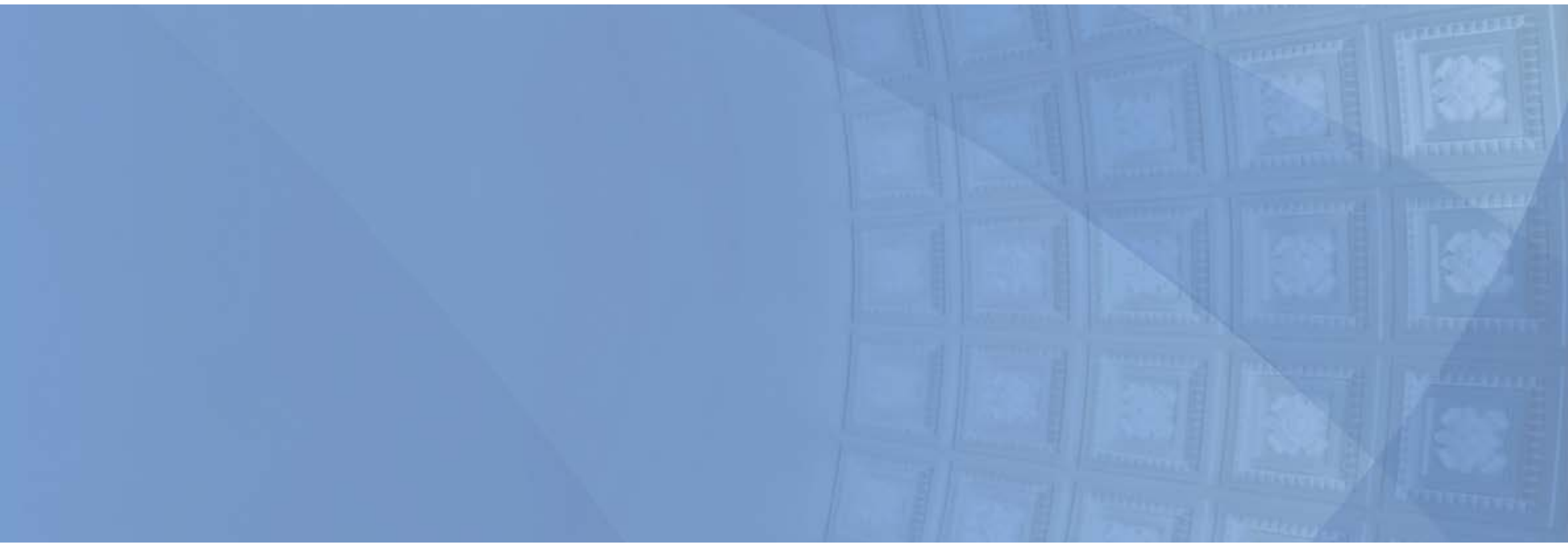
30 YEARS OF THE CROATIAN NATIONAL BANK

Working Papers W-63

Assessment of Readiness of Croatian Companies to Introduce I4.0 Technologies

Rajka Hrbić, Tomislav Grebenar

Zagreb, March 2021



HRVATSKA NARODNA BANKA · CROATIAN NATIONAL BANK

Procjena spremnosti hrvatskih poduzeća na uvođenje tehnologija I4.0

Sažetak

Tema je rada procjena mogućnosti i spremnosti hrvatskih poduzeća za jačanje tehnološko-inovativnog potencijala te analiza prednosti, ograničenja i rizika koje donosi značajan tehnološki skok. Analizirano je 7147 hrvatskih poslovnih subjekata iz različitih djelatnosti. Polazna je točka istraživanja prepoznavanje poduzeća koja potencijalno rabe tehnologiju I4.0, na temelju sličnosti njihovih pokazatelja s pokazateljima 58 poduzeća iz uzorka nedvojbeno identificiranih korisnika i proizvođača tehnologije I4.0. U tu je svrhu razvijen i upotrijebljen model strojnog učenja s pomoću algoritma *eXtreme Gradient Boosting*, što do sada nije bilo primijenjeno u sličnim istraživanjima.

Istraživanjem je pokazano da su glavni razlikovni elementi između I4.0 i tradicionalnih poduzeća najizraženiji kao osjetno bolje poslovne performanse u pokazateljima investiranja, troškovne efikasnosti, tehničke opremljenosti i tržišne konkurentnosti. Rizičnost I4.0 poduzeća značajno je niža od rizičnosti tradicionalnih poduzeća.

U istraživanju je identificirano 141 poduzeće (1,97% analiziranih subjekata) s potencijalom za I4.0, što čini oko 27% aktive analiziranog uzorka te oko 26% poslovnih prihoda.

Ključne riječi: industrija 4.0, *eXtreme Gradient Boosting* (*XGBoost*), umjetna inteligencija, robotika, visokotehnološka poduzeća, strojno učenje, utjecaj I4.0 na poslovne rezultate

JEL: C45, D22, D24, O14, O32, O33

Abstract

The main topic of this paper is to estimate the possibility and inclination of Croatian companies towards technology and inovation as well as to analize advantages, limitations and risks involved with this significant technological leap. In this paper, we analized 7.147 of Croatian business entities operating in different industries. Starting point in this research is to identify other subjects which could be users of I4.0 or its elements, based on the simmilarity of indicators with indicators of a sample of 58 identified I4.0 companies. We developed machine learning model by using eXtreme Gradient Boosting algoritm (XGBoost) for this purpose, an approach which has not been used in any similar reserches.

This research shows that the main difference between I4.0 and traditional industry is mostly observable in significantly better businesess performance of investment indicators, cost efficiency, technical equipment and market competitivness. Riskiness of I4.0 companies is significantly lower than the riskiness of traditional ones.

We identified 141 companies (1,97% of total analized sample) as potential users of I4.0, which make around 27% of total assets of the analised sample and around 26% of revenues.

Keywords: Industry 4.0, eXtreme Gradient Boosting (XGBoost), artificial intelligence, robotics, high-tech companies, machine learning, impacts of I4.0 on bussines results

JEL: C45, D22, D24, O14, O32, O33

Contents

1 Introduction

This research estimates Croatian companies' readiness to strengthen their technological and innovation potential as well as the advantages, limitations and impact on company riskiness involved in the fourth industrial revolution. The paper analyses the key business indicators and the risk characteristic to I4.0[1] companies in Croatia and compares them against "traditional" companies operating in the same or similar industries.

I4.0 affects the development of companies, the financial sector and thus the economy as a whole. Investments in new technologies have a positive impact on GDP growth through increased investment and productivity (competitiveness). Investing in technology requires substantial financial resources, which leads to an increase in demand for loans. The core of the fourth industrial revolution is artificial intelligence, i.e. the application of machine learning, and especially the so-called deep learning: algorithms for system state identification and autonomous decision making with the aim of process optimisation. These are sophisticated devices that use artificial intelligence and technologies that shorten the duration of research and development projects in design (CAD), prototype development, simulations and process control in production or communication. The technology provided by I4.0 is one of the greatest opportunities for economic development today. The interest of this paper lies in the fact that it applies the deep learning model used by advanced I4.0 technology systems – deep machine learning – on a sample of registered users or manufacturers of I4.0 technology.

An analysis of previous research and a review of the literature revealed that there is a need for such research in order to use sophisticated machine learning techniques to estimate the readiness of companies and their potential for the introduction of I4.0 in the Republic of Croatia. The structure of the paper is such that the first, introductory chapter, is followed by the second chapter, which explains the concept and role of the fourth industrial revolution and provides an overview of theory and previous research. The third chapter deals with identifying the potential for I4.0 using machine learning. The analysis of results constitutes the fourth chapter. The concluding chapter elaborates on the implications for economics and economic policy, i.e. it establishes the interconnections between the main results of the paper (detected potentials) and policies that should be able to support the development of companies in terms of introducing new technological solutions of the fourth industrial revolution. The annexes contain more detailed information on the variables used and a brief presentation of the basics of machine learning.

[1] Artificial intelligence, robotics and other technologies with a high degree of autonomy

2 The impact of Industry 4.0 on company operations

2.1 The concept and application of Industry 4.0

Industry 4.0, or I4.0 and I4, is based on automated technology networked via sensors and communication elements (Blunck and Werthmann, 2017), which thus connects the real and virtual world in the form of a cyber-physical system, such as e.g. autonomous robots. Unlike traditional production systems with centralised control, which considers each individual machine an independent unit, the so-called 4.0 factory connects machines into a type of community that is interacting and collaborating autonomously and "intelligently". The use of advanced prediction tools enables continuous processing of big data for the purpose of decision making that is based on all available information at all times, which is the basis for the development of artificial intelligence (AI).

There are different definitions of Industry 4.0, but what they have in common is that they include technologies that lead to the automation of certain processes in production and/or provision of services. These are 3D printing, artificial intelligence, augmented reality, robots, big data, Blockchain, cloud technology, "cobotic" systems involving human-robot cooperation – collaborative systems, cybersecurity, drones, GPS (Global Positioning System), the Industrial Internet of Things, mobile technology, nanotechnology, RFID (technology that uses wireless communication and automatically tracks and identifies specific objects), sensors and simulations (Dalenogare et al., 2018; Lu, 2017; Wan et al., 2015; Posada et al., 2015, cited in Bai et al., 2020).

In this paper, companies that use certain elements of the fourth industrial revolution or plan to modify their business operations in line with the concept of Industry 4.0 are identified based on the following technologies (BCG – Boston Consulting Group):

1. big data and analytics
2. autonomous robots
3. simulations
4. horizontal and vertical system integration
5. the Industrial Internet of Things
6. cybersecurity
7. the cloud
8. 3D printing
9. augmented reality.

A holistic approach to the technologies it uses is important for determining whether a company is an I4.0 company. Certain companies use some of these technologies, but this does not mean that they are fully considered part of Industry 4.0. Depending on the degree of application of these technologies, we can conclude whether a certain company is on track to realise the I4.0 concept. Differences among companies can be significant, ranging from a fully automated company that uses robots to manufacture robots (Japan as a synonym for robots and robotics) to companies that are gradually embracing certain segments of the new industrial revolution.

The technology characteristic of each of the previous three industrial revolutions (steam engines, electricity, information technology) was an extraordinary discovery and advancement, which is also the case in the current revolution. These changes will affect a number of areas such as business administration, finance, the health sector, energy, transport, industry, service activities, intellectual services and many other areas such as genetics and biotechnology. The research conducted by Frey and Osborne, 2013, assesses the susceptibility of current jobs to technological development. According to that estimate, 47% of total US employment is in the high risk category. The model applied in the aforementioned paper predicts a different trend of polarisation of the labour market from the existing one. As technology advances, according to the aforementioned research on the future of employment, workers with a lower level of skills are reallocated to tasks that require creative and social intelligence. However, the changes will also affect highly educated professionals (the example of IBM Watson), which will influence the field of law or healthcare (diagnostics). Google has utilised artificial intelligence as Google Duplex, a virtual assistant that can schedule meetings or appointments by communicating with real people, even those who do not know the language well. Due to the topic of this paper, it is necessary to emphasise the area of creative artificial intelligence that can provide new creative technological solutions by processing big data.[2]

Economic benefits cited as a kind of "drivers" of the fourth industrial revolution (McKinsey, cited in Blunck and Werthmann, 2017) include using resources and optimising business processes (for example, decreasing material costs due to real-time monitoring of the production process, reduction of waiting time between different production steps in manufacturing and acceleration of research and development processes result in increased productivity). Optimal utilisation of assets, management of inventories, increased productivity, improvement of the quality of products and services, reducing the time to market, reducing the costs of aftersales and customer support, service and product maintenance using virtual assistants and the like are just some of the benefits of Industry 4.0.

The industrial revolution of the fourth generation is mostly characterised by technologies listed below.

Artificial intelligence is mostly used for interaction with the environment, image recognition (static or in motion), human speech and the state of the environment

[2] For example, in the aerospace industry, when designing profiles that are extremely strong and light (the example of the Airbus A-320 concept, which reduces the weight of certain components by up to 45% compared to traditional models, which in turn significantly reduces fuel consumption and CO2 and other GHG emissions, while in combination with the use of 3D printing it also reduces the consumption of raw materials up to 95%).

(temperature, humidity, position, speed, direction of movement, etc.) and processing of collected data in real time with the aim of autonomous and experientially optimised management of a process. There is no generally accepted definition of artificial intelligence. The widest application of artificial intelligence is in robotics, which is used mostly in production processes, transport, design, engineering, finance, IT, diagnostics, and increasingly in households and the entertainment industry.

Big Data are becoming the standard in real-time support in decision making. Data are collected from multiple sources, such as production equipment and systems, and company and customer management systems. In order for the use of big data to be meaningful in terms of utilisation, it is necessary to consolidate and evaluate such data in an intelligent way (Sauter et al., 2015, p. 5, cited in Blunck and Werthmann, 2017).

Robots interact with each other and operate "in collaboration" with people and learn from them. Costs will be lower and opportunities more plentiful than in today's production. Robotics is one of the foundations of Industry 4.0, and robots and humans are increasingly becoming equal in business processes.

Simulation is mostly used to transpose the physical world to a virtual model for the purpose of reducing costs and increasing quality. They allow operators to test and optimise machine settings for the next product before physical production.

Horizontal and vertical system integration allows for better cohesion between departments and functions, as comprehensive data networks develop automated value chains.

The Internet of Things (IoT) in Industry 4.0 means that computers will be integrated into devices in order to enable them to communicate with each other. Blunck and Werthmann (2017) describe it as an "ecosystem" of technologies monitoring the status of physical objects. At the same time, they capture meaningful data and communicate that information through networks to software applications. Each definition of the Internet of Things includes: smart objects, machine to machine communication (M2M) and radio frequency technologies (Thrasher, 2014, cited in Blunck and Werthmann, 2017).

Cybersecurity is a necessity arising from increased connectivity and the use of standard communication protocols. Secure, reliable communications as well as identity and access management of machines are essential. According to the latest available European Investment Bank Activity Report (EIB, 2018), the topic of cybersecurity was highlighted. The report points out that over the past period, cyber-attacks have threatened thousands of companies and the data of billions of people.

Cloud technology enables connectivity in production and requires greater data exchange. The performance of cloud technologies will improve in terms of response time, resulting in the provision of more data services.

3D printing is increasingly used due to its construction advantages for the production of prototypes and individual components or for the production of small batches of customised products. It is an extraordinary revolution, one that is akin to that of Gutenberg's printing press 570 years ago. The possibilities of 3D printing are impressive, from utilising it for NASA technology in the aerospace industry to manufacturing of organs (e.g. ear, kidney, etc.) using patient's cells. 3D printing of food is another economically interesting area.

Augmented reality supports a variety of services and provides real-time information. This technology can result in better decision making and/or performance.

2.2 Previous research and overview of literature

The study conducted by PwC (2014) not only demonstrates how industrial companies can create new opportunities for economic development using I4, but also discusses possible challenges. The study was conducted in the form of a survey of five core industrial sectors[3] using a database of 235 German industrial companies. The authors estimate that the share of investments in I4.0 technology will account for more than 50% of planned capital investments in the five-year period. Likewise, German industry will invest approximately 40 billion euro in I4 every year by 2020. The companies surveyed expect an 18% increase in productivity over the next five years. The Internet of Things or Services will contribute to an increase in revenues of 2% to 3% per year, which will represent an increase of 30 billion euro at the level of German industry.

The Cerved SMEs Report (2017) analyses the Italian government's plan for I4.0 to stimulate innovation, investment and research and development. A method of dividing companies into clusters based on the inclination of companies towards innovation and investments was applied, and companies that are inclined towards innovation generating higher revenue growth and better profit margins, while at the same time facing higher bankruptcy rates and higher labour turnover.

In addition to the aforementioned research on the future of employment (Frey and Osborne, 2013), it is also important to mention the Acemoglu and Restrepo study (2017), which examines the impact of robots and computer technology on the future of the labour market based on data on the increase in the use of robots between 1990 and 2007 in the US. By using a model in which robots compete against people in performing various jobs and tasks, they demonstrate that the introduction of robots can reduce employment and wages depending on the industry. Therefore, they conclude that automation, robotisation and artificial intelligence have a strong adverse impact on the labour market. According to their estimates, the introduction of an additional robot per

[3] C – manufacturing, D – electricity, gas, steam and air conditioning supply, H – transportation and storage, J – information and communication, and M – professional, scientific and technical activities are the activities in which advanced I4.0 technologies are introduced the most.

one thousand employees reduces the employment rate by 0.18 – 0.34 percentage points, and wages by 0.25 – 0.5 percent.

Veža et al. (2018) examine the position of Croatian manufacturing companies in relation to Industry 4.0, i.e. “whether a company can survive in the market without taking strategic guidelines towards Industry 4.0 by 2020”. According to the research, the industrial maturity of Croatian companies is at a very low level (only slightly higher than the level of the second industrial revolution). A fundamental weakness was also expressed, namely insufficient monitoring of developments in technology due to the low level of employee training, established on the sample of surveyed companies (rarely more than five days per year). The results obtained are in line with the conclusion presented in the study by Roland Berger (cited in Veža et al., 2018), according to which Croatia has a very low Industry 4.0 readiness index[4] and belongs to the group of hesitators, along with Bulgaria, Poland, Portugal, Estonia, Spain and Italy. Such a conclusion stems from the relationship between the share of manufacturing in GDP and the readiness of European countries to introduce Industry 4.0, of which only Bulgaria had a weaker result than Croatia.

The McKinsey study (Novak et al., 2018) mentions digitisation as a new impetus in the development of Central and Eastern European (CEE) countries, which they call digital challengers. CEE is one of the most attractive regions for investments at the global level, providing an opportunity that Croatia must seize in order to reduce the gap with regard to developed Western European countries. The attractiveness of these countries stems from high mathematical literacy (which is almost identical to that of front-runner countries), a large STEM[5] talent pool and high-quality digital infrastructure with excellent 4G network coverage. They call the CEE region a “vibrant emerging digital ecosystem” and estimate that digitisation could be a driver for the region, which could contribute 200 billion euro in additional GDP by 2025 (8.3 billion euro, or approximately 2,000 euro of GDP per capita for Croatia (McKinsey, 2018)).

3 Identifying potentials for I4.0 using machine learning

This research estimates Croatian companies’ readiness to strengthen their technological and innovation potential as well as the advantages, limitations and risks involved in the fourth industrial revolution. The analysis is based on the estimation of the potential for the introduction of I4.0 technologies in a wider set of Croatian companies. The potential for I4.0 is defined as the similarity of a company to companies that are autonomously

[4] Measured by the degree of production complexity of the business process, automation, innovation and knowledge (readiness of the labour force)

[5] Science, technology, engineering and mathematics

identified as users of I4.0 technology. This section describes the set of data used and the method that was utilised.

The first challenge is to identify companies whose business operations or product (service) is related to the fourth industrial revolution (I4.0). There is no single systematic record of users of new generation high technology in Croatia. Companies that consistently use I4.0 technology were identified by individual verification of each entity from the list of companies that are users of high technology from various sources, according to the criteria described below (Chapter 3.2 and Annex A, Table A.1).

The potential, i.e. the readiness of a company to introduce I4.0 was estimated probabilistically by applying the classification algorithm of supervised machine learning with a binomial dependent variable. Based on the model estimate of probability, the other observed companies are classified in group I4.0 (probability > 50%), which means that they are very similar to companies that are unequivocally identified as users or producers of some of the listed technologies of the fourth industrial revolution (Annex A, Table A.1) or in the group of traditional companies (probability <= 50%) if this is not the case.

New technologies are often associated with the perception of increased risk, which makes it difficult or at least further increases the cost of funding research and development projects. This paper demonstrates that there is no objective basis for the perception of higher riskiness of I4.0 companies, whose developmental path is based on high technology. On the contrary, investing in development and new technologies increases their competitiveness in increasingly demanding markets that set quality and reliability as the new standard ahead of price.

3.1 Data

Analysis of the potentials for the introduction of I4.0 takes into account companies that operate in five industries, including C – manufacturing, D – electricity, gas, steam and air conditioning supply, H – transportation and storage, J – information and communication, and M – professional, scientific and technical activities.

The sample of companies consists of entities whose annual financial statements[6] had been made public for 2017 and 2012, to which, depending on availability, financial indicators and certain items from the balance sheet and profit and loss statement for 2008 have been added (non-probabilistic sample). The non-probabilistic sample makes up for approximately 35% of the total number of companies operating in the analysed industries, 88% of assets, 85% of operating income and 78% of the total number of employees in these branches of business activity.

[6] Source: GFI-POD (annual financial statements of companies) database of the Financial Agency

These criteria are met by a total of 7147 companies (Table 1), of which 58 were identified by expert assessment[7] as actively using or offering technology and services according to the criteria for I4.0. The indicators and items from the balance sheet and profit and loss statement that have been tested for inclusion in the model for estimating the potential for the introduction of I4.0 are described in the Annex in Table B.1 and Table B.2.

Table 1 Number of analysed entities according to different samples

Type of sample	Number of entities analysed	Share in the non-probabilistic sample
Training sample	512	7.16%
Test sample	501	7.00%
Non-probabilistic sample	7147	100%

Table 2 Number of analysed entities according to different activities

Industry	Number of entities analysed
C – manufacturing	2803
D – electricity, gas, steam and air conditioning supply	103
H – transportation and storage	747
J – information and communication	989
M – professional, scientific and technical activities	2505
TOTAL	7147

It can be seen that the largest number of companies operates in industry C – manufacturing and industry, and M – professional, scientific and technical activities (Table 2).

Table 3 Non-probabilistic sample according to company size

Company size	Number of entities analysed
1 (small and micro enterprises)	6457
2 (medium-sized)	528
3 (large-scale)	162
TOTAL	7147

[7] Expert assessment of the use of I4.0 technologies was made on the basis of available data from various sources (see References), with additional verification on the websites of the analysed entities, thus identifying 110 companies, of which 58 (Annex A, Table A.1) were retained in the final non-probabilistic data set of "7147".

3.2 Hypothesis and assumptions

The initial hypothesis is that companies whose financial performance indicators are similar to those of identified I4.0 users are at a similar level of technological equipment and organisational structure, which enables the identification of potential I4.0 users in a wider set of companies.

Since there is currently no systematic record of users of I4.0 technology (such as Cerved in Italy) in Croatia, the collected data on high-tech companies and users/producers of I4.0 technologies were verified for each entity individually. The criterion for designating a company an unequivocal user of I4.0 is to find evidence that the company uses or produces/provides products or services based on at least one technology of the fourth industrial revolution, namely: big data and analytics, robots, simulation, horizontal and vertical system integration, the Internet of Things, cybersecurity, cloud technologies, 3D printing or augmented reality. The list of these companies is provided in Annex A, Table A.1. Identifying other potential users of I4.0 technology relies on similarities in the structure of financial statements and indicators of such companies in relation to identified I4.0 companies, especially the share of intangible assets in fixed assets and investments in research and development, as applied, for example, in the Cerved (2017) research. The difference in relation to the mentioned research is that these indicators were not selected exclusively by expert assessment, but, among other indicators, were confirmed as statistically significant so that in the final classification model their branches have the highest information gain in classifying companies as I4.0 companies.

For this purpose, a binomial logistic (logit) classification model calculated using the Extreme Gradient Boosting (XGB) technique of deep machine learning was used through the application of the supervised learning method. XGB has proven to be a superior model in binomial logistic classification also in the area of risk assessment (Petropoulos et al., 2018), and among other deep learning algorithms in relation to logistic regression. Their results were tested empirically, by comparative evaluation using logistic regression, which also resulted in somewhat weaker discriminant properties of the model compared to XGB.

The XGB method starts from the basic linear model (Chen et al., 2016):

$$\hat{y}_i = \sum_j w_j x_{ij} \tag{1}$$

that is, its logistical transformation given by the expression:

$$\Pr(Y = 1 \mid X) = \frac{1}{1 + e^{-\hat{y}_i}} \tag{2}$$

the parameters of which

$$\theta = \{w_j \mid j = 1, \ldots, d\} \tag{3}$$

are optimised so as to minimise the error on the training sample, but also on other data that are "unseen" by the model:

$$Obj(\theta) = L(\theta) + \Omega(\theta) \tag{4}$$

In the objective function, $Obj(\theta)$ $L(\theta)$ represents the function of minimising the error on the training data, while $\Omega(\theta)$ represents regularisation, most often using the L_2 Euclidean norm in order to "smooth" the regression and adjust it to "unseen" data. The applied form of the objective function is a binomial logistic function (objective = "binary:logistic").

For K of decision trees, the model takes the form of

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \tag{5}$$

and similar for any t-th tree

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) \tag{6}$$

The loss function for binomial logistic classification takes the form of

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log(p_i) + (1 - y_i)\log(1 - p_i)\right) \tag{7}$$

where y_i is the target value, and p_i the predicted value. The regularisation function for T number of leaves in a tree is defined by the form

$$\Omega = \gamma T + \frac{\lambda}{2}\sum_{j=1}^{T} w_j^2 \tag{8}$$

where γ is the minimum degree of loss reduction the increase of which contributes to the conservatism of the model. XGB uses a gradient descent algorithm to minimise the objective function and tree branching, using the predicted value from the previous step, which is simplified into:

$$y^{(t)} = y^{(t-1)} + \eta f_t(x_i) \tag{9}$$

where η is the learning rate, which reduces the impact of each new tree in the iteration, and thus the overfitting of the model.

Since very few companies in the entire sample were identified as I4.0 companies, model training and testing samples were made using random sampling from the non-probabilistic sample so that I4.0 companies were divided in a ratio of 60: 40 in favour of the training sample, while the remaining candidate companies were selected at random (Figure 1).

Figure 1 Distribution of the training sample

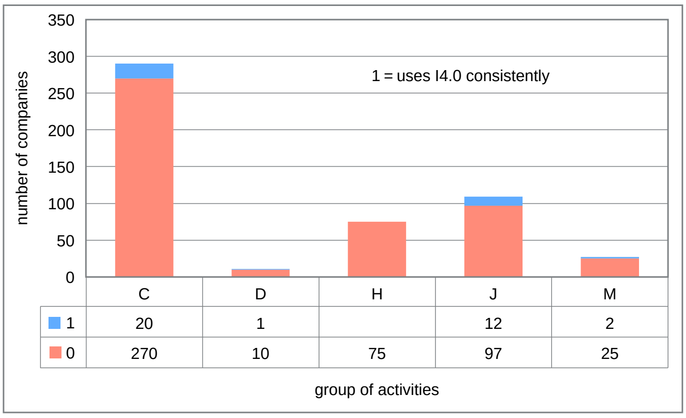


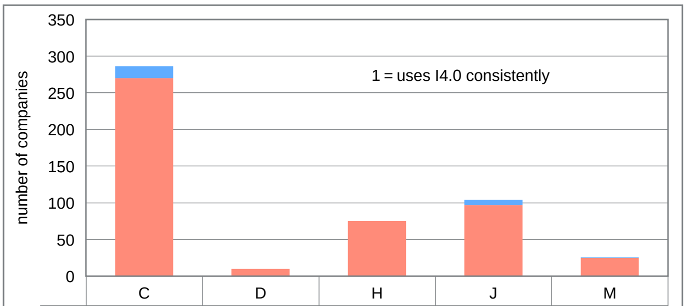
Table 4 shows the companies identified as I4.0. The group of small enterprises has largest share among them in the training sample (44.7%), followed by large-scale enterprises (31.6%) and medium-sized enterprises (23.7%).

Table 4 Structure of the I4.0 training sample according to company size

Company size	Number of I4.0 companies	Share, %
1 – small enterprises	17	44.7%
2 – medium-sized enterprises	9	23.7%
3 – large-scale enterprises	12	31.6%
Total training sample	38	100.0%

The test sample was selected so that its structure by activities corresponds to the training sample (Figure 2).

Figure 2 Distribution of the test sample



	C	D	H	J	M
1	16			7	1
0	270	10	75	97	25

Table 5 shows the number of entities with potential for I4.0 allocated to the test sample, where the number of small enterprises is the greatest (43.8%), followed by medium-sized enterprises (37.5%) and large-scale enterprises (18.8%).

Table 5 Structure of the I4.0 test sample according to company size

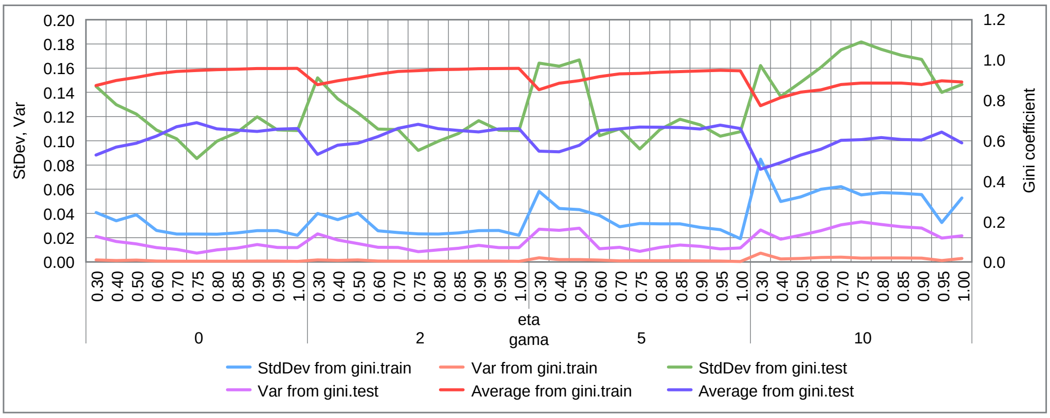
Company size	Number of companies with potential	Share, %
1 – small enterprises	14	43.8%
2 – medium-sized enterprises	12	37.5%
3 – large-scale enterprises	6	18.8%
Total training sample	32	100.0%

3.3 The model for estimating the potentials for I4.0 application

The model for estimating the potentials for application of fourth industrial revolution technologies was subject to learning on the training sample, and it was verified using the test sample. In order to avoid excessive adaptation of the model to the training sample (overfitting), an iterative method of sampling from a non-probabilistic sample was applied for both samples, the training sample and the test sample, the so-called bootstrapping method that creates different samples from the initial sample through random selection, which examines the performance of the out-of-sample model. The procedure was repeated 20 times, during which the companies entering the samples were replaced through random selection and the set distribution of samples by groups of activities was retained (Figure 1 and Figure 2). The XGBoost methodology uses several parameters in model evaluation (see chapter 3.2). A binomial logistic objective function (objective = “binary: logistic”) was applied, given the objective dependent variable that takes only two values: 0 or 1, and the result of the classification is the probability of using I4.0 technologies. Four (nrounds = 4) was selected as the number of learning iterations, and since the sample is small, there is a small number of companies that meet the criterion of the dependent variable, as well as the number of independent variables of potential classifiers. The depth of learning was kept at low values for those reasons: 2 (max.depth = 2). Each pair of training-test samples was taken to evaluate and verify the discriminant power of the model and its stability with regard to sample change for different gamma parameters (γ ... minimum degree of loss reduction; see equation (8)) and eta (η ... learning rate; see equation (9)) of the XGB classification.

The results obtained are presented graphically (Figure 3), whereby the criteria for selecting the optimal parameters for model evaluation are the maximisation of the discriminant power of the model measured by the Accuracy Ratio (AR or GINI) and minimising its standard deviation at the same time. The obtained average values of the Gini coefficient for 20 iterations with the given combinations of gamma and eta parameters and the corresponding standard deviations were ranked according to the optimisation criteria (Table 6). The average Gini coefficients were ranked from highest to lowest (highest = 1), while standard deviation was ranked from lowest to highest (lowest = 1). The total rank is the sum of these two ranks, and the best (lowest) rank is the optimum combination of gamma and eta parameters.

Figure 3 Discriminant power and stability of the XGB model with regard to changes in gamma and eta parameters



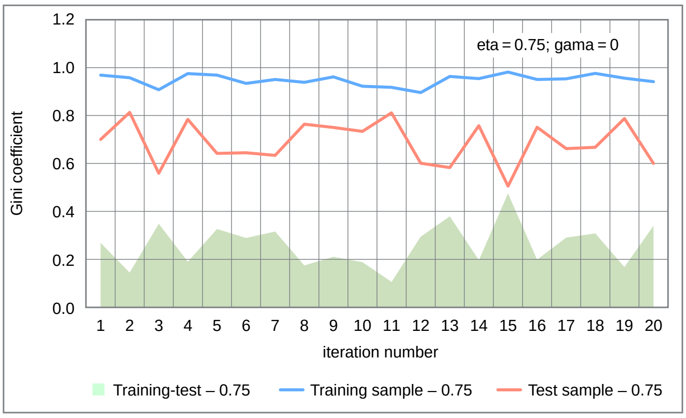
The optimal combination of gamma and eta parameters was obtained for gamma = 0 and eta = 0.75. The resulting Gini coefficients for the training sample and the test sample with the optimal parameters of gamma and eta in 20 performed iterations are shown graphically (Figure 4), where the movement of the Gini coefficient from approximately 0.9 to 0.98 for training samples is noticeable (average 0.95) and between 0.55 and 0.81 (average 0.69) for test samples.

Table 6 Ranking the results of iterations according to optimisation criteria

	Training sample		Test sample					
gama eta	Rank 1 (avg. gini)	Rank 2 (–StdDev)	Rank 3 (avg. gini)	Rank 4 (–StdDev)	Rank 5 (avg. gini)	Rank 6 (–StdDev)	Rank 5+6	Best rank?
0	0.959	0.02191	0.690	0.08533	25	25	50	
0.3	11	10	11	11	47	34	81	
0.4	10	8	10	10	41	29	70	
0.5	9	9	9	9	37	26	63	
0.6	8	7	8	6	28	15	43	
0.7	7	2	2	3	4	6	10	
0.75	6	3	1	1	1	1	2	yes
0.8	5	1	4	2	14	4	18	
0.85	4	4	6	4	18	10	28	
0.9	3	5	7	8	21	24	45	
0.95	2	6	5	7	17	16	33	
1	1	0	3	5	9	13	22	
2	0.958	0.02198	0.682	0.09201	26	27	53	
0.3	11	9	11	10	45	37	82	
0.4	10	8	10	9	38	31	69	
0.5	9	10	9	8	36	28	64	
0.6	8	5	8	6	29	19	48	
0.7	7	4	2	5	8	17	25	
0.75	6	2	1	0	2	2	4	
0.8	5	1	4	1	12	5	17	
0.85	4	3	6	2	19	9	28	
0.9	3	6	7	7	22	22	44	
0.95	2	7	5	4	16	14	30	
1	1	0	3	3	10	12	22	
5	0.949	0.01919	0.677	0.09328	24	30	54	
0.3	11	10	10	9	43	41	84	
0.4	10	9	11	8	44	39	83	
0.5	9	8	9	10	39	42	81	
0.6	8	7	8	2	20	8	28	
0.7	7	3	6	5	13	20	33	
0.75	6	6	2	0	5	3	8	
0.8	5	5	3	4	6	18	24	
0.85	4	4	4	7	7	23	30	
0.9	3	2	7	6	15	21	36	
0.95	1	1	1	1	3	7	10	
1	2	0	5	3	11	11	22	
10	0.897	0.03252	0.643	0.13665	40	44	84	
0.3	11	10	11	5	49	40	89	
0.4	10	1	10	0	48	32	80	
0.5	9	3	9	3	46	36	82	
0.6	8	8	8	4	42	38	80	
0.7	6	9	6	8	34	46	80	
0.75	4	4	4	10	32	48	80	
0.8	3	7	2	9	30	47	77	
0.85	5	6	3	7	31	45	76	
0.9	7	5	5	6	33	43	76	
0.95	1	0	1	1	23	33	56	
1	2	2	7	2	35	35	70	
	0.959	0.01919	0.690	0.08533				

Source: Author's work.

Figure 4 Gini coefficients of training and test samples with optimal gamma and eta parameters (XGB)



In the 20 iterations performed, the variables with the highest total information gain[8] are: operating in high-technology industry (Eurostat classification)[9], share of development expenditure in long-term assets, relative change in the share of concessions, patents and licences in total long-term assets in the period 2012 – 2017, share of exports in income, relative change in the share of intangible assets in long-term assets in the period 2012 – 2017, ratio of market to nominal capitalisation, age of the company, investments in new long-term assets per employee, long-term financial assets in total assets, operating expenses in income, etc. (Figure 5). For a detailed list of all variables and indicators used, see Annex B. Most of the relevant classification variables are of a structural nature (ratios in the balance sheet and profit and loss statement, such as share of research and development expenditure in long-term assets averaging approximately 16% in I4.0 companies, and less than 0.3% in traditional companies, analysed in the non-probabilistic sample, see Table 13), which is expected given the high share of research and development assets, as well as technological assets in the total assets of I4.0 companies in relation to traditional companies.

---

[8] Sum of information gain in 20 iterations

[9] The level of technological intensity of an industry is determined according to the Organization for Economic Cooperation and Development (OECD) and Eurostat classification of research and development intensity of individual industries as follows:

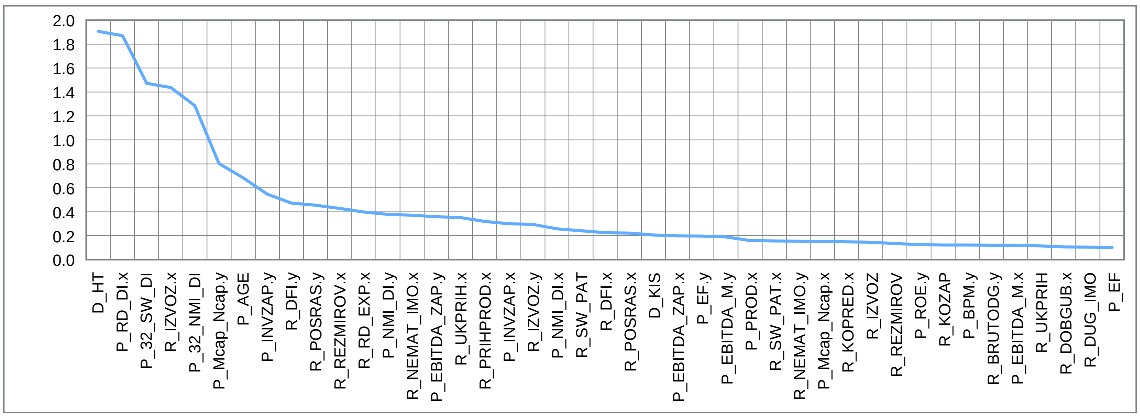
a) high-technology (HT): C21 – Manufacture of basic pharmaceutical products and pharmaceutical preparations, C26 – Manufacture of computer, electronic and optical products

b) medium high-technology (MHT): C20 – Manufacture of chemicals and chemical products, C27 – Manufacture of electrical equipment, C28 – Manufacture of machinery and equipment, C29 – Manufacture of motor vehicles, trailers and semi-trailers, C30 – Manufacture of other transport equipment

c) medium low-technology (MLT): C19, C22 – C25, C33 – manufacture of coke and refined petroleum products, rubber and plastic products, mineral products and basic metals, repair and installation of machinery and equipment

d) low-technology (LT): C10 – C18, C31 – C32 – manufacture of food products, tobacco products, beverages, textiles and wearing apparel, leather products, wood, paper and paper products, printing and reproduction of recorded media, manufacture of furniture and other manufacturing.

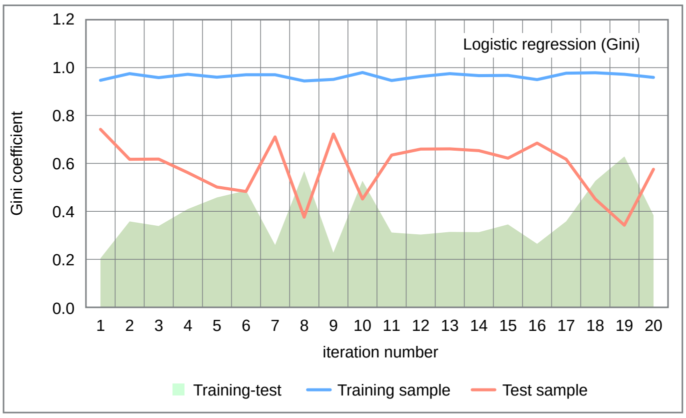
Figure 5 Total information gain of individual variables in 20 iterations



Note: Description of the variables is provided in Annex B.

For the purpose of comparison, the model was also tested using logistic regression on the same data (Annex C). The results obtained with regard to the discriminant power of the model are comparable with XGB models, but overfitting is much greater, which is noticeable at lower values of Gini coefficient of the test sample (Figure 6 and Figure 4). Details regarding the models including logistic regression are provided in Annex C.

Figure 6 Discriminant power of the logistic model

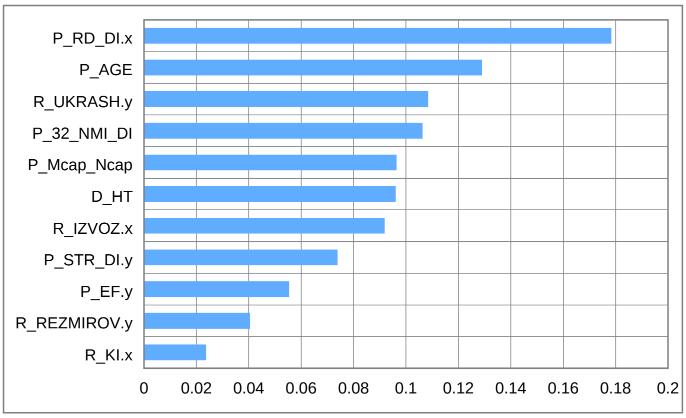


3.3.1 Model evaluation using machine learning

The model from the second iteration for gamma = 0 and eta = 0.75 was selected as the final model for estimating the potential for application of I4.0, in which model overfitting is the lowest (the highest discriminant power of the model on the test sample).

The evaluated XGB decision tree is shown in Annex B (Figure B.1), and the variables with the highest information gain included in the model are the following (from the

Figure 7 Information gain of variables of the final model



Note: Description of the variables is provided in Annex B.

highest information gain to the lowest, Figure 7): share of development expenditure in long-term assets (positive effect), age of the company (positive effect), ratio of total expenses to operating income (negative effect), relative change in the share of intangible assets in long-term assets in the period 2012 – 2017 (positive effect), ratio of market to nominal capitalisation (positive effect), operating in high-technology industry (indicator variable, positive effect), share of exports in income (positive effect), share of plant and machinery in long-term assets (positive effect), efficiency indicator: operating income per employee (positive effect), share of provisions for pensions, severance pay and similar liabilities in assets (positive effect) and the share of short-term assets in assets (negative effect).

The high share of development expenditure (long-term intangible assets) in long-term assets is a consequence of significant initial investments in development and continuous improvement of high technologies. The results of such development expenditures are expected over a longer period of a company's operations. The model shows that entities with a higher share of development expenditure in long-term assets have a greater inclination towards Industry 4.0. Since older high-technology companies had started investing at an earlier date and had invested more in high technology, according to this model, this sets them apart from traditional companies. These are larger companies that have stable business models and substantial investments in technology. Operating income arises from the core business and according to the model, if total expenses of a company are lower than its operating income, there is a greater inclination towards Industry 4.0.

Companies that increased the share of intangible assets in long-term assets in the period under observation, between 2012 and 2017, show a greater inclination towards Industry 4.0 because intangible assets, inter alia, include the value of patents, software, licences and various types of intellectual property. Market capitalisation is considered to be the market value of company's shares, and if it is higher than nominal capitalisation, it is an indicator that the market has recognised the company in question as successful. Higher exports are a consequence of greater market competitiveness and innovation. Operating income per employee is a high-quality indicator of efficiency. The share of provisions for pensions, severance pay and similar liabilities in total assets in the model has a positive contribution. As the number of employees increased in companies with potentials for I4.0 (Table 8) during 2017, the high share of provisions for pensions, severance pay and similar liabilities suggests that companies with potentials were “rejuvenating” the human resources structure.

The discriminant power of the model on the training sample is exceptional – the Gini coefficient is 0.95 (Figure 8), while it is slightly lower on the test sample (0.8; Figure 9), with the highest obtained value in 20 iterations.

Figure 8 Cumulative accuracy profile (CAP) curve of the final model on the training sample

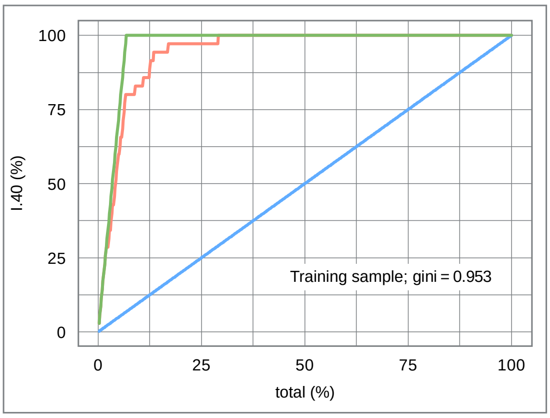
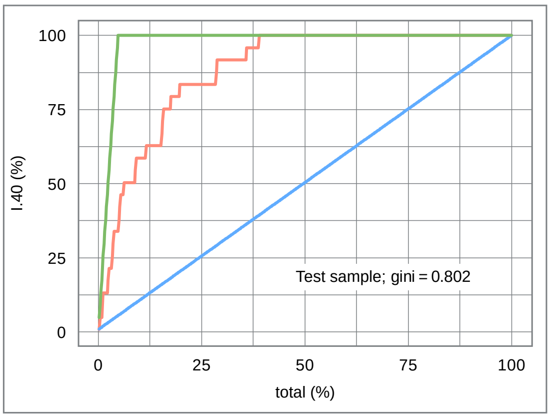


Figure 9 Cumulative accuracy profile (CAP) curve of the final model on the test sample

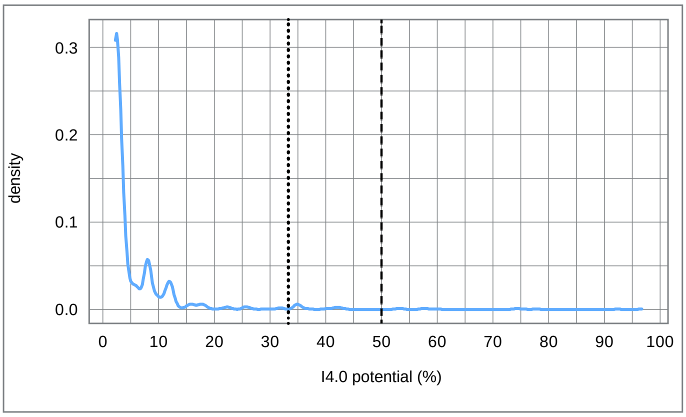


The exceptionally high discriminant power of the model on the test sample, as well as on the training sample, confirms that there is no significant overfitting of data to the training sample, and the estimates obtained by the model can be considered unbiased.

4 Analysis of results

Figure 10 shows the probability distribution density function of potential for I4.0 on the set of analysed companies defined as the probability that the model assigns in the classification of I4.0 companies. The figure shows the highest concentration of companies within the first 20% of probability, while above 50% of probability the distribution density is very low, which indicates a very low readiness for the application of I4.0 technologies in Croatia. Of the 7147 companies analysed, 141 companies (including 58 identified through expert assessment) were classified as companies with potential for I4.0, which makes up 1.97% of all analysed entities.

Figure 10 I4.0 potential distribution density function



4.1 Analysis of I4.0 potential on the whole set (non-probabilistic sample)

Table 7 shows the potentials of Industry 4.0 that include potentials identified through expert assessment and model-detected potentials by activities. The largest share by number of companies in the total potential for I4.0 has the group of activities C – manufacturing (44.7%), followed by the group of activities J – information and communication (42.6%). The share of activities of group M – professional, scientific and technical activities (9.9%), H – transportation and storage, and D – electricity, gas, steam and air conditioning supply (1.4% each) is significantly smaller.

According to this research, most of the companies that have the potential for Industry 4.0 in Croatia perform the activity J-6201 computer programming activities (38), C-2620 manufacture of computers and peripheral equipment (11), J-6202 computer consultancy activities, J-6209 other information technology and computer service activities, and M-7219 other research and experimental development on natural sciences and engineering (five entities in each category, Figure 11).

Figure 12 shows the marked and model-detected potentials according to the share of company size. The share of small enterprises (52%) in the potentials of the group is the greatest, while medium-sized (25%) and large-scale enterprises (23%) account for similar shares.

However, although the number of companies with potential for I4.0 is not large, they account for approximately 27% of assets of the non-probabilistic sample (approximately 24% of the population of analysed activities) and approximately 26% of operating income (22% of the population of analysed activities), Table 8.

Table 7 Number and share of companies with potential for I4.0 across industries – non-probabilistic sample

Activity	Number	Share
C – manufacturing	63	44.7%
Manufacture of electricity distribution and control apparatus	4	2.8%
Other manufacture (of tools, electric motors...)	38	27.0%
Manufacture of pharmaceutical preparations	3	2.1%
Manufacture of instruments and appliances for measuring, testing and navigation	2	1.4%
Manufacture of communication equipment	3	2.1%
Computer activities, data processing, hosting and related activities	11	7.8%
Services (other software publishing, other information service activities, management activities...)	2	1.4%
D – electricity, gas, steam and air conditioning supply	2	1.4%
Electricity distribution and transmission	2	1.4%
H – transportation and storage	2	1.4%
Services (other software publishing, other information service activities, management activities...)	2	1.4%
J – information and communication	60	42.6%
Wired and wireless telecommunications activities	3	2.1%
Computer activities, data processing, server services and related activities	53	37.6%
Services (other software publishing, other information service activities, management activities...)	4	2.8%
M – professional, scientific and technical activities	14	9.9%
Engineering activities and related technical consultancy	4	2.8%
Other research and experimental development on natural sciences and engineering	5	3.5%
Services (other software publishing, other information service activities, management activities...)	5	3.5%
Total	141	100.0%

Figure 11 Distribution of I4.0 potential across classes of activity

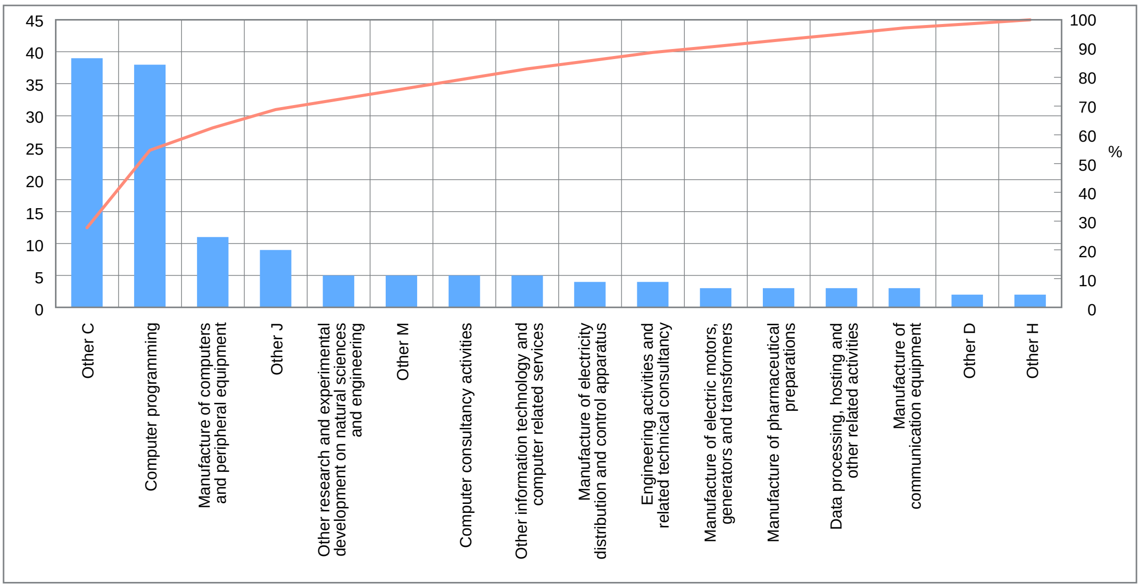
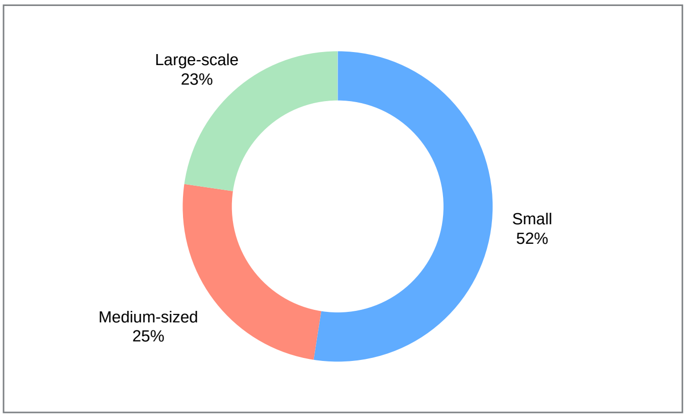


Figure 12 I4.0 potentials according to company size



Source: FINA (author's work).

Table 8 Industry 4.0 potentials

Potential/size	Number of employees in 2016	Total employees 2017	Number of companies	Share in sample assets	Share in assets	Share in sample operating income	Share in operating income
Traditional	227,256	231,710	7,006	72.7%	100.0%	74.1%	100.0%
SMALL	76,223	78,470	6,383	12.5%	17.2%	17.3%	23.35%
<=10	14,491	14,401	4,393	3.8%	5.2%	4.1%	5.59%
11-50	39,913	41,196	1,740	7.4%	10.1%	10.2%	13.78%
51-250	19,885	20,809	244	1.4%	1.9%	2.9%	3.89%
>250	1,934	2,064	6	0.0%	0.0%	0.1%	0.09%
MEDIUM-SIZED	65,075	66,456	493	15.3%	21.0%	20.7%	27.98%
<=10	63	42	7	0.5%	0.6%	1.1%	1.46%
11-50	2,124	1,938	57	2.8%	3.9%	2.4%	3.29%
51-250	46,730	47,269	384	10.6%	14.6%	14.8%	20.00%
>250	16,158	17,207	45	1.4%	1.9%	2.4%	3.23%
LARGE-SCALE	85,958	86,784	130	45.0%	61.8%	36.0%	48.67%
<=10	8	8	1	0.6%	0.9%	0.1%	0.16%
11-50	164	134	5	3.2%	4.4%	3.6%	4.83%
51-250	2,640	2,720	16	2.5%	3.5%	3.5%	4.71%
>250	83,146	83,922	108	38.6%	53.1%	28.9%	38.97%
I4.0 potential	47,696	48,372	141	27.3%	100.0%	25.9%	100.0%
SMALL	1,958	2,132	74	0.4%	1.5%	0.5%	2.09%
<=10	96	92	16	0.0%	0.1%	0.0%	0.10%
11-50	1,121	1,203	45	0.3%	1.2%	0.4%	1.37%
51-250	741	837	13	0.1%	0.3%	0.2%	0.62%
MEDIUM-SIZED	5,146	5,489	35	1.6%	5.7%	1.5%	5.72%
11-50	110	120	3	0.1%	0.2%	0.1%	0.31%
51-250	2,884	3,157	27	1.3%	4.9%	1.1%	4.13%
>250	2,152	2,212	5	0.2%	0.6%	0.3%	1.27%
LARGE-SCALE	40,592	40,751	32	25.3%	92.8%	23.9%	92.19%
51-250	736	786	4	0.7%	2.4%	0.7%	2.88%
>250	39,856	39,965	28	24.6%	90.4%	23.2%	89.32%
Total	274,952	280,082	7,147	100.0%		100.0%	
Share of groups of activities C, D, H, J, M in the population		78.2%	34.9%	88.0%		84.5%	

Source: FINA (author's work).

Most of the entities (48 companies) are grouped as having 11 to 50 employees. The least of them are grouped as having less than 10 employees (16), which can be explained by reference to having more difficulties regarding the availability of sources of funding for development/investment projects due to increased risk and lack of human resources for the implementation of complex, high-technology projects.

Table 9 Comparative presentation of samples and potentials

Type of industry/Company size	Non-prob. sample	Non-prob. sample – potentials	Test sample	Test sample – potentials	Training sample	Training sample – potentials
C – manufacturing	2803	63	286	19	290	21
1 – small enterprises	2334	21	240	7	237	8
2 – medium-sized enterprises	364	18	29	6	38	5
3 – large-scale enterprises	105	24	19	6	15	8
D – electricity, gas, steam and air conditioning supply	103	2	10	0	11	1
1 – small enterprises	79	0	7	0	9	0
2 – medium-sized enterprises	14	0	0	0	1	0
3 – large-scale enterprises	10	2	3	0	1	1
H – transportation and storage	747	2	75	1	75	0
1 – small enterprises	662	2	68	1	66	0
2 – medium-sized enterprises	63	0	5	0	5	0
3 – large-scale enterprises	22	0	2	0	4	0
J – information and communication	989	60	104	11	109	14
1 – small enterprises	934	43	96	6	99	8
2 – medium-sized enterprises	38	13	6	5	6	3
3 – large-scale enterprises	17	4	2	0	4	3
M – professional, scientific and technical activities	2505	14	26	1	27	2
1 – small enterprises	2448	8	24	0	26	1
2 – medium-sized enterprises	49	4	2	1	1	1
3 – large-scale enterprises	8	2	0	0	0	0
Total	7147	141	501	32	512	38

Source: FINA (author's work).

Table 9 summarises the number of companies for the samples analysed, according to group of activities and company size. The potential for I4.0 is defined for model-estimated probabilities greater than or equal to 50%, while for the training sample and test sample, the potential relates to companies that have been unequivocally identified as using some of the technologies of the fourth industrial revolution. The number of companies that are estimated to have the potential for I4.0 is relatively small compared to the total number of companies performing the analysed activities, but all of them exhibit a high degree of automation of production processes. The most technologically advanced countries apply various support mechanisms for the introduction of I4.0 technologies such as tax reliefs, both in the introduction of technological infrastructure and for investing in education and training of employees (France), high (hyper) depreciation rates and special funds for financing investment and development projects (Italy, Germany, Finland). Along with the direct financial support, developed countries additionally enable and encourage the launch of and investment in I4.0 in various manners. Construction and development of infrastructure is encouraged and regulatory frameworks are adjusted to enable the establishment of start-ups, equal access to available data and the use of high I4.0 technology such as autonomous vehicles, drones and robots. An example of such a country is Estonia, and similar practices are followed by Sweden, Norway and Finland.

4.2 Business performance and riskiness of I4.0 companies

The main indicators showing differences in the potential of I4.0 companies and traditional companies in the analysed sample are of a structural nature, such as share of intangible assets or business equipment and machinery in long-term assets, investment in research and development, share of short-term assets in total assets etc. In addition to structural differences, financial statements of companies with I4.0 potential show significantly better business indicators, some of which are included in the model itself because they have a significant effect in distinguishing companies within the context of estimating potential (Figure 6). Significantly better business performance is most pronounced in terms of indicators of investments, cost efficiency, technical equipment and market competitiveness, while profitability indicators, although higher on average, are not significantly better.

Despite higher capital per employee (marginal rate of technical substitution), the average is not significantly different in relation to companies with traditional technical equipment, as shown by single-factor analysis of variance, ANOVA (Figure 13 and Table 10). The growth of the number of employees during 2017 and in the period 2012

Figure 13 Distribution of the marginal rate of technical substitution (MRTS)

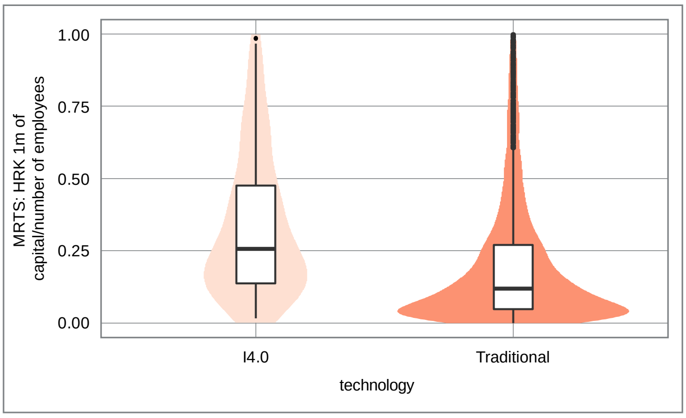


Table 10 ANOVA of the marginal rate of technical substitution

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
I4.0	141	95.65396	0.678397	1.624471
Traditional	7006	2738.178	0.390833	31.6031

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	11.42966	1	11.42966	0.368512	0.543836	3.842761
Within Groups	221607.1	7145	31.01569			
Total	221618.5	7146				

Figure 14 Distributions of average annual staff costs per employee

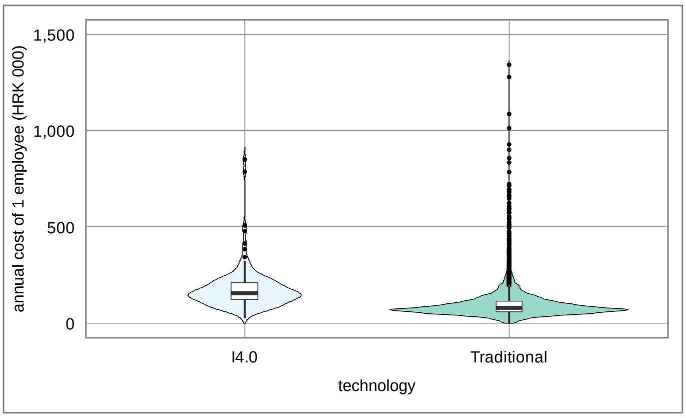


Table 11 ANOVA of average employee cost

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Traditional	7006	685,248,603	97,809	6,998,915,742
I4.0	141	24,938,859	176,871	12,066,478,002

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	863,986,034,781	1	863,986,034,781	121.7188578	0.00000	3.84276064
Within Groups	50,716,711,694,672	7145	7,098,210,174			
Total	51,580,697,729,453	7146				

– 2017 is slightly higher than the growth of the number of employees in traditional industry, but is not statistically significant, while salaries of employees of companies with I4.0 potential are higher, which is statistically significant (Figure 14 and Table 11).

Although the application of high technologies shows negative effects on employment (so-called Keynes' "technological unemployment"), this view ignores other, positive effects of technology, such as creating new (different, more complex) jobs, fostering innovation and productivity and other benefits, e.g. in healthcare, retail, security, Bughin et al., 2017).

Companies with potential for I4.0 are also more competitive in the international market, which is why their share of export revenues in operating income is significantly higher than that of traditional companies (Figure 15 and Table 12).

Figure 15 Distributions of the share of exports in operating income

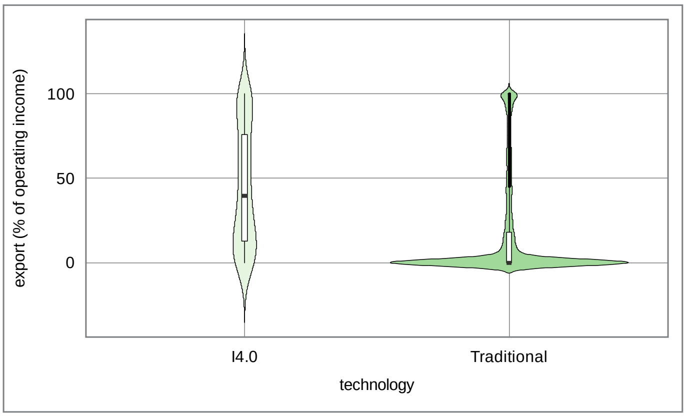


Table 12 ANOVA of the share of exports in operating income

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Traditional	7006	1195.753	0.17067556	0.0918911
I4.0	141	62.9177885	0.44622545	0.12434291

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	10.4946011	1	10.4946011	113.422087	0.00000	3.84276064
Within Groups	661.105142	7145	0.09252696			
Total	671.599743	7146				

Companies with potential for I4.0 invest significantly more than traditional companies in research and development of new technologies in relation to other long-term assets.

Figure 16 Distributions of the share of research and development in long-term assets

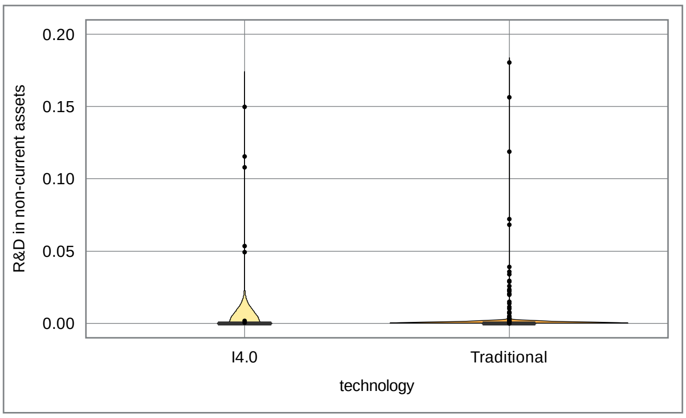


Table 13 ANOVA of the share of investment in research and development in long-term assets

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Traditional	6355	22.670588	0.00357	0.00221
I4.0	141	22.4446999	0.15918	0.07486

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	3.34034285	1	3.34034285	885.29853	0.00000	3.84289117
Within Groups	24.5026798	6494	0.00377313			
Total	27.8430227	6495				

All of the advantages and positive effects resulting from development, investment and use of high technologies according to I4.0 criteria, are reflected in the increase in the value of such companies in the capital market (analysed for companies listed on official stock exchanges) relative to the nominal value of shares, which is also one of the indicators that was included in the model. Efficiency, competitiveness and development strategy are recognised by investors in the securities market, and this has a positive effect on the price thereof.

The riskiness of I4.0 companies is significantly lower than the riskiness of traditional companies (at the level of significance of 1%; Table 14) since none of the I4.0 companies among the approximately 7000 companies under consideration recorded payment delays exceeding 90 days (default, i.e. D rating according to FINA), unlike traditional companies (Figure 17 and Figure 18).

Figure 17 Distribution of probability of default (PD)
probability of default

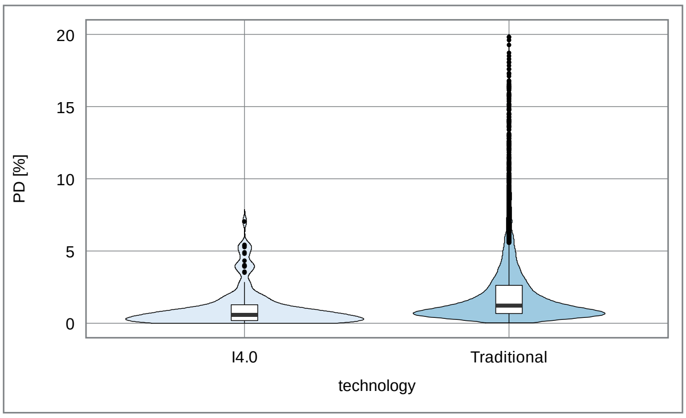


Table 14 ANOVA of probability of default
Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
I4.0	139	1.45472593	0.01046565	0.00017906
Traditional	6994	185.875407	0.02657641	0.00283621

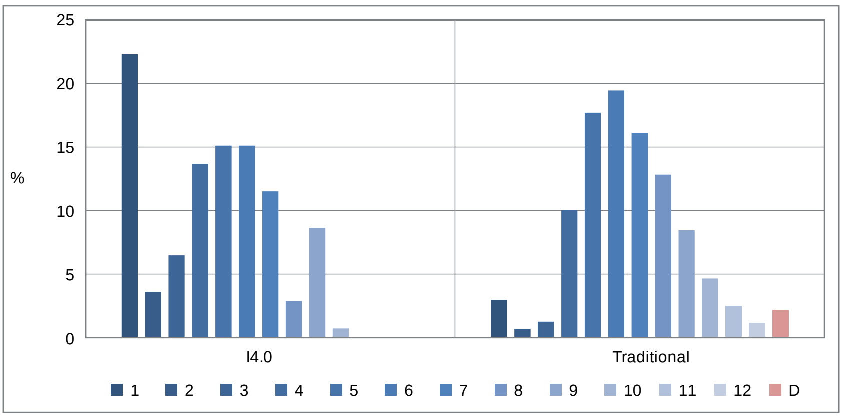
ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.03537529	1	0.03537529	12.7030478	0.00036745	3.84276319
Within Groups	19.8583201	7131	0.00278479			
Total	19.8936954	7132				

Distribution of I4.0 companies and traditional companies across FINA's rating grades (FINA, 2019) is presented in the figure (Figure 18), where an increased proportion of the number of I4.0 companies in better grades is noticeable, particularly in grade 1, with none of them being classified in one of the grades worse than 10, including D (default).

A company's investments in research, development and use of new technologies of the fourth industrial generation demonstrate that there is no objective basis for the perception of I4.0 companies, whose developmental path is based on high technology, as high-risk ventures or high-risk investments. On the contrary, investing in development and application of new technologies opens new markets for such companies, increases the competitiveness of their products and services, raises the level of knowledge and ensures business stability and better profitability and efficiency in the long run.

Figure 18 Distribution of company proportions across rating grades[10]



Note: "D" indicates the rating for companies in default.

Source: FINA (author's work).

5 Conclusion and implications of results in terms of economic policy

Nine technologies constitute Industry 4.0[11] and, depending on their use, we can conclude whether a particular company is on track to realise the I4.0 concept. There are various motivations for the application of I4.0 technology, from increasing the efficiency and productivity of a company, reducing operating costs and increasing profitability in the long run, to market positioning, meeting higher standards regarding quality, etc. Balance sheets of such successful companies exhibit differences in asset structure and business indicators compared to traditional companies. The financial sector must also be ready to finance the development of Industry 4.0.

---

[10] FINA rating classes and associated ranges of probability of default

Rating	PD min.	PD maks.
1	0.00%	0.09%
2	0.09%	0.19%
3	0.19%	0.31%
4	0.31%	0.51%
5	0.51%	0.82%
6	0.82%	1.33%
7	1.33%	2.14%
8	2.14%	3.46%
9	3.46%	5.59%
10	5.59%	9.04%
11	9.04%	14.60%
12	14.60%	99.99%
D	100.00%	100.00%

[11] According to BCG criterion

The initial hypothesis in this paper is that balance sheet structure and business indicators of companies that use I4.0 technologies are similar, which enables the identification of potential users of I4.0, or estimation of the probability that a company is already applying or is in the process of introducing I4.0 technologies. A binomial logistic (logit) classification model calculated using the Extreme Gradient Boosting (XGBoost) technique of deep machine learning was used through the application of supervised learning methodology. Of 7147 analysed companies, 141 companies with potential for I4.0 (1.97% of analysed entities) accounting for approximately 27% of assets of the non-probabilistic sample (approximately 24% of the population of analysed activities) and approximately 26% of operating income (22% of the population of analysed activities) were identified, of which the predominant share was that of large business entities.

The main indicators showing differences in the potential of I4.0 companies in relation to traditional companies are of a structural nature, such as share of intangible assets or business equipment and machinery in long-term assets, investment in research and development, the proportion of short-term assets in total assets etc. Significantly better business performance is most pronounced in terms of investment, cost efficiency, technical equipment and market competitiveness, while in this phase of the introduction of I4.0 technologies, which is still an early one, profitability indicators are higher on average, but the difference is not statistically significant. Although companies with potential for I4.0 have a higher capital-to-labour ratio (capital equipment of labour), the cost of their employee is almost twice as high as in traditional industry. Companies with potential for I4.0 are also more competitive in the international market, which is why their share of export revenues in operating income is significantly higher than that of traditional companies. Increasing cost efficiency, effectiveness and profitability requires significantly greater investment in research and development of new technologies, but due to the period of return on investment, the differences in relation to traditional companies at this stage of development are not significant. Furthermore, the business activity plays an important role and most companies with I4.0 potential are grouped as companies performing computer-related activities, computer activities, data processing, etc. due to easier availability of I4.0 technology in the IT segment, i.e. the fact that they already operate within the scope of a certain segment of I4 0.

The results suggest that the riskiness of I4.0 companies is significantly lower than the riskiness of traditional companies on account of the fact that, out of 7147 companies under consideration, none of I4.0 companies recorded a payment default exceeding 90 days, unlike traditional companies. The concentration of I4.0 companies better rating grades is higher[12], particularly in grade 1, with none of them being classified in one of the grades worse than 10, including D (default). Investment in development and application of new technologies opens new markets for Croatian I4.0 companies, increases the competitiveness of their products and services, raises the level of

[12] FINA's rating, see note 10.

knowledge and ensures business stability, better profitability and efficiency in the long run, making them less risky and more stable than traditional companies. This is also proven empirically and is reflected in the resulting structure of the model: the variables identified by the model that characterise I4.0 companies suggest a higher level of investment in development (higher share of development expenditure in long-term assets), higher relative change of share of intangible assets in long-term assets in the period 2012 –2017 and a higher share of plant and machinery in long-term assets in relation to traditional companies. The model also proved that I4.0 companies are characterised by variables that show a positive effect of I4.0 on their competitiveness and efficiency: ratio of market to nominal capitalisation, share of exports in income and operating income per employee, and a lower ratio of total expenses to operating income (they are more cost-efficient).

The results obtained show that an increase in labour efficiency can be expected (higher revenues per employee) with increased investments in research and development, procurement of new and modernisation of existing plants and equipment, and investments in software solutions for autonomous machine control or artificial intelligence. Boosting competitiveness, exports and a positive investment climate is very important for a small and open European economy that has the opportunity and capacity for development.

Given the stated advantages of I4.0 companies, it is desirable that the government encourages investment in research and development, i.e. in I4.0 technologies, whereby the approach used by developed countries can be applied. These are the establishment of special funds to finance investments and development projects as in Italy, Germany and Finland, adaptation of regulatory frameworks (encouraging the establishment of start-ups, regulating the use of I4.0 technology such as autonomous vehicles, drones and robots), changes to and adaptation of the education system to new work skills that are needed and encouraging the application of new technologies. In a few years, a large part of the population will work in jobs that do not yet exist today. In order to make use of the potential of Croatian companies, it is necessary to create stimulating conditions for development and growth of companies whose activity is related to Industry 4.0, regardless of whether it is an activity that uses I4.0 in its production or produces products and services for Industry 4.0. As this research shows, the companies of the fourth industrial revolution are high-quality companies that, by engaging in this global development trend, have the potential to improve the growth and development of the entire economy, with this being possible if investments in such companies are increased and encouraged.

Annex A

Table A.1 List of I4.0 technologies identified in Croatian companies

No.	ID	Produces I4.0 elements	Uses I4.0 elements	**Head office**	Applied I4.0 technology
1	I001	1		Osijek	Intelligent machines, automated production lines or robotic arms
2	I002	1		Zagreb	Highly automated production processes
3	I003	1		Alaginci	Use of intelligent machines and robots in production
4	I004		1	Orahovica	Computer controlled, robots in production
5	I005	1		Rijeka	Highly automated production, separation of automated warehouse in the second phase of the project
6	I006	1		Rugvica	High degree of automation of production processes
7	I007		1	Vodnjan	Provider of the service of cloud mobile communications for business customers
8	I008		1	Zagreb	Pharmaceutical industry
9	I009	1		Čakovec	Customer Relationship Management (CRM) software solutions for the pharmaceutical industry
10	I010	1		Velika Gorica	A development and investment cycle worth several million euro is being prepared, and it is rapidly moving towards the concept of Industry 4.0.
11	I011		1	Zagreb	*BI, Big Data*
12	I012	1		Zadar	An innovative exporter in the field of high-speed technology and robotic automation. They belong to the very top of global production of sophisticated engines, and their products are used in the automotive industry, mostly in Germany, but also in the markets ranging from USA to Korea. During the period of crisis, they expanded their business operations to industrial and robotic automation.
13	I013		1	Zadar	Robots in glass production
14	I014	1		Samobor	The entire know-how in this sector is the result of own research and development, and almost everything is produced in own highly automated and robotised plants.
15	I015		1	Zagreb	Robots in manufacture and restoration of aircraft equipment
16	I016	1		Zagreb	A global leader in inspection and manufacture of parts and robots for nuclear power plants
17	I017		1	Zagreb	Cloud systems, vertical integration, IoT
18	I018	1		Bakar	Automated production using robots
19	I019	1		Osijek	Highly automated production
20	I020	1		Zagreb	Highly automated lines; management system integration
21	I021		1	Sveta Nedelja	Computer simulations in engineering and design
22	I022	1		Hum Na Sutli	In production they use, for example, robot welding
23	I023	1		Zagreb	Horizontal and vertical system integration
24	I024	1		Sesvete	Intelligently networked compressed air systems in extremely flexible production environments resulting from Industry 4.0; IoT
25	I025	1		Pustodol Začretski	High technology, new production facility with state-of-the-art machines

No.	ID	Produces I4.0 elements	Uses I4.0 elements	Head office	Applied I4.0 technology
26	I026	1		Zagreb	Industrial automation, the first super-fast self-healing grid in Croatia set up as part of a pilot project. For the first time in Europe, such a grid has been set up on a decentralised system with a communication protocol via wireless communication, with real time communication. This technology is the first step towards fully autonomous systems with artificial intelligence within the framework of Smart Grid technologies.
27	I027	1		Zagreb	Optimised production of automatic doors and introduction of the Industry 4.0 to its factory in Zusmarshausen (DE).
28	I028	1		Solin	3D printing; 3D digitalisation, or 3D scanning, a 3D digital database from which one can quickly start designing or redesigning a product. Rapid prototyping or 3D printing is a technique by which a virtual CAD design is transformed into a physical model in a few hours. The 3D printing device reads data from the 3D model, arranging them into micron layers that, depending on the type of device, merge into a compact product.
29	I029		1	Osijek	Robotic team; in pilot projects, their robots will be used by some domestic companies.
30	I030		1	Rijeka	Educational hexapod robot, a system based on which everyone, especially students, can independently create and program a hexapod robot.
31	I031		1	Zagreb	Artificial intelligence and machine learning in text identification and recognition
32	I032	1		Osijek	Manufacture of drone equipment
33	I033	1		Kutina	A private and secure cloud computing environment ensures maximum data protection and privacy. Integration of IoT in agriculture for central management of all plantations and crops within the company
34	I034	1		Zagreb	IoT - the Internet of Things
35	I035		1	Zagreb	Cloud computing tools for microfinance
36	I036		1	Zagreb	IoT - the Internet of Things
37	I037	1		Zagreb	The first Croatian manufacturer of intelligent sensors for Industry 4.0
38	I038	1		Zagreb	Cloud computing platform with geodata
39	I039		1	Zagreb	Data science, machine learning, IoT – the Internet of Things
40	I040		1	Zagreb	Data storage, information security, analytics and cloud computing are just some of their services
41	I041	1		Zagreb	Development of the first national IoT network, and connecting and encouraging companies and individuals to get involved in the development of solutions based on Sigfox technology.
42	I042		1	Zagreb	Application of AI in product quality control on the production line, 3D printers, virtual simulations
43	I043		1	Zagreb	Development and installation of robotic systems, within existing or new production lines. Complete robotic palletisation/depalletisation systems and any other form of robotic manipulation
44	I044	1		Zagreb	A combination of digitalisation, automation and artificial intelligence
45	I045		1	Zagreb	Extensive experience in various areas of implementation of solutions for big data analytics
46	I046		1	Solin	IoT - the Internet of Things
47	I047		1	Zagreb	Applies cloud computing
48	I048	1		Karlovac	Development of complex AI & Analytics solutions for a German car manufacturer
49	I049		1	Zagreb	Digitalisation of business operations (by implementing IT solutions, whether hosting infrastructure or solutions using cloud computing)

No.	ID	Produces I4.0 elements	Uses I4.0 elements	**Head office**	Applied I4.0 technology
50	I050	1		Zagreb	Operates within a high-technology company; production of modular data centres for the global market (cloud computing)
51	I051	1		Zagreb	*Technology that represents an exceptional potential in the construction of a highly manageable and automated IT system that has thus come close to, in terms of its characteristics, what is today popularly called Private Cloud*
52	I052	1		Zagreb	One of the largest private cloud projects in the region, they apply advanced cognitive analytics in various business segments – rapid detection of correlations between data. In terms of technology, they are considering the possibilities of applying IoT technology in the production and sales channel and machine learning in marketing investment management; they are monitoring the development of blockchain and artificial intelligence (AI) technologies.
53	I053	1		Zagreb	Leading role in shaping the trend of Industry 4.0; a member of the German government's "I4.0" initiative, offering comprehensive automation solutions, from the cloud to control technology, IoT gateways, sensors and actuator technology.
54	I054	1		Zagreb	Networking and integration of "smart manufacturing" (Industry 4.0)
55	I055	1		Karlovac	The parent company has one of the largest installed bases in the global robotics industry.
56	I056		1	Zagreb	Real time technology on the way to fully autonomous artificial intelligence (AI) systems
57	I057		1	Zagreb	Automation and robotisation of production, cloud smart factories, three-dimensional engineering and visualisation, smart wiring, fully automated, fast and precise machining of control panels and complete products. Production of control panels – they are also used in leading technologies. Introduction of comprehensive design software that can automate a large number of everyday tasks in design and production, ensuring a level of quality and the ability to quickly adapt to engineering changes that will automatically affect the final product.
58	I058	1		Zagreb	A company specialising in the use of artificial intelligence and its application in data processing in robotics, drones and unmanned aerial vehicles.

Source: Author's work.

Annex B

Table B.1 List of independent variables from the financial statements and the dependent variable

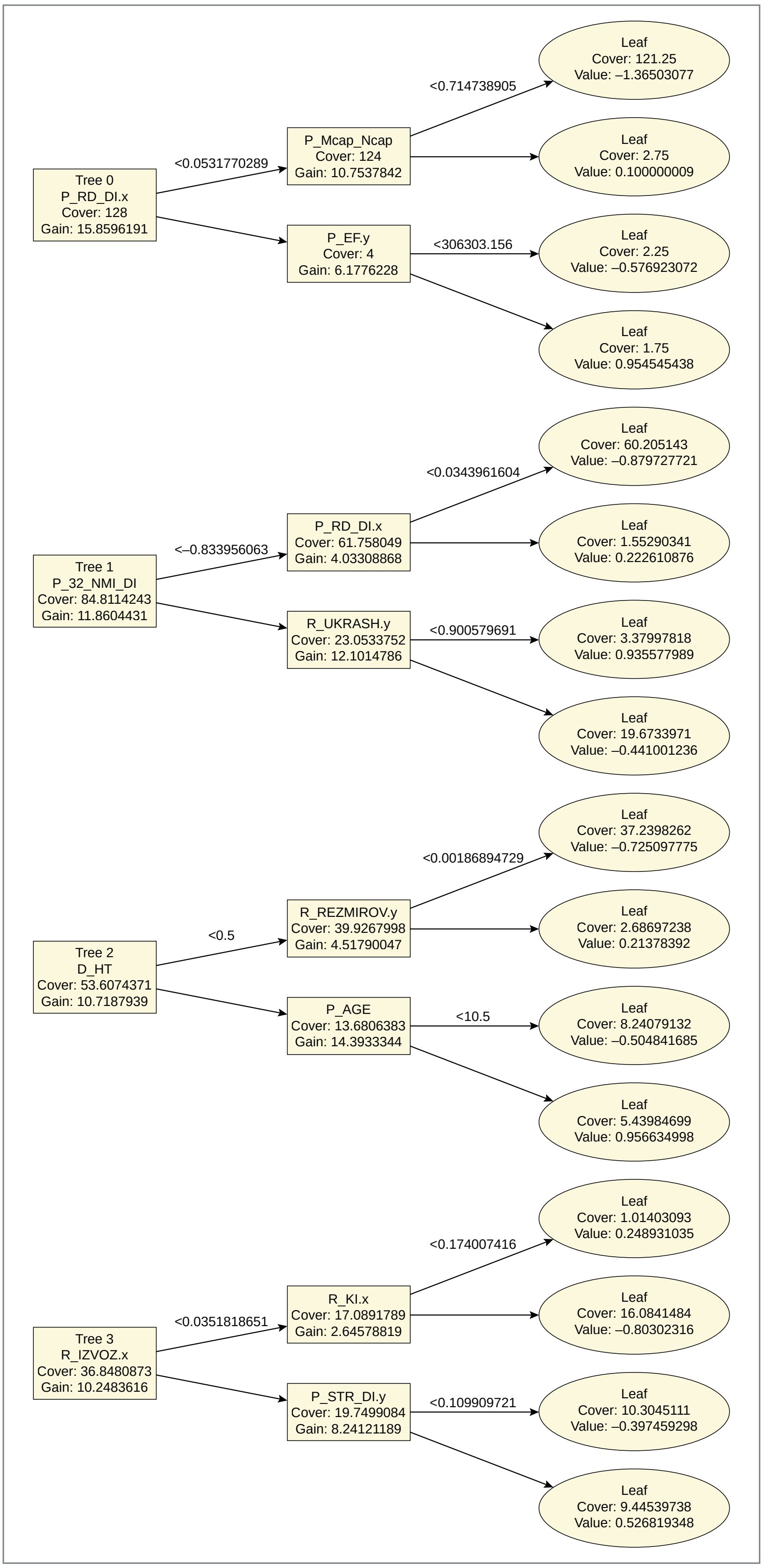
Name of variable	Description	Name of the relative variable	In relation to
DUG_IMO	LONG-TERM ASSETS (AOP 003+010+020+031+036)	R_DUG_IMO	Total assets (TOTAKT)
NEMAT_IMO	INTANGIBLE ASSETS (AOP 004 to 009)	R_NEMAT_IMO	Total assets (TOTAKT)
RD_EXP	Development expenditure	R_RD_EXP	Total assets (TOTAKT)
SW_PAT	Concessions, patents, licences, trademarks and service marks, software and other rights	R_SW_PAT	Total assets (TOTAKT)
STROJEVI	Plant and equipment	R_STROJEVI	Total assets (TOTAKT)
ALATI	Tools, operating inventory and transport assets	R_ALATI	Total assets (TOTAKT)
DFI	Long-term financial assets	R_DFI	Total assets (TOTAKT)
DTPOT	Long-term receivables	R_DTPOT	Total assets (TOTAKT)
DTPOTKRE	Receivables from sales on credit	R_DTPOTKRE	Total assets (TOTAKT)
KI	Short-term assets	R_KI	Total assets (TOTAKT)
ZAL	Inventories	R_ZAL	Total assets (TOTAKT)
KTPOT	Short-term receivables	R_KTPOT	Total assets (TOTAKT)
KTPOTKUP	Of which: accounts receivable	R_KTPOTKUP	Total assets (TOTAKT)
KTFI	Short-term financial assets	R_KTFI	Total assets (TOTAKT)
NOVAC	Cash	R_NOVAC	Total assets (TOTAKT)
TOTAKT	ASSETS		
KAPREZ	Equity and reserves	R_KAPREZ	Total assets (TOTAKT)
REZMIROV	Provisions for pensions, severance pay and similar liabilities	R_REZMIROV	Total assets (TOTAKT)
DUGOBV	LONG-TERM LIABILITIES (084 to 092)	R_DUGOBV	Total assets (TOTAKT)
DOFI	Long-term liabilities to financial institutions	R_DOFI	Total assets (TOTAKT)
DOPRED	Liabilities for advances	R_DOPRED	Total assets (TOTAKT)
DODOB	Trade payables	R_DODOB	Total assets (TOTAKT)
DOVP	Long-term liabilities arising from issued securities	R_DOVP	Total assets (TOTAKT)
KO	Short-term liabilities	R_KO	Total assets (TOTAKT)
KOFI	Short-term liabilities to financial institutions	R_KOFI	Total assets (TOTAKT)
KOPRED	Liabilities for advances	R_KOPRED	Total assets (TOTAKT)
KODOB	Trade payables	R_KODOB	Total assets (TOTAKT)
KOVP	Liabilities arising from securities	R_KOVP	Total assets (TOTAKT)
KOZAP	Short-term liabilities to employees	R_KOZAP	Total assets (TOTAKT)
TOTPAS	LIABILITIES		
PP	Operating income		
PRIHPROD	Sales revenue	R_PRIHPROD	Operating income (PP)
POSRAS	Operating expenses	R_POSRAS	Operating income (PP)
MATTR	Material costs (117 to 119)	R_MATTR	Operating income (PP)
TRSIROV	Costs of raw materials and consumables	R_TRSIROV	Operating income (PP)

Name of variable	Description	Name of the relative variable	In relation to
TRPRODROB	Cost of goods sold	R_TRPRODROB	Operating income (PP)
TRZAP	Employee costs	R_TRZAP	Operating income (PP)
NETOPL	Net salaries and wages	R_NETOPL	Operating income (PP)
AMORT	Amortisation and depreciation	R_AMORT	Operating income (PP)
VUDI	Value adjustment of long-term assets (excluding financial assets)	R_VUDI	Operating income (PP)
VUKI	Value adjustment of short-term assets (excluding financial assets)	R_VUKI	Operating income (PP)
UKPRIH	TOTAL INCOME	R_UKPRIH	Operating income (PP)
UKRASH	TOTAL EXPENSES	R_UKRASH	Operating income (PP)
BRUTODG	PROFIT OR LOSS BEFORE TAX (146-147)	R_BRUTODG	Operating income (PP)
DOBGUB	PROFIT OR LOSS FOR THE PERIOD (148-151)	R_DOBGUB	Operating income (PP)
SUBV	Income from grants, government assistance and subsidies	R_SUBV	Operating income (PP)
KAPPROD	Capitalised production for own needs	R_KAPPROD	Operating income (PP)
IZVOZ	Income from sales abroad	R_IZVOZ	Operating income (PP)
INVDI	Investment in new long-term assets	R_INVDI	Operating income (PP)
ZAPOSL_SATITK	Number of employees based on working hours in the current period		
ZAPOSL_SATIPR	Number of employees based on working hours in the previous period		
Mcap	Market capitalisation		
Ncap	Nominal value of capitalisation		
i40	Designation indicating that a company uses (or produces) the technology of the fourth industrial revolution (dependent variable of the model; source: web)		
D_HT	Company performing a high-technology activity		
D_MHT	Company performing a medium high-technology activity		
D_MLT	Company performing a medium low-technology activity		
D_KIS	Company performing a knowledge-intensive activity		

Table B.2 List of indicators

Designation of indicator	Name of indicator	Numerator	Denominator
P_EBITDA_ZAP	EBITDA per employee	(PP – POSRAS + AMORT)	ZAPOSL_SATITK
P_NMI_DI	Share of intangible assets in long-term assets	(NEMAT_IMO)	DUG_IMO
P_RD_DI	Share of development expenditure in long-term assets	(RD_EXP)	DUG_IMO
P_SW_DI	Share of concessions, patents, licences... in total long-term assets	(SW_PAT)	DUG_IMO
P_STR_DI	Share of plant and machinery in long-term assets	(STROJEVI)	DUG_IMO
P_AL_DI	Share of tools, operating inventory and transport assets in long-term assets	(ALATI)	DUG_IMO
P_EBITDA_M	EBITDA margin	(PP – POSRAS + AMORT)	PP
P_INVEST	Share of amortisation and depreciation in long-term assets	(AMORT)	DUG_IMO
P_PROD	Share of employee costs in operating income	(TRZAP)	PP
P_ROE	Return on equity (ROE)	(DOBGUB)	KAPREZ
P_EBITDA_EQ	EBITDA return on equity	(PP – POSRAS + AMORT)	KAPREZ
P_BPM	Gross profit margin	(BRUTODG)	UKPRIH
P_EF	Efficiency (operating income per employee)	(PP)	ZAPOSL_SATITK
P_POKR	Coverage degree I	(KAPREZ)	DUG_IMO
P_SUBV_PP	Share of income from government subsidies and grants in operating income	(SUBV)	PP
P_EKON	Cost efficiency	(UKPRIH)	UKRASH
P_DFI	Share of long-term financing	(DUGOBV)	(TOTPAS – KAPREZ)
P_INVZAP	Investments in new long-term assets per employee	(INVDI)	ZAPOSL_SATITK
P_Mcap_Ncap	Ratio of market to nominal capitalisation	(Mcap)	
P_AGE	Age of the company		
P_32_NMI_DI	Relative change in indicator P_NMI_DI in the period 2012 – 2017	$(P\_NMI\_DI_{2017} – P\_NMI\_DI_{2012})$	$P\_NMI\_DI_{2012}$
P_32_RD_DI	Relative change in indicator P_RD_DI in the period 2012 – 2017	$(P\_RD\_DI_{2017} – P\_RD\_DI_{2012})$	$P\_RD\_DI_{2013}$
P_32_SW_DI	Relative change in indicator P_SW_DI in the period 2012 – 2017	$(P\_SW\_DI_{2017} – P\_SW\_DI_{2012})$	$P\_SW\_DI_{2014}$
P_32_STR_DI	Relative change in indicator P_STR_DI in the period 2012 – 2017	$(P\_STR\_DI_{2017} – P\_STR\_DI_{2012})$	$P\_STR\_DI_{2015}$
P_32_AL_DI	Relative change in indicator P_AL_DI in the period 2012 – 2017	$(P\_AL\_DI_{2017} – P\_AL\_DI_{2012})$	$P\_AL\_DI_{2016}$
P_32_EBITDA_ZAP	Relative change in indicator P_EBITDA_ ZAP in the period 2012 – 2017	$(P\_EBITDA\_ZAP_{2017} – P\_EBITDA\_ZAP_{2012})$	$P\_EBITDA\_ZAP_{2017}$

Figure B.1 XGBoost model tree



Annex C

On machine learning

Machine learning and data-based approaches are becoming increasingly important in recent years and are applied in many areas: process management and automation, computer science, security (pattern recognition), e-mail classification, fraud detection, anomaly detection, speech recognition, forecasts and process simulations (in finance, healthcare, transportation) and many other areas. There are two key factors on which successful application of machine learning depends: the use of efficient statistical models that reveal complex dependencies between different data and adaptive learning systems that learn from large data sets. Machine learning systems can be supervised and unsupervised. Supervised machine learning systems learn by input data for learning containing the target value of a variable, where the form of data is (input, output) = ($\mathbf{x}$, y). The goal of machine learning is to find the functional connection f between input data $\mathbf{x}$ and the target value y: $y = f(\mathbf{x})$. When y is a continuous variable, the use of regression is more appropriate, and when y is a discrete variable, certain classification algorithm is more efficient.

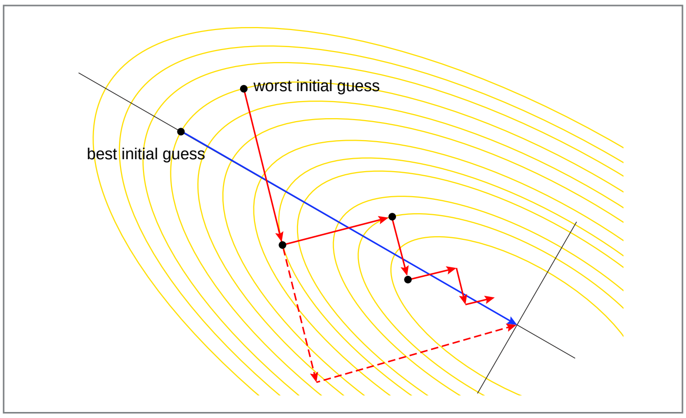
There are several methods of machine learning that are applied in practice (decision tree, random forest, neural network, K-nearest neighbours, decision tree ensemble, support-vector machines, gradient boosting...), among which in recent times, eXtreme Gradient Boosting (XGB) stands out to a more significant degree, see Petropoulos et al. (2018) and Chen et al. (2016). The basis of XGB methodology is the algorithm for boosting a decision tree that builds new decision trees by learning from the errors of the previous tree using sequential learning, achieving higher algorithm speed (fewer iterations) and scalability that allows for lower processor and memory requirements even when dealing with big data.

XGB uses a gradient descent for optimisation. Gradient descent includes the theorem that the function $f(x)$ at an extreme point (minimum) has a gradient $\nabla f(x) = 0$, while at other points the value of the gradient $\nabla f(x)$ corresponds to the direction of increase of the function. Starting from an initially selected point x, we can find the minimum of the function by an iterative procedure by updating the value of x in the direction opposite to the gradient ∇f until we approach zero at the given precision ε:

$$x_{n+1} = x_n - \eta \nabla f(x) \tag{10}$$

η is the learning rate, for which, if too high, the procedure diverges, and if it is too low, the procedure converges slowly.

Figure C.1 Gradient descent



Source: Šnajder (2017).

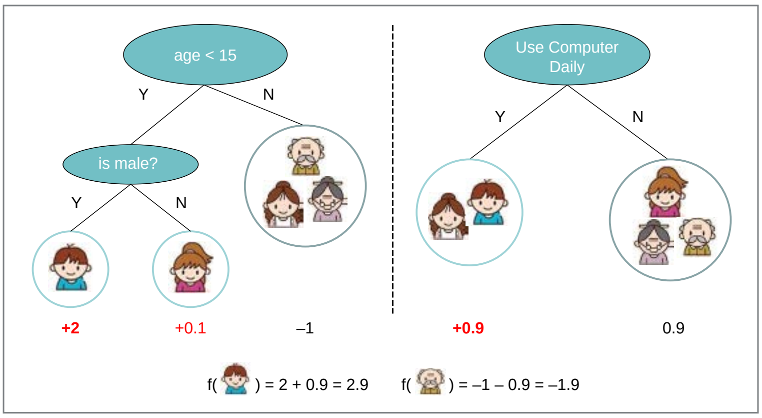
If the function is convex, the minimum found is also the global minimum, otherwise it may be local.

XGB uses the decision tree ensemble, whereby the model is trained in an additive or boosting way, and XGB includes a greedy algorithm that greedily adds the f_t function to the model, which improves the model the most with respect to the regularisation function (see Chen et al., 2016). Ensembles are a common method in building a machine learning algorithm within which a single meta-classifier is built by combining basic classifiers, which results in better classification properties and higher learning speed. The following brief example explains the algorithm of model training using an ensemble of trees, while a detailed explanation is available in the paper by Chen et al. (2016).

The example (Chen et al., 2016): we are seeking a model that will recognise whether a person likes computer games. Inputs are data on age, gender and occupation of a person. The algorithm checks different trees and greedily searches for the optimum for each tree and finally adds the best trees to the model, optimising the objective function, consisting of the loss function l and the regularisation function Ω:

$$Obj = \sum_{i=1}^{n} l(y_i \hat{y}_t) + \sum_{k=1}^{K} \Omega(f_k) \tag{11}$$

Figure C.2 Tree ensemble



Source: Chen et al. (2016).

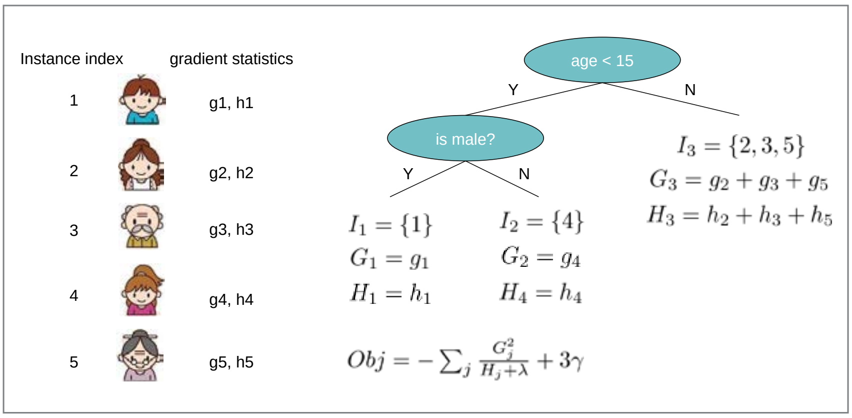
Each leaf in the tree is assigned with a score. Boosting (additive) learning in each iteration t is contained in the sum of the functions retained in the previous iteration $t-1$:

$$\begin{gathered} y_i^{(0)} = 0 \\ \hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\ \hat{y}_i^{(2)} = f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \\ \cdots \\ \hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \end{gathered} \tag{12}$$

The logistic cross-entropy loss function l was used in the research (see equation (7)), and the regularisation function Ω is given by the expression (8), with a learning curve $\eta = 0.75$ (equation (9) and Figure 4). If (12) is included in the objective function (11), with inclusion of (7) and (8)) and after its approximation by the second-order Taylor polynomial, the objective function of the following form can be obtained (for details see Chen et al., 2016):

$$Obj = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j + \lambda} + \gamma T \tag{13}$$

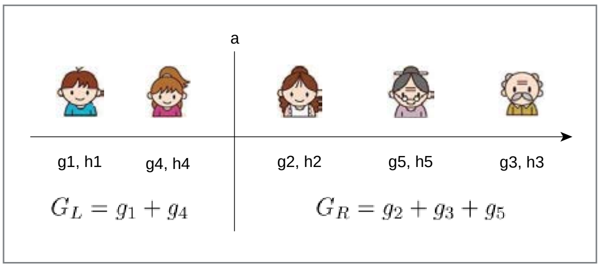
Figure C.3 Structure score calculation



Source: Chen et al. (2016).

where g_i and h_i denote the components of the gradient function of the Taylor polynomial $(g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}),\ h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}))$, while G_i and H_i are their sums. Trees are defined using leaf score vectors, and tree complexity is defined by leaf number and L_2 score norm ((8)). Optimal leaf division is obtained by linear scanning of instances from left to right, for example for the age rule $x_j < a$:

Figure C.4 Optimal division by linear scanning



Source: Chen et al. (2016).

Table C.1 Exact greedy algorithm for split finding

Algorythm 1: Exact greedy algorithm for split finding

Input: I, instance set of current node

Input: d, feature dimension

dobit 0

$G \leftarrow \sum_{i \in I} g_i,\ H \leftarrow \sum_{i \in I} h_i$

for $k = 1$ **to** m **do**

  $G_L \leftarrow 0,\ H_L \leftarrow 0$

  for j **in** sorted(I, by x_{jk}) **do**

    $G_L \leftarrow G_L + g_j,\ H_L \leftarrow H_L + h_j$

    $G_R \leftarrow G - G_L,\ H_R \leftarrow H - H_L$

    $score \leftarrow \max\left(score, \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{G^2}{H + \lambda}\right)$

  end

end

Output: Split with max score

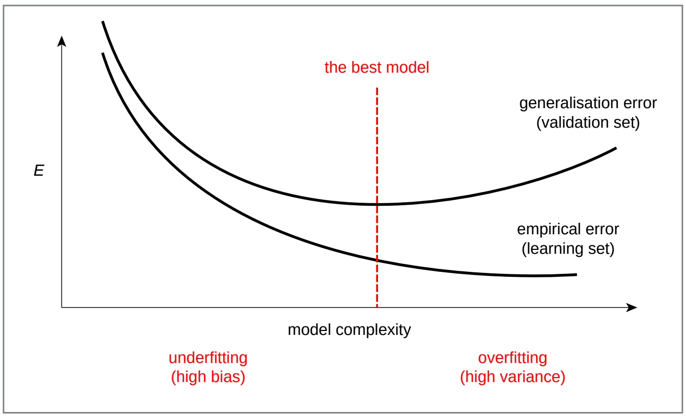
Source: Chen et al. (2016).

where information gain is

$$Gain = \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} - \gamma \tag{14}$$

The exact greedy algorithm for split finding is shown in Table C.1, and XGB uses a version of this algorithm that includes missing values. In this manner, trees of shallower and deeper structures are built, which are then arranged in a tree ensemble, thus forming a network of learned knowledge. By choosing the right learning rate and depth of trees, a compromise can be reached between model overfitting and underfitting, which is most often checked by cross-validation: the model learns based on the learning set (training data), and it is checked using the validation set (test data). Since the classifier is not trained on the validation set data, we can estimate very well how the classifier will behave on unseen data, and the optimum of the model is one in which the empirical error and generalisation error are the smallest:

Figure C.5 Components of a supervised learning algorithm



Sources: Šnajder and Dalbelo Bašić (2014).

Machine learning can be used today to analyse large amounts of data and find dependencies among them, even though their structures are too complex or seem insufficiently connected to draw a conclusion therefrom. Another problem that arises in the application of deep machine learning (excluding overfitting) is unclear interpretation of cause-and-effect and logical connections between data. However, precisely because of their complexity, machine learning techniques generally achieve better results, as evidenced by machine learning competitions such as Kaggle, within which competitors often use ensembles of several different models that achieve greater precision at the cost of making the interpretation of causality more difficult. The problem is less pronounced in flatter structures, which do not branch too deeply, while individual trees ensure sufficient intelligibility for a segmented interpretation of cause-and-effect phenomena. Therefore, this research does not use too great a depth of learning, and at the same time achieves better results than the comparatively examined classical logistic regression.

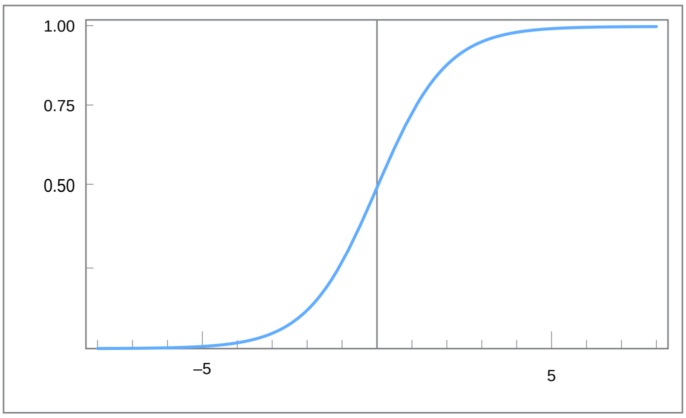
Logistic regression

Logistic regression is a probabilistic discriminant model. Despite the name, it is not a regression, but a classification, the output of which has a probabilistic interpretation, with posterior probability $P(y|x)$:

$$P(x \mid y) = \sigma(\alpha) = \frac{1}{1 + e^{-\alpha}} \tag{15}$$

The logistic function maps all values from the domain of real numbers to the interval <0 , 1> (Figure C.6).

Figure C.6 Logistic or sigmoid function



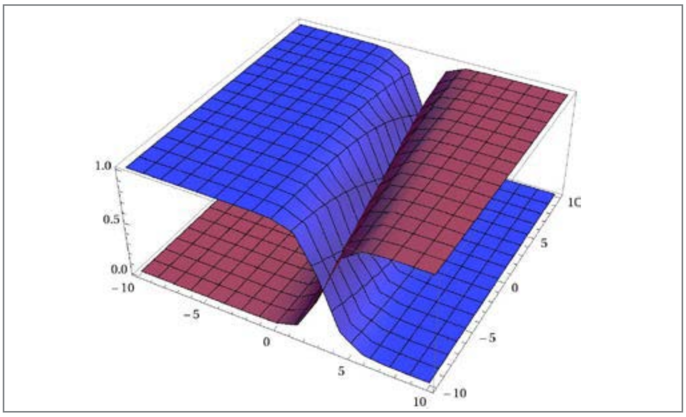
Sources: Šnajder and Dalbelo Bašić (2014).

The value α is a linear combination of weights:

$$\alpha = \ln \frac{p(x \mid y=1)\, P(y=1)}{p(x \mid y=0)\, P(y=0)} = \boldsymbol{w}^T \boldsymbol{x} + w_0 \tag{16}$$

A model for two classes (e.g. I4.0 or traditional company) is shown in Figure C.7.

Figure C.7 A model for two classes, posterior probability modelled by logistic function

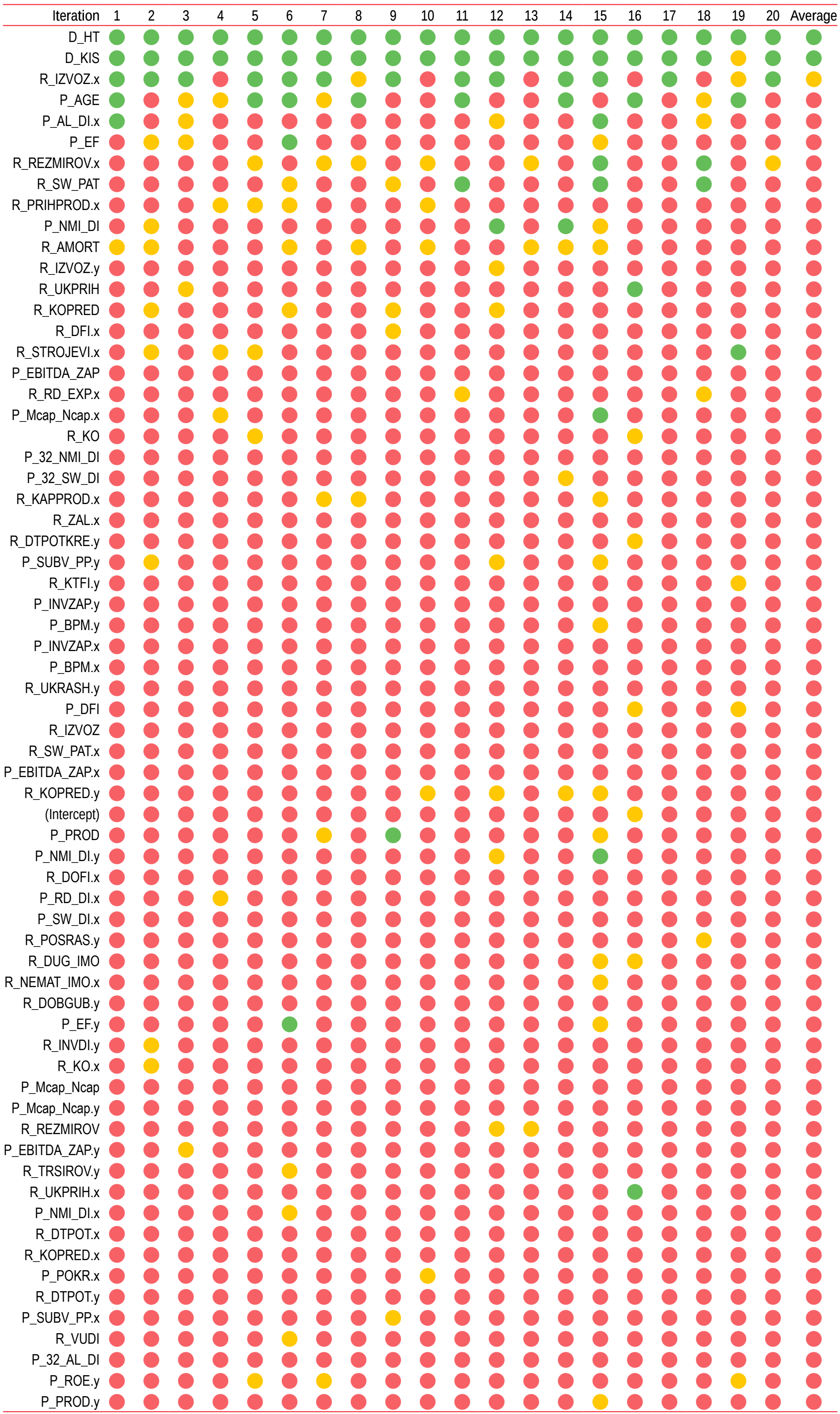


Sources: Šnajder and Dalbelo Bašić (2014).

Learning the logistic regression model comes down to determining the parameters w from the expression (16), or estimators $\tilde{\boldsymbol{w}}$ for $h(x) = \sigma(\tilde{\boldsymbol{w}}^T \tilde{\boldsymbol{x}})$. Solving the logistic regression optimisation problem most commonly uses gradient descent and typically L_2 regularisation.

In this research, logistic regression was applied in an iterative procedure to 20 learning samples and the same number of validation samples. Although the average obtained

Figure C.8 Significance of variables in 20 iterations of logistic regression



Note: ●... >5%, ●... 1–5% ●...<1%.

discriminant properties of logistic regression are comparable to the discriminant properties obtained by the XGB method, overfitting through logistic regression is higher, which can be seen in Figure 4 and Figure 6. Figure C.8 shows the significance of the variables in iterations of logistic regression, ordered from the highest average significance to the lowest. Logistic regression diverged more often during iterations, on average more than 70% of the time, so the procedure had to be repeated several times to achieve convergence on 20 different samples.

Figure C.8 shows that the statistically most significant variables are precisely those that were included in the XGB model, but logistic regression does not synthesise them into a single model, unlike the decision tree ensemble.

References

Acemoglu, D. and P. Restrepo (2017): *Robots and jobs: evidence from US labor markets*, NBER working paper series, available at: http://www.nber.org/papers/w23285 [accessed on 7 February 2019]

Bai, C. et al. (2020.): *Industry 4.0 technologies assessment: A sustainability perspective*, International Journal of Production Economics

BCG: *Embracing Industry 4.0 and Rediscovering Growth*, Boston Consulting Group, available at: https://www.bcg.com/capabilities/operations/embracing-industry-4.0-rediscovering-growth.aspx [accessed on 8 March 2020]

Blunck, E. and H. Werthmann (2017) (2017.): *Industry 4.0 – an opportunity to realize sustainable manufacturing and its potential for a circular economy*, Dubrovnik International Economic Meeting, (3) 1

Bughin, J., J. Staun, J. Andersen, M. Schultz-Nielsen, P. Aagaard, and T. Enggaard (2017): *Digitally-enabled automation and artificial intelligence: Shaping the future of work in Europe's digital front-runners*, McKinsey & Company, October, available at: https://www.mckinsey.com/featured-insights/europe/shaping-the-future-of-work-in-europes-nine-digital-front-runner-countries [accessed on 6 May 2019]

Cerved, 2017 (2017): *Cerved SMEs report*, available at: https://know.cerved.com/imprese-mercati/2017-cerved-smes-report/ [accessed on 5 February 2019]

Chen, T. et al. (2016): *XGBoost: A Scalable Tree Boosting System*, KDD, San Francisco, CA, USA, available at: https://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf [accessed on 14 February 2020]

Eurostat (2016): *Eurostat indicators on High-tech industry and Knowledge – intensive services*, Annex 3 – High-tech aggregation by NACE Rev.2, Eurostat Reference Metadata in Euro SDMX Metadata Structure (ESMS), available at: https://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an3.pdf [accessed on 25 February 2019]

FINA (2019): *Dokumentacija PD modela i metodologije za izračun kreditnog rejtinga poduzetnika u skladu s Basel III smjernicama ver. 2.1* (Documentation of the PD model and methodology for calculating the credit rating of enterprises in accordance with the Basel III guidelines v. 2.1), available at: https://www.fina.hr/documents/52450/138867/Dokumentacija+PD+modela+i+metodologije+za+izracun.pdf [accessed on 23 July 2020]

Frey, C. and M. Osborne (2013): *The future of employment: how susceptible are jobs to computerisation?*, University of Oxford

EIB (2018): *2017 Activity Report*, European Investment Bank, EIB

McKinsey (2018): *Croatia – Emerging Digital Challenger Digitization as the new growth engine for Croatia*, presentation, McKinsey & Company, November, available at: https://digitalchallengers.mckinsey.com/files/Digital-Challengers-Perspective-on-Croatia.pdf [accessed on 19 January 2020]

Novak, J. et al. (2018): *The rise of Digital Challengers: How digitalization can become the next growth engine for Central and Eastern Europe*, McKinsey, available at: https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Europe/Central%20and%20Eastern%20Europe%20needs%20a%20new%20engine%20for%20growth/The-rise-of-Digital-Challengers.ashx [accessed on 19 January 2020]

Petropoulos, A. et al. (2018): *A robust machine learning approach for credit risk analysis of large loan level datasets using deep learning and extreme gradient boosting*, BIS, available at: https://www.bis.org/ifc/publ/ifcb49_49.pdf [accessed on 22 March 2020]

PwC (2014): *Industry 4.0 – Opportunities and Challenges of the Industrial Internet*, available at: https://www.pwc.nl/en/assets/documents/pwc-industrie-4-0.pdf [accessed on 12 February 2019]

Šnajder, J. (2017): *Strojno učenje – natuknice za predavanja* (Machine Learning – Notes for Lectures, Faculty of Electrical Engineering and Computing

Šnajder, J. and B. Dalbelo Bašić (2014): Strojno učenje – *skripta* (Machine Learning – Course Notes), Faculty of Electrical Engineering and Computing

Veža, I., N. Gjeldum and M. Mladineo (2018): *Inovativno pametno poduzeće* (Innovative Smart Company), Faculty of Mechanical Engineering and Naval Architecture

List of figures and tables

List of figures

List of tables

PUBLISHER
Croatian National Bank
Trg hrvatskih velikana 3
10000 Zagreb
T. +385 1 4564 555
www.hnb.hr

ISSN 1334-0131 (online)

Assessment of Readiness of Croatian Companies to Introduce I4.0 Technologies

ISSN 1334-0131 (online)