in all the business indicators analyzed in the Kohonen SOM, in which they underperform the average robotic company considerably. These companies do not reach the average productivity and performance levels characteristic of a robotic company, and they finally abandon robotization. These findings support H5 on the late adoption of robotics in SMEs. In the case of large companies, this effect is not as evident, as their business indicators are above average, but, for the effective hours worked per employee, they are 1.32% below the average, with 1.74 thousand hours worked per employee. Our results partially confirm those of Acemoglu and Restrepo (2019, 2020) and Doraszelski and Jaumandreu (2018).

This means, as suggested by Frey and Osborne (2017) and Gaimon (1985), that having the right workforce characteristics plays an important role in the successful adoption of robotics and the subsequent expected benefits. This supports H4, especially in the case of SMEs.

4.1.3. Cluster of companies with late adoption of robotics that achieve mixed results

Cluster 3 is a small group of large companies (5.24%, 151 records) that adopt robotics late and obtain mixed results. In this cluster, 68.91% of the companies remain loyal to their transformation to robotics, but 31.09% abandon the attempt and become non-robotic again.

All their business indicators are below the average for a large robotic company. For example, the productivity per employee is €51.34 K, which is 12.60% lower, and the sales per employee is €188.00 K, 15.25% lower than those for an average large robotic company. These findings support H5 regarding the late adoption of robotics in large firms.

In addition to this, the labor conditions of the workers are lower than those in the average large robotic company, with 13.77% being temporary workers, 13.32% above the average, and an average cost per employee of €30.51 K, 13.49% lower. The same applies to the level of education of the employees, with 7.55% of employees with a BA, 18.55% lower than the average, and 6.55% of employees with an MA, 11.25% lower than the average. These results also support H4 for large firms, that is, having the right characteristics of the workforce plays an important role in the successful adoption of robotics.

4.2. Group 2: Cases from companies with unexpectedly early adoption of robotics

This group consists of companies that adopt robotics unexpectedly early. The MLP-ANN predicts that these companies should not have been robotic in the period of observation (900 records belonging to 633 companies).

This group is classified into four clusters based on the same nine business indicators from categories such as productivity and labor, firm performance, and the robotics adoption life cycle (Table 5). These clusters are described in the following sections using the results from the MLP-ANN and the Kohonen SOM model in Tables 4 and 8.

4.2.1. Clusters of premature adopters of robotics that remain loyal to its adoption

Cluster 1 is made up of SMEs (25.67%, 231 records) that adopt robotics earlier than predicted but remain loyal to their transformation. These companies have been unable to maximize the return on their investment in robotization despite remaining committed to it.

The transition to robotics improves their performance in all the business indicators, for example €40.2 K average productivity per employee and €143.28 K average sales per employee, which are, respectively, 36.79% and 45.60% above those of the average non-robotic company. Nevertheless, they have been unable to achieve as great a benefit overall as expected from being a robotic company. For example, the average productivity per employees is 24.62% less and the average sales figure per employee is 31.44% less than those of the average SME robotic company. These findings confirm H5 on the premature adoption of robotics in SMEs. The MLP-ANN considers them not to be fully ready to adopt robotics because their workforce characteristics are superior to the average non-robotic SME but they are not sufficiently advanced to adopt robotics.

These companies have an average of 52.12 employees, being 40.07% larger than the average non-robotic SME but still 67.96% smaller than the average robotic SME. In addition, the percentage of temporary workers is still too high (16.07%), being 16.77% higher than that of the average robotic SME, and the percentages of employees with a BA (7.03%) and an MA (5.8%) are too low, 14.59% and 2.59% below the average for robotic SMEs. Consequently, the cost per employee (€24.90 K) is also 23.79% below the average. These findings complement those reported by Doraszelski and Jaumandreu (2018) and confirm H4 for SMEs.

4.2.2. Clusters of premature adopters of robotics that later cease robotization

Cluster 2 contains SMEs (26.44%, 238 records) that adopt robotics too early and cease to become non-robotic again after the transition. This cluster represents SMEs that are at an earlier stage of the evolution of their workforce compared with Cluster 1. Being at an earlier stage of the evolution of the workforce has a negative impact on all the business indicators compared with Cluster 1.

The average size of the companies in Cluster 2, with 43.24 employees, is 17.04% smaller than that in Cluster 1, 16.20% larger than the average non-robotic SME but also very far from the average robotic SME (73.42% smaller). In Cluster 2, the average percentage of temporary workers (19.41%) is 20.84% higher than that in Cluster 1. The average percentages of employees with a BA (5.56%) and an MA (4.45%) are 20.88% and 24.36% lower than those in Cluster 1. These differences are even clearer when compared with the average robotic SME. The percentage of temporary workers is 41.10% higher and the percentages of employees with a BA and an MA are 32.42% and 26.32% lower, respectively.

The improvement in the performance of the main business indicators due to robotization is much smaller in Cluster 2 than in Cluster 1, and the companies are very far from the average robotic SME company. For example, the €35.55 K average productivity per employee and €120.67

K average sales per employee are, respectively, 33.03% and 42.26% lower than those of the average robotic SME and 11.16% and 15.78% lower than those in Cluster 1, respectively.

Finally, the average costs per employee (€22.57 K) are 9.36% lower than those in Cluster 1 and 30.93% lower than those of the average robotic SME. These results conform to those of Cottelleer and Bendoly (2006) and Gaimon (1985) regarding IT implementation. All these findings justify why companies in Cluster 2 abandon the robotics adoption process and confirm H4 and H5 vis-à-vis the premature adoption of robotics in SMEs.

4.2.3. Clusters of premature adopters of robotics that have mixed results

Clusters 3 and 4 represent companies that adopt robotics too early according to the MLP-ANN and achieve mixed results. Some of them remain loyal to their transformation to robotics, but others become non-robotic thereafter.

Cluster 3 comprises large companies (15.67%, 141 records). In this cluster, 44.23% of the companies abandon robotics and become non-robotic again, while 55.77% of the companies remain loyal to their robotization process. Cluster 4 consists of SMEs (32.22%, 290 records). In this cluster, 36.6% of the companies become non-robotic again and 63.39% of the companies remain loyal to their robotization process.

Companies from Clusters 3 and 4 have that they underperform compared with the average robotic company in common. For both SMEs and large companies, all their business indicators except the hours worked by employees are below the average. For example, the productivity per employee is €52.42 K for large firms (Cluster 3) and €36.89 K for SMEs (Cluster 4), that is, 10.76% and 30.51% lower than that of the average robotic company, respectively; the sales per employee are €187.18 K for large firms (Cluster 3) and €117.09 K for SMEs (Cluster 4), which are 15.63% and 43.97% lower than those of the average robotic company, respectively.

In both large firms (Cluster 3) and SMEs (Cluster 4), the number of employees, their level of education, and the costs per employee are below the average of robotic companies of each kind. Clusters 3 and 4 have 245.36 and 42.59 employees on average, respectively, which are 75.85% and 73.82% lower than those for the average large robotic company and robotic SME, respectively. Clusters 3 and 4 have averages of 5.11% and 5.15% of employees with a BA, being 44.77% and 37.37% lower than those of the average large robotic company and robotic SME, respectively.

Consequently, this is reflected in the costs per employee, which are €30.39 K for large firms in Cluster 3 and €23.81 K for SMEs in Cluster 4; these figures are 13.83% and 27.11% lower than those of the average large robotic company and robotic SME, respectively. These findings confirm H4 and H5 about the premature adoption of robotics in large companies and SMEs.

It is relevant to highlight how the adoption of robotics by SMEs evolves among the different clusters, tracing a journey from Cluster 2, representing the group of companies with the most immature workforce, to Cluster 4 and finally to Cluster 1. The increase in the maturity of the workforce is linked to a decrease in the abandonment of the technology and an improvement in the performance of the main company business indicators, even if they are unable to maximize the adoption of robotization because they undertake the transition without being fully ready.

5. Conclusions

Our goal in this research is to elucidate how the adoption of robotization affects industrial firms, focusing, for the first time, on the dynamics of transformation and complementing the growing literature on the effects of automation technologies (Acemoglu & Restrepo, 2020; Ballestar et al., 2021; Doraszelski & Jaumandreu, 2018; Gregory et al., 2019). Thus, we analyze data on manufacturing firms in Spain within a 26-year period, focusing on one of the most sensitive elements of the transformation process toward robotization, which is the workforce.

First, we develop a predictive ML model using a supervised ANN multilayer perceptron model that can predict whether a company is ready to make the transition to robotics according to its workforce structure, characteristics, and composition (Athey & Imbens, 2019; Ballestar, Doncel, Sainz, & Ortigosa-Blanch, 2019; Ballestar, Grau-Carles, & Sainz, 2019). Hence, this model provides a real-time tool for companies to evaluate the need to make structural changes in their workforce according to their relationship with the adoption of robotics at the right time. It allows them to develop a detailed workforce transformation plan and avoid the risks—technical, economic, and human—associated with premature or late adoption of the technology or no adoption at all.

This model also shows how the transformation to robotics implies higher average costs per employee both in SMEs and in large firms. Nevertheless, these costs result from different changes in the structure of the workforce depending on the size of the company. In SMEs, this cost increase comes from the larger investment in employees with higher levels of education, the improvement in the labor conditions, and the type of professional relationship between the employees and the company. Meanwhile, in large firms, this cost increase is related not to the employees' level of education but only to an improvement in the labor conditions and the type of professional relationship between the employees and the company.

The robotization process implies an increase in the average percentage of employees with a BA in SMEs and large firms. Nevertheless, this effect works differently on the level of education in SMEs compared with large firms. In SMEs, it represents an increase in the employees' level of education, as it is caused by a reduction in the average percentage of non-graduates. In large firms, the situation is different and means a reduction in the level of education as it results from a decrease in the average percentage of employees with an MA. Hence, the changes involved in the robotization process are related to not only the technology but also the workforce. Investment in new technology is associated with higher-paid workers in SMEs and large companies and, in the case of SMEs, more specialized and qualified human capital (Acs et al., 2009; Ballestar et al., 2021; Díaz-Chao et al., 2016; Koch et al., 2021).

In the second phase of this research, we focus on the characterization and segmentation of the companies that are misclassified by the AMM-MLP to understand the reasons for such unexpected behavior regarding the adoption of robotics. We group the misclassified cases, analyzed their main business indicators using an ANN model, specifically the Kohonen SOM, as an unsupervised clustering method, and expand the theories first described by Gaubert and Cottrell (2007) on labor market classification. We obtain relevant results about the relationship between a company's workforce structure and characteristics and its degree of adoption of robotics, which have an impact on the performance of the main business indicators that are affected by these factors.

We find a common characteristic in all the clusters of misclassified cases. Both SMEs and large companies have a smaller number of employees than the average robotic company in each category. Hence, smaller companies in both categories are more prone to making unexpected decisions regarding robotization than companies of an average size. As Pratt (2015) reports, the transition requires changes in the management of organizations, which are easier to implement in large firms with greater access to talent.

This finding is also linked to the evidence that the workforce plays an important role in the process of adopting robotics and the generation of the expected benefits. The unexpected adoption of robotics, too late or too soon, in combination with an inadequately structured workforce, risks the success of the transformation process. Consequently, the company generates lower profits than expected in terms of the main business indicators. In the cases in which the underperformance is greater, there is a high risk that the company will abandon the process of robotization. We have shown that the greater the underperformance of the company, the higher the probability of dropping the process of adopting robots. These findings apply to both SMEs and large firms;

nevertheless, the negative effects are more evident in SMEs than in large firms.

Hence, it is very relevant for companies and decision makers to have predictive tools, such as the ANN-MLP provided in this research, to ascertain whether their workforce is ready for the adoption of robotics and maximize the expected benefits of the transformation process and minimize the dropping-off risk. An additional benefit of this predictive model is that it can be implemented for real-time evaluations, allowing decision makers to act when needed (Ballestar et al., 2021).

Additionally, we find that some companies, large companies and SMEs, are reluctant to adopt robotics. These companies transform their workforce, assuming the additional human resource costs, in a similar way to robotic companies. Nevertheless, they do not proceed further in their transformation process and remain non-robotic. Their aim is to maintain a competitive advantage in terms of productivity and competitiveness among non-robotic companies without assuming the technical costs of the transformation process. This means that the workforce itself represents a competitive advantage among companies with the same level of robotics adoption, specifically non-robotic companies. These findings are valid for both SMEs and large companies.

This research has some limitations. The ESEE database focuses on manufacturing firms, but further research should analyze service industries wherein the workforce structure and characteristics could also be transformed by robotization. We have also seen that our findings are applicable to all the sectors, but an analysis by sector would provide additional details according to the specific activity of companies.

Our research has clear implications for management and will help firms to hire the right human capital for their characteristics, invest in training, decide how they will conduct knowledge management of their human capital, and anticipate the effects of these changes linked to training. Thus, it is essential to identify the complementarities of change in employment and productivity to design the training policies needed to avoid unemployment and focus on the most disadvantaged groups.

CRediT authorship contribution statement

María Teresa Ballestar: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft. **Aida García-Lazaro:** Writing – review & editing, Visualization, Investigation, Supervision. **Jorge Sainz:** Conceptualization, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. **Ismael Sanz:** Writing – review & editing, Visualization, Supervision, Investigation, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

See the Table A1.

Table A1
Activity sectors.

Activity Sector	Description
1	Meat industry
2	Food products and tobacco
3	Drinks
4	Textiles and clothing manufacturing
5	Leather and footwear
6	Wood industry
7	Paper industry

Table A1 (*continued*)

Activity Sector	Description
8	Graphic arts
9	Chemical industry and pharmaceutical products
10	Rubber and plastic products
11	Non-metallic mineral products
12	Ferrous and non-ferrous metals
13	Metal products
14	Agricultural and industrial machines
15	Computer, electronic, and optical products
16	Machinery and electrical equipment
17	Motor vehicles
18	Other transport equipment
19	Furniture industry
20	Other manufacturing industries

References

Aad, G., Abbott, B., Abdallah, J., Abdelalim, A. A., Abdesselam, A., Abdinov, O., Abi, B., Abolins, M., Abramowicz, H., Abreu, H., Acerbi, E., Acharya, B. S., Adams, D. L., Addy, T. N., Adelman, J., Aderholz, M., Adomeit, S., Adragna, P., Adye, T., … Zwalinski, L. (2012). Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s}$ = 7 TeV with atlas. *European Physical Journal C, 72* (1), 1–35. https://doi.org/10.1140/epjc/s10052-011-1844-6

Acemoglu, D., & Restrepo, P. (2019). *Automation and new tasks: How technology displaces and reinstates labor* (NBER Working Paper No. 25684). National Bureau of Economic Research. https://doi.org/10.3386/w25684.

Acemoglu, D., & Restrepo, P. (2020). *Unpacking skill bias: Automation and new tasks* (NBER Working Paper No. 26681). National Bureau of Economic Research. https://doi.org/10.3386/w26681.

Acs, Z. J., Braunerhjelm, P., Audretsch, D. B., & Carlsson, B. (2009). The knowledge spillover theory of entrepreneurship. *Small Business Economics, 32*(1), 15–30. https://doi.org/10.1007/s11187-008-9157-3

Afolabi, M. O., & Olude, O. (2007). Predicting stock prices using a hybrid Kohonen self organizing map (SOM). In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)* (pp. 48).

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., & Li, H. (2019). Missing growth from creative destruction. *American Economic Review, 109*(8), 2795–2822. https://doi.org/10.1257/aer.20171745

Angrist, J. D., & Pischke, J. S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives, 24*(2), 3–30. https://doi.org/10.1257/jep.24.2.3

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics, 11*(1), 685–725. https://doi.org/10.1146/annurev-economics-080217-053433

Autor, D., & Salomons, A. (2017). Does productivity growth threaten employment? In *European Central Bank Annual Conference*, Sintra, Portugal (Vol. 27).

Autor, D., & Salomons, A. (2018). Is automation labor-displacing? Productivity growth, employment, and the labor share. *Brookings Papers on Economic Activity.* https://www.brookings.edu/bpea-articles/is-automation-labor-displacing-productivity-growth-employment-and-the-labor-share/.

Ballestar, M. T., Díaz-Chao, Á., Sainz, J., & Torrent-Sellens, J. (2020). Knowledge, robots and productivity in SMEs: Explaining the second digital wave. *Journal of Business Research, 108*, 119–131. https://doi.org/10.1016/j.jbusres.2019.11.017

Ballestar, M. T., Díaz-Chao, Á., Sainz, J., & Torrent-Sellens, J. (2021). Impact of robotics on manufacturing: A longitudinal machine learning perspective. *Technological Forecasting and Social Change, 162*, Article 120348. https://doi.org/10.1016/j.techfore.2020.120348

Ballestar, M. T., Doncel, L. M., Sainz, J., & Ortigosa-Blanch, A. (2019). A novel machine learning approach for evaluation of public policies: An application in relation to the performance of university researchers. *Technological Forecasting and Social Change, 149*. https://doi.org/10.1016/j.techfore.2019.119756

Ballestar, M. T., Grau-Carles, P., & Sainz, J. (2018). Customer segmentation in e-commerce: Applications to the cashback business model. *Journal of Business Research, 88*, 407–414. https://doi.org/10.1016/j.jbusres.2017.11.047

Ballestar, T., Grau-Carles, P., & Sainz, J. (2019). Predicting customer quality in e-commerce social networks: A machine learning approach. *Review of Managerial Science.* https://doi.org/10.1007/s11846-018-0316-x

Beane, M., & Brynjolfsson, E. (2020). Working with robots in a post-pandemic world. *MIT Sloan Management Review, 62*(1), 1–5. https://mitsmr.com/2Rv0zD0.

Blanas, S., Gancia, G., & (Tim) Lee, S. Y. (2019). Who is afraid of machines? *Economic Policy, 34*(100), 627–690. https://doi.org/10.1093/epolic/eiaa005

Brynjolfsson, E., & Oh, J. (2012). The attention economy: Measuring the value of free digital services on the Internet. In *ICIS 2012 Proceedings.* https://aisel.aisnet.org/icis2012/proceedings/EconomicsValue/9.

Buckler, F. (2001). *NEUSREL-neuer Kausalanalyseansatz auf Basis neuronaler Netze als Instrument der Marktforschung.* Cuvillier.

Chen, Y., Ghosh, A., Mcafee, R. P., & Pennock, D. (2008). Sharing online advertising revenue with consumers. In *International Workshop on Internet and Network Economics* (pp. 556–565).

Choudhary, B., & Bhattacharyya, P. (2002). Text clustering using semantics. In *Proceedings of the 11th International World Wide Web Conference* (pp. 1–4).

Chui, M., Manyika, J., & Miremadi, M. (2016). Where machines could replace humans—And where they can't (yet). *McKinsey Quarterly, 30*(2), 1–9. http://www.oregon4biz.com/assets/e-lib/Workforce/MachReplaceHumans.pdf.

Currie, J., Kleven, H., & Zwiers, E. (2020). *Technology and big data are changing economics: Mining text to track methods* (NBER Working Paper No. 26715). National Bureau of Economic Research. http://www.nber.org/papers/w26715.

Cotteleer, M. J., & Bendoly, E. (2006). Order Lead-Time Improvement following Enterprise Information Technology Implementation: An Empirical Study. *MIS Quarterly. Management Information Systems Research Center, University of Minnesota, 30*(3), 643–660. https://doi.org/10.2307/25148743

Díaz-Chao, Á., Sainz-González, J., & Torrent-Sellens, J. (2016). The competitiveness of small network-firm: A practical tool. *Journal of Business Research, 69*(5). https://doi.org/10.1016/j.jbusres.2015.10.053

Dickson, K. E., & Hadjimanolis, A. (1998). Innovation and networking amongst small manufacturing firms in Cyprus. *International Journal of Entrepreneurial Behaviour & Research, 4*(1), 5–17. https://doi.org/10.1108/13552559810203939

Doraszelski, U., & Jaumandreu, J. (2018). Measuring the bias of technological change. *Journal of Political Economy.* https://doi.org/10.1086/697204

Dželihodžić, A., & Jonko, D. (2016). Comparison of ensemble classification techniques and single classifiers performance for customer credit assessment. *Modeling of Artificial Intelligence, 3*, 140–150.

Eisfeldt, A. L., Falato, A., & Xiaolan, M. Z. (2021). *Human capitalists* (NBER Working Paper No. w28815). National Bureau of Economic Research.

Faber, M. (2020). Robots and reshoring: Evidence from Mexican local labor markets. *Journal of International Economics, 127*, 103384. https://www.sciencedirect.com/science/article/pii/S0022199620300994%3Fcasa_token%3DhvsWYNPXpFEAAAAA:n0dGFDsOnG_7K_R8JUll-0qHmWnO3jNOQv8VLJ1Tplhs0Va_xes6hyVyR8QQ9ltRCba0F8DS0A&hl=es&sa=T&oi=gsb-gga&ct=res&cd=0&d=13376749078247885573&ei=9cPwX_OgB9LKsQKt2aXI.

Flexer, A. (1996). Statistical evaluation of neural network experiments: Minimum requirements and current practice. *Cybernetics and Systems Research*, 1005–1008.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Gaimon, C. (1985). The optimal acquisition of automation to enhance the productivity of labor. *Management Science, 31*(9), 1175–1190. https://doi.org/10.1287/mnsc.31.9.1175

Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. *International Journal of Emerging Technology and Advanced Engineering, 2*(4), 42–47.

Garbe, J. N., & Richter, N. F. (2009). Causal analysis of the internationalization and performance relationship based on neural networks – Advocating the transnational structure. *Journal of International Management, 15*(4), 413–431.

Gaubert, P., & Cottrell, M. (2007). A dynamic analysis of segmented labor market. *ArXiv Preprint.* Math/0703011.

Gómez, A. R. (2019). Digital fame and fortune in the age of social media: A classification of social media influencers. *ADResearch, 19*, 8–29.

Goolsbee, A. D., & Klenow, P. J. (2018). Internet rising, prices falling: Measuring inflation in a world of e-commerce. *AEA Papers and Proceedings, 108*, 488–492. https://doi.org/10.1257/pandp.20181038

Goos, M., Manning, A., & Salomons, A. (2009). Job polarization in Europe. *American Economic Review, 99*(2), 58–63. https://doi.org/10.1257/aer.99.2.58

Goswami, B. N., Wheeler, M. C., Gottschalck, J. O. N. C., & Waliser, D. E. (2011). Intraseasonal variability and forecasting: A review of recent research. In *The global monsoon system: Research and forecast* (pp. 389–407). World Scientific.

Graetz, G., & Michaels, G. (2018). Robots at work. *Review of Economics and Statistics, 100* (5), 753–768. https://doi.org/10.1162/rest_a_00754

Gregory, T., Salomons, A., & Zierahn, U. (2019). *Racing with or against the machine? Evidence from Europe* (Discussion Paper No. 12063). IZA Institute of Labor Economics.

Guo, H., & Viktor, H. L. (2004). Learning from imbalanced data sets with boosting and data generation: The databoost-IM approach. *ACM Sigkdd Explorations Newsletter, 6* (1), 30–39.

Ha, S. H. (2007). Applying knowledge engineering techniques to customer analysis in the service industry. *Advanced Engineering Informatics, 21*(3), 293–301.

Hosmer, Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression.* John Wiley & Sons.

Hu, X., & Weng, Q. (2009). Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. *Remote Sensing of Environment, 113*(10), 2089–2102.

Jensen, H. L. (1992). Using neural networks for credit scoring. *Managerial Finance., 18* (6), 15–26. https://doi.org/10.1108/eb013696

Kavzoglu, T., & Mather, P. M. (2003). The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing, 24*(23), 4907–4938. https://doi.org/10.1080/0143116031000114851

Koch, M., Manuylov, I., & Smolka, M. (2021). Robots and firms. *Economic Journal, 131* (638), 2553–2584. https://doi.org/10.1093/EJ/UEAB009

Kohonen, T. (1995). *Self-organizing maps.* Springer-Verlag.

Kraus, S., Palmer, C., Kailer, N., Kallinger, F. L., & Spitzer, J. (2019). Digital entrepreneurship: A research agenda on new business models for the twenty-first century. *International Journal of Entrepreneurial Behaviour and Research, 25*(2), 353–375. https://doi.org/10.1108/IJEBR-06-2018-0425

Kugler, A. D., Kugler, M., Ripani, L., & Rodrigo, R. (2020). *U.S. robots and their impacts in the tropics: Evidence from Colombian labor market* (NBER Working Paper No. 28034). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w28034/w28034.pdf.

Li, Z., & Eastman, J. R. (2006). Commitment and typicality measurements for fuzzy ARTMAP neural network. *Geoinformatics 2006: Geospatial Information. Science, 6420,* 64201I.

Maravall, A., & del Río, A. (2007). Temporal aggregation, systematic sampling, and the Hodrick-Prescott filter. *Computational Statistics and Data Analysis, 52*(2), 975–998. https://doi.org/10.1016/j.csda.2007.08.001

Muro, M., Maxim, R., Whiton, J., & Hathaway, I. (2019). *Automation and artificial intelligence: How machines are affecting people and places.* Brookings Institution.

Paliwal, M., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications, 36*(1), 2–17.

Pratt, G. A. (2015). Is a Cambrian explosion coming for robotics? *Journal of Economic Perspectives, 29*(3), 51–60.

Ribeiro, D., Martinez-Climent, C., & Tur-Porcar, A. M. (2018). Innovation, knowledge, judgment and decision-making as virtuous cycles: Editorial. *International Journal of Entrepreneurial Behaviour and Research, 24*(6), 1063–1064. https://doi.org/10.1108/IJEBR-10-2018-531

Richardson, A. J., Risien, C., & Shillington, F. A. (2003). Using self-organizing maps to identify patterns in satellite imagery. *Progress in Oceanography, 59*(2–3), 223–239.

Ritter, H. (1999). *Self-organizing maps on non-Euclidean spaces. In Kohonen maps* (pp. 97–109). Elsevier.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*(C), 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Seamans, R., & Raj, M. (2018). AI, labor, productivity and the need for firm-level data. In *The economics of artificial intelligence: An agenda* (pp. 553–565). University of Chicago Press.

Sun, Y., Wong, A. K. C., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence, 23*(4), 687–719.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature, 49*(2), 326–365. https://doi.org/10.1257/jel.49.2.326

Tian, J., Azarian, M. H., & Pecht, M. (2014). Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm. In *Proceedings of the European Conference of the Prognostics and Health Management Society* (pp. 1–9).

Torrent-Sellens, J. (2018). *Las empresas industriales en 2014 y 2015. Encuesta sobre Estrategias Empresariales (ESEE).* Fundación SEPI. https://www.researchgate.net/publication/323150903_Las_empresas_industriales_en_2014_y_2015_Encuesta_sobre_Estrategias_Empresariales_ESEE.

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology, 49*(11), 1225–1231.

Wagstaff, K. (2004). Clustering with missing values: No imputation required. In *Classification, clustering, and data mining applications* (pp. 649–658). Springer.

Williams, R. N., de Souza Jr, P. A., & Jones, E. M. (2014). Analysing coastal ocean model outputs using competitive-learning pattern recognition techniques. *Environmental Modelling & Software, 57*, 165–176.

Yu, H., Su, T., & Zeng, X. (2014). A three-way decisions clustering algorithm for incomplete data. *International Conference on Rough Sets and Knowledge Technology,* 765–776.

María Teresa Ballestar de las Heras is researcher at ESIC Business & Marketing School and faculty at Universidad Rey Juan Carlos.

She has developed her professional career in different sectors such as IT, consultancy, banking, startups and pharma. She spent more than seven years as IT consultant and later more than eleven years in areas related with data analysis in banking, such as risk, finance, marketing and business intelligence. She has been holding managing positions, driving digitalization, innovation and data analytics in startups and pharmaceutical companies.

She holds two B.Sc. in Statistics and M.Sc.Market research, a M.A. in Knowledge and Information Society and a Ph.D in Applied Economics. Her main field of research is the analysis of customer behavior in e-commerce by using Big Data methods and techniques where I she has published several papers in top tier journals within the last years.

Aida Garcia-Lazaro Economist and researcher at the Institute for Policy Research, University of Bath. She attained the PhD Economics from the University of York in 2019 and the MSc in Economics from the National Autonomous University of Mexico (UNAM).

Her research agenda includes topics related to the labour market, technological change, macroprudential policies, Brexit and migration. Aida's expertise is on Dynamic Stochastic General Equilibrium (DSGE) models and econometrics. More recently, she has engaged on machine learning techniques.

Aida has experience as a Lecturer in the MRes program of the University of Bath. She was a Graduate Teaching scholar at the University of York from 2014 to 2019 and a Lecturer at the Department of Economics in the UNAM from 2009 to 2014. In addition, she possesses experience working as a consultant in projects funded by the Mexican government and the United Nations from 2011 to 2014.

Jorge Sainz is Professor of Applied Economics at Rey Juan Carlos University and Fellow at the Institute of Policy Research at the University of Bath. He has extensive experience in private public management, holding positions as chief economist at CIMD, Director of E-Commerce at Yahoo! Spain and Secretary of Universities at the Spanish Government, etc.

He holds a B.Sc. in Economics from the Complutense University of Madrid, a LLB (Hons) from UNED, and MBA from the Simon School of the University of Rochester and a Ph.D. in Economics from URJC. His research is linked to his management experience and has focused on financial markets and technological innovation in both private enterprise and education, having published in journals such as Journal of Business Research, Kyklos, International Review of Economics and Finance, etc.

Ismael Sanz PhD in Applied Economics from the Faculty of Economic and Business Sciences of the Complutense University of Madrid (UCM). 2007 Extraordinary Thesis Award from the Faculty of Economic and Business Sciences (UCM).

Professor of the Department of Applied Economics I of the Rey Juan Carlos University.

General Director of Innovation and Scholarships of the Ministry of Education of the Community of Madrid (July 2015- August 2019). Director of the National Institute for Educational Assessment (INEE) of the Central Government Ministry of Education (2012-2015). President of the OECD PISA Strategic Development Group (November 2014-July 2015).

He has taken extended study courses and research stays at the University of California Santa Bárbara, Victoria University of Wellington (New Zealand), New Zealand Department of the Treasury, University of Otago (Dunedin, New Zealand), Australian National University, Kennedy School of Government (Harvard) or University of Nottingham.

He is the author of articles in research journals such as Economic Journal, Oxford Bulletin of Economics and Statistics, Scandinavian Journal of Economics, Canadian Journal of Economics, European Journal of Political Economy, Public Choice or IZA Journal of Labor Studies.