SpringerBriefs in Applied Sciences and Technology

Vitor Joao Pereira Domingues Martinho

Machine Learning Approaches for Evaluating Statistical Information in the Agricultural Sector

Springer

SpringerBriefs in Applied Sciences and Technology

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical publications can be:

- A timely report of state-of-the art methods
- An introduction to or a manual for the application of mathematical or computer techniques
- A bridge between new research results, as published in journal articles
- A snapshot of a hot or emerging topic
- An in-depth case study
- A presentation of core concepts that students must understand in order to make independent contributions

SpringerBriefs are characterized by fast, global electronic dissemination, standard publishing contracts, standardized manuscript preparation and formatting guidelines, and expedited production schedules.

On the one hand, **SpringerBriefs in Applied Sciences and Technology** are devoted to the publication of fundamentals and applications within the different classical engineering disciplines as well as in interdisciplinary fields that recently emerged between these areas. On the other hand, as the boundary separating fundamental research and applied technology is more and more dissolving, this series is particularly open to trans-disciplinary topics between fundamental science and engineering.

Indexed by EI-Compendex, SCOPUS and Springerlink.

Vitor Joao Pereira Domingues Martinho

Machine Learning Approaches for Evaluating Statistical Information in the Agricultural Sector

Springer

Vitor Joao Pereira Domingues Martinho
Agricultural School (ESAV)
and CERNAS-IPV Research Centre
Polytechnic Institute of Viseu (IPV)
Viseu, Portugal

ISSN 2191-530X ISSN 2191-5318 (electronic)
SpringerBriefs in Applied Sciences and Technology
ISBN 978-3-031-54607-5 ISBN 978-3-031-54608-2 (eBook)
https://doi.org/10.1007/978-3-031-54608-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Acknowledgements

This work is funded by National Funds through the FCT—Foundation for Science and Technology, I.P., within the scope of the project Refª UIDB/00681/2020 (https://doi.org/10.54499/UIDP/00681/2020). This research is also funded by the Enovo Company. This study was carried out under the international project "Agriculture 4.0: Current reality, potentialities and policy proposals" (CERNAS-IPV/2022/008). Furthermore, we would like to thank the CERNAS Research Centre and the Polytechnic Institute of Viseu for their support. This work is too co-financed by the PRR—Plano de Recuperação e Resiliência (República Portuguesa) and the European Next Generation EU Funds (https://recuperarportugal.gov.pt) through application PRR-C05-i03-I-000030—"Carb2Soil—Reforçar a Complementaridade entre agricultura e pecuária para aumentar a fertilidade dos solos e a sua capacidade de sequestro de carbon".

I would also like to thank all those who have contributed in some way to this work.

A special thanks to my wife Lúcia Domingues Martinho and my two daughters Inês Domingues Martinho e Isabel Domingues Martinho.

About This Book

This book presents suggestions, based on machine learning approaches, for identifying the most important predictors of crucial variables for dealing with the challenges of managing production units and designing policies. The book focuses on the agricultural sector in the European Union and considers statistical information from the Farm Accountancy Data Network (FADN). In other words, nowadays, statistical databases present a lot of information for many indicators and, in these contexts, one of the main tasks is to identify the most important predictors of certain indicators. In this way, the book presents approaches to identifying the most relevant variables that best support the design of adjusted farming policies and management plans. These subjects are currently important, namely for the students, public institutions and farmers. To achieve these objectives, the IBM SPSS Modeler procedures were considered, as well as the respective models suggested by this software.

Contents

About the Author

Vitor Joao Pereira Domingues Martinho is Coordinator Professor with Habilitation at the Polytechnic Institute of Viseu Portugal, and holds a Ph.D. in Economics from the University of Coimbra, Portugal. He was President of the Scientific Council, President of the Directive Council and President of the Agricultural Polytechnic School of Viseu Portugal, from 2006 to 2012. He has published several scientific papers, is referee of some scientific journals and participates in the evaluation of national and international projects. Occasionally, he is Guest Professor at Portuguese Higher Education Institutions and participates in several technical and scientific projects.

Chapter 1
Predictive Machine Learning Approaches to Agricultural Output

Abstract The agricultural sector needs to increase agricultural production to guarantee food security worldwide, however, to achieve these objectives agriculture must improve the sustainability of its activities and processes, specifically improving the efficiency of the sector. In these frameworks, adjusted agricultural planning and management is crucial, where the availability of information plays a determinant role, as well as the consideration of new technologies and methodologies. In the context of the new approaches of analysis, digital methodologies may bring relevant added value, namely those associated with predictive machine learning technologies. From this perspective, this study intends to identify the most adjusted models to predict the European Union farming output, taking into account machine learning approaches and statistical information from the Farm Accountancy Data Network. The results obtained highlight the most important farming variables that must be taken into account to predict the total output in the European Union farms.

Keywords IBM SPSS modeler · Farm accountancy data network · European Union farms

1.1 Introduction

Agricultural output is influenced by several factors, some of which are related to water conditions and soil characteristics [1], specifically salinity [2] and soil organic matter [3]. The machine learning approaches may bring relevant contributions to the assessment of these variables, particularly those associated with water quality [4], for example. The consideration of new technologies in the analysis of water dimensions has motivated different research, including on water conservation [5] and groundwater [6].

Another dimension with a great impact on the farming output is climate change and the consequent global warming. The new knowledge related to the digital transition has been used to assess these frameworks, namely in Africa [7], where the consequences of the negative impacts on agriculture may be more severe, in some

V. J. P. D. Martinho, *Machine Learning Approaches for Evaluating Statistical Information in the Agricultural Sector*, SpringerBriefs in Applied Sciences and Technology, https://doi.org/10.1007/978-3-031-54608-2_1

circumstances, because of the current problems of food security. These methodologies enable us to work with Big Data and information collected with alternative and modern technologies, such as the Internet of Things (IoT) and sensors [8]. For agricultural income prediction is crucial to deal with the data volatility in the models as a consequence of global warming and economic tendencies [9]. In these scenarios of climate change, the greenhouse gas emissions mitigation from farming practices, through more precise and innovative procedures, is crucial [10].

Plant diseases appear also between the factors that may have negative implications on agricultural dynamics [11]. The artificial methodologies, biosensors [12] and IoT sensors [13] allow for early assessment of problems associated with plant diseases and this is fundamental for adjusted farming management that permits maintaining, or improving, the expected performance of agriculture.

Digital methodologies offer similarly new opportunities to estimate and project agricultural output during the growing season with important added value for the food conditions and sustainability of the farming sector, especially in contexts with more difficulties. The predictive models have here a crucial support [14]. These approaches may be also taken into account to identify crop types [15] in contexts where is more difficult to collect information.

These digital approaches may bring still useful insights into individual attitudes on food understanding, namely on genetically modified products [16]. These products may contribute to improve the sustainability of the food sector, however, there is still some work to do to better understand the public opinion on genetically modified food.

Considering the context described before, this chapter intends to bring more insights into the agricultural output prediction in the European Union farm contexts, using data from the Farm Accountancy Data Network (FADN) [17] and taking into account machine learning approaches to identify accurate models and important indicators, following the procedures proposed by the software IBM SPSS Modeler [18].

1.2 Data Analysis

The data considered in this research was obtained from the Farm Accountancy Data Network database for European Union countries and the respective agricultural regions. This statistical information is available for the representative farms of each country and agricultural region (when the data are available at member-state, or region, respectively). These representative microeconomic data are found through harmonised bookkeeping principles.

Table 1.1 shows that the European Union countries with the highest/lowest growth rates for the total farming output, over the period 2018–2021, are different, revealing the vulnerability of agriculture to external (market conditions and climate) and internal circumstances. These results are influenced by the effects of the prices. In any case, the intention here is to analyse the changes in the revenues of the farmers and to highlight some variability of the agricultural incomes.

Table 1.1 Growth rate (%) results for the agricultural output of the European Union countries, with data at the farm level, over the period 2018–2021

Member state	Year		
	2019	2020	2021
Austria	*1.356*	2.896	11.857
Belgium	5.578	– 0.754	9.007
Bulgaria	*1.104*	– 2.281	**34.124**
Croatia	*– 3.618*	**8.974**	16.203
Cyprus	1.895	2.277	*– 2.747*
Czechia	4.651	**22.543**	10.014
Denmark	**17.248**	**17.693**	*– 0.698*
Estonia	**20.500**	*– 2.438*	4.772
Finland	**27.861**	2.841	8.449
France	1.777	– 1.749	12.836
Germany	7.678	– 2.177	16.325
Greece	1.464	2.164	15.845
Hungary	4.778	6.744	15.398
Ireland	*– 1.536*	5.833	15.985
Italy	*– 1.656*	3.843	4.746
Latvia	**15.438**	5.726	*2.402*
Lithuania	6.574	**13.981**	9.614
Luxembourg	4.769	– 2.165	5.213
Netherlands	6.040	*– 3.298*	9.692
Poland	5.492	– 2.353	**18.005**
Portugal	3.363	*– 13.644*	**18.064**
Romania	1.579	*– 8.476*	**27.692**
Slovakia	4.483	**8.317**	*4.060*
Slovenia	8.238	*– 3.407*	7.369
Spain	14.096	6.188	*2.404*
Sweden	**14.490**	2.578	**24.148**
Average	6.679	2.687	11.568

Note Bold corresponds to the highest values and italic to the lowest

For a better assessment, Table 1.2 presents the normalised values (obtained through $(x_i - x_{minimum})/(x_{maximum} - x_{minimum})$) for the total output disaggregated at the European Union agricultural region level. These results allow identifying some leader countries/regions from Belgium, Denmark, Germany, Netherlands and Slovakia. Some of the frameworks with the lowest results are from Croatia, Greece, Poland and Romania.

Table 1.2 Normalised values for the agricultural output of the European Union farming regions, with data at the farm level, over the period 2018–2021

Member state	Region	Year			
		2018	2019	2020	2021
Austria	Austria	0.083	0.084	0.086	0.086
Belgium	Vlaanderen	**0.344**	**0.362**	**0.344**	0.341
Belgium	Wallonie	0.167	0.177	0.184	0.175
Bulgaria	Severen tsentralen	0.100	0.099	0.090	0.120
Bulgaria	Severoiztochen	0.120	0.124	0.093	0.145
Bulgaria	Severozapaden	0.102	0.098	0.116	0.129
Bulgaria	Yugoiztochen	0.063	0.068	0.063	0.081
Bulgaria	Yugozapaden	0.016	0.018	0.021	0.013
Bulgaria	Yuzhen tsentralen	0.019	0.019	0.025	0.023
Croatia	Jadranska Hrvatska	*0.005*	*0.006*	*0.007*	*0.006*
Croatia	Kontinentalna Hrvatska	0.009	*0.007*	0.012	0.012
Cyprus	Cyprus	0.034	0.035	0.036	0.029
Czechia	Czechia	0.304	0.318	**0.384**	**0.381**
Denmark	Denmark	**0.490**	**0.576**	**0.661**	**0.592**
Estonia	Estonia	0.119	0.147	0.140	0.131
Finland	Etelä-Suomi	0.087	0.109	0.111	0.108
Finland	Pohjanmaa	0.121	0.164	0.171	0.172
Finland	Pohjois-Suomi	0.126	0.172	0.179	0.167
Finland	Sisä-Suomi	0.105	0.156	0.146	0.136
France	Alsace	0.176	0.170	0.166	0.151
France	Aquitaine	0.186	0.176	0.164	0.158
France	Auvergne	0.104	0.106	0.104	0.102
France	Basse-Normandie	0.238	0.234	0.236	0.248
France	Bourgogne	0.212	0.213	0.210	0.232
France	Bretagne	0.307	0.334	0.315	0.304
France	Centre	0.200	0.204	0.193	0.226
France	Champagne-Ardenne	0.241	0.217	0.188	0.203
France	Corse	0.115	0.119	0.121	0.118

(continued)

Table 1.2 (continued)

Member state	Region	Year			
		2018	2019	2020	2021
France	Franche-Comté	0.195	0.216	0.207	0.203
France	Guadeloupe	0.035	0.055	0.046	0.033
France	Haute-Normandie	0.326	0.307	0.271	0.277
France	Île-de-France	0.275	0.276	0.257	0.289
France	La Réunion	0.096	0.087	0.091	0.080
France	Languedoc-Roussillon	0.128	0.128	0.136	0.121
France	Limousin	0.093	0.094	0.090	0.085
France	Lorraine	0.203	0.192	0.205	0.215
France	Midi-Pyrénées	0.103	0.109	0.107	0.109
France	Nord-Pas-de-Calais	0.240	0.260	0.262	0.254
France	Pays de la Loire	0.284	0.315	0.280	0.297
France	Picardie	0.274	0.289	0.247	0.270
France	Poitou–Charentes	0.243	0.238	0.248	0.265
France	Provence-Alpes-Côte d'Azur	0.196	0.195	0.196	0.173
France	Rhône-Alpes	0.160	0.163	0.162	0.154
Germany	Baden-Württemberg	0.170	0.174	0.158	0.156
Germany	Bayern	0.158	0.167	0.162	0.174
Germany	Brandenburg	**0.756**	**0.969**	**0.907**	**0.954**
Germany	Hessen	0.167	0.176	0.164	0.175
Germany	Mecklenburg-Vorpommern	**0.700**	**0.751**	**0.773**	**0.928**
Germany	Niedersachsen	**0.327**	**0.358**	0.328	**0.344**
Germany	Nordrhein-Westfalen	0.268	0.290	0.268	0.286
Germany	Rheinland-Pfalz	0.193	0.190	0.187	0.191
Germany	Saarland	0.143	0.135	0.141	0.153
Germany	Sachsen	**0.744**	**0.812**	**0.887**	**0.772**
Germany	Sachsen-Anhalt	**0.730**	**0.763**	**0.714**	**0.802**
Germany	Schleswig–Holstein/Hamburg	0.311	0.334	0.303	0.343
Germany	Thüringen	**1.000**	**1.000**	**1.000**	**1.000**
Greece	Ipiros-Peloponissos-Nissi Ioniou	*0.003*	*0.003*	*0.004*	*0.004*
Greece	Makedonia-Thraki	*0.007*	*0.007*	*0.009*	*0.008*
Greece	Sterea Ellas-Nissi Egaeou-Kriti	*0.002*	*0.002*	*0.005*	*0.005*
Greece	Thessalia	*0.007*	*0.008*	*0.009*	*0.009*
Hungary	Alföld	0.057	0.059	0.069	0.070
Hungary	Dunántúl	0.101	0.107	0.107	0.113
Hungary	Észak-Magyarország	0.049	0.063	0.052	0.055
Ireland	Ireland	0.062	0.060	0.064	0.067
Italy	Abruzzo	0.024	0.024	0.025	0.023

(continued)

Table 1.2 (continued)

Member state	Region	Year			
		2018	2019	2020	2021
Italy	Alto Adige	0.081	0.058	0.066	0.056
Italy	Basilicata	0.036	0.036	0.033	0.027
Italy	Calabria	*0.007*	0.012	*0.008*	0.014
Italy	Campania	0.038	0.041	0.044	0.038
Italy	Emilia-Romagna	0.099	0.101	0.100	0.095
Italy	Friuli-Venezia Giulia	0.112	0.089	0.079	0.077
Italy	Lazio	0.051	0.057	0.064	0.051
Italy	Liguria	0.040	0.041	0.044	0.046
Italy	Lombardia	0.180	0.166	0.178	0.167
Italy	Marche	0.028	0.026	0.034	0.030
Italy	Molise	0.027	0.026	0.026	0.025
Italy	Piemonte	0.079	0.088	0.095	0.087
Italy	Puglia	0.027	0.030	0.030	0.030
Italy	Sardegna	0.028	0.029	0.035	0.031
Italy	Sicilia	0.023	0.021	0.022	0.020
Italy	Toscana	0.092	0.068	0.067	0.050
Italy	Trentino	0.045	0.039	0.041	0.040
Italy	Umbria	0.034	0.040	0.045	0.037
Italy	Valle d'Aosta	0.066	0.064	0.060	0.054
Italy	Veneto	0.109	0.106	0.114	0.096
Latvia	Latvia	0.050	0.061	0.065	0.057
Lithuania	Lithuania	0.025	0.028	0.036	0.034
Luxembourg	Luxembourg	0.235	0.247	0.235	0.222
Netherlands	The Netherlands	**0.578**	**0.613**	**0.575**	**0.570**
Poland	Malopolska i Pogórze	*0.002*	*0.003*	*0.002*	*0.003*
Poland	Mazowsze i Podlasie	0.013	0.015	0.014	0.014
Poland	Pomorze i Mazury	0.039	0.045	0.050	0.053
Poland	Wielkopolska and Slask	0.032	0.033	0.034	0.036
Portugal	Açores e Madeira	0.010	0.011	0.014	*0.007*
Portugal	Alentejo e Algarve	0.034	0.036	0.019	0.033
Portugal	Norte e Centro	0.010	0.011	0.012	*0.009*
Portugal	Ribatejo e Oeste	0.027	0.031	0.024	0.026
Romania	Bucuresti-Ilfov	0.054	0.041	0.021	0.014
Romania	Centru	*0.006*	0.010	0.012	0.012
Romania	Nord-Est	0.007	*0.008*	*0.006*	0.011
Romania	Nord-Vest	*0.001*	*0.002*	*0.003*	*0.002*

(continued)

Table 1.2 (continued)

Member state	Region	Year			
		2018	2019	2020	2021
Romania	Sud-Est	0.027	0.027	0.017	0.027
Romania	Sud-Muntenia	0.019	0.019	0.015	0.021
Romania	Sud-Vest-Oltenia	*0.000*	*0.000*	*0.000*	*0.000*
Romania	Vest	0.016	0.013	0.017	0.016
Slovakia	Slovakia	**0.549**	**0.573**	**0.604**	**0.567**
Slovenia	Slovenia	0.011	0.014	0.014	0.011
Spain	Andalucía	0.059	0.068	0.061	0.056
Spain	Aragón	0.088	0.087	0.129	0.101
Spain	Asturias	0.039	0.040	0.043	0.040
Spain	Canarias	0.120	0.125	0.128	0.127
Spain	Cantabria	0.043	0.044	0.045	0.046
Spain	Castilla y León	0.092	0.113	0.117	0.106
Spain	Castilla-La Mancha	0.056	0.090	0.097	0.078
Spain	Cataluña	0.104	0.104	0.106	0.101
Spain	Comunidad Valenciana	0.033	0.046	0.061	0.060
Spain	Extremadura	0.065	0.072	0.081	0.071
Spain	Galicia	0.038	0.049	0.052	0.043
Spain	Islas Baleares	0.042	0.037	0.049	0.045
Spain	La Rioja	0.073	0.071	0.110	0.108
Spain	Madrid	0.058	0.055	0.055	0.059
Spain	Murcia	0.078	0.088	0.104	0.096
Spain	Navarra	0.090	0.105	0.103	0.101
Spain	País Vasco	0.072	0.078	0.068	0.066
Sweden	Län i norra Sverige	0.113	0.138	0.119	0.125
Sweden	Skogsoch mellanbygdslän	0.164	0.176	0.166	0.177
Sweden	Slättbyggdslän	0.181	0.214	0.221	0.253

Note Bold corresponds to the highest values and italic to the lowest

1.3 Main Results

For the period 2018–2021 considered, the models with the highest accuracy for the training set, to predict the total output of the European Union agricultural framework, are presented in Tables 1.3, 1.5, 1.7 and 1.9, respectively, for the years 2018, 2019, 2020 and 2021. In the four years taken into account in this assessment, there are some similarities in the most accurate models.

The relationships between the observed values and the predicted ones, for each year, are those presented in Figs. 1.1, 1.2, 1.3 and 1.4. In general, these figures confirm the accuracy of the models considered.

The most important predictors are those revealed in Tables 1.4, 1.6, 1.8 and 1.10. The importance presented in these tables ranges among 0 and 1. These findings confirm the variability of the farming variables over the years and the difficulty of predicting agricultural output based on internal indicators of the farms. In fact, some of the most important predictors are different in the years considered. There are, however, some predictors that appear in more than one year, bringing relevant insights for the several stakeholders.

Table 1.3 Models with the highest accuracy (the lowest relative error) for the agricultural output of the European Union farming regions, with data at the farm level, for the year 2018

Model	Build time	Correlation	No. fields	Relative error
Linear	1	1.000	13	0.000
Neural net	1	1.000	171	0.001
C&R tree	1	0.988	31	0.026
Random forest	1	0.992	178	0.041
Random trees	1	0.979	178	0.044

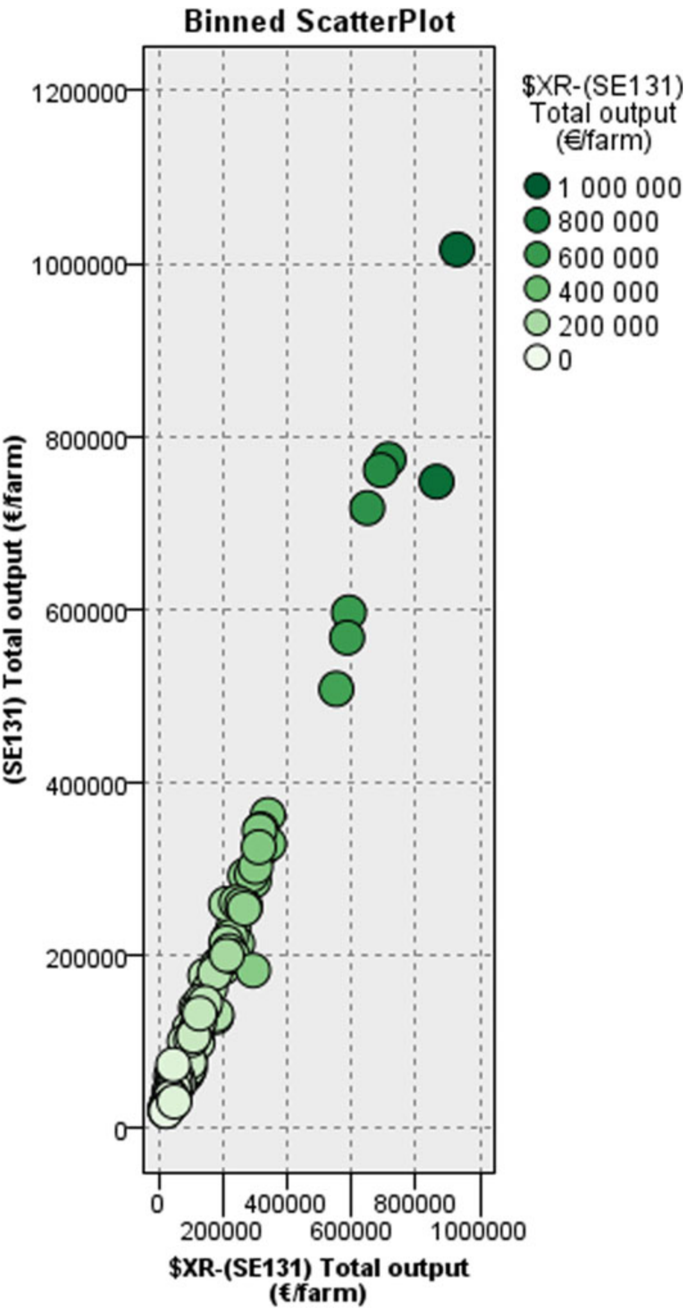

Fig. 1.1 Relationships between the observed values and the predicted ones for the agricultural output of the European Union farming regions, with data at the farm level, for the year 2018

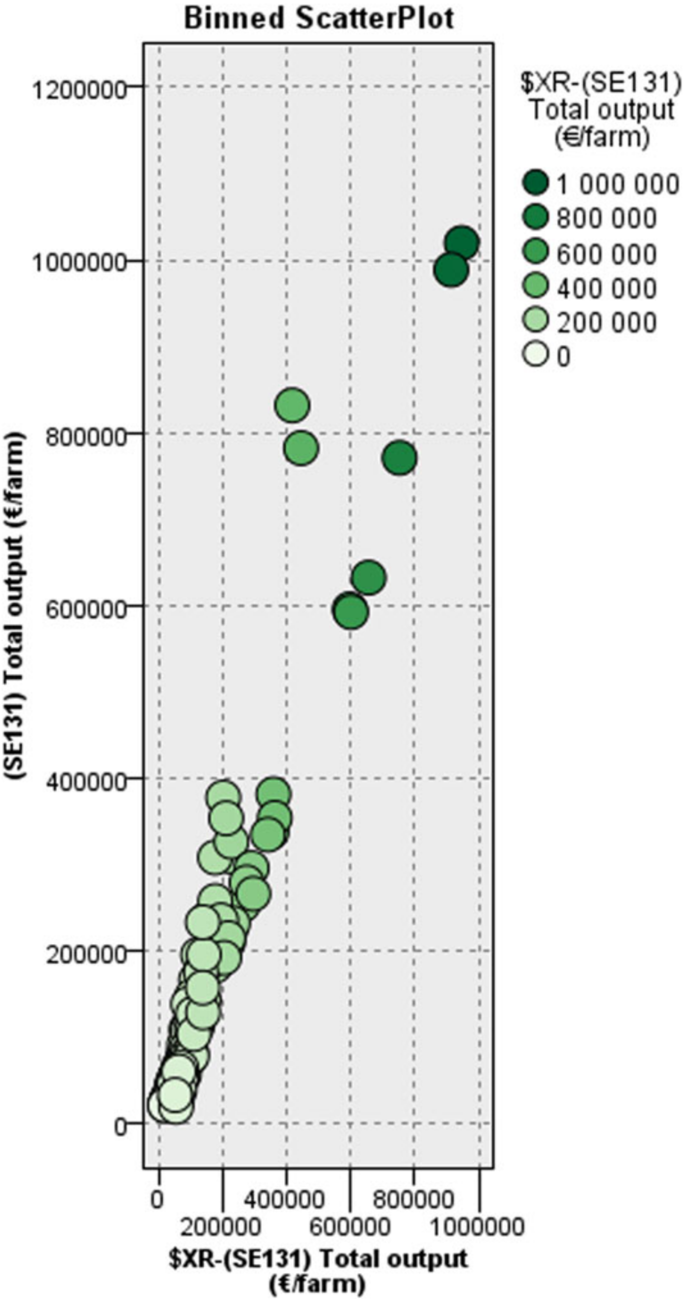

Fig. 1.2 Relationships between the observed values and the predicted ones for the agricultural output of the European Union farming regions, with data at the farm level, for the year 2019

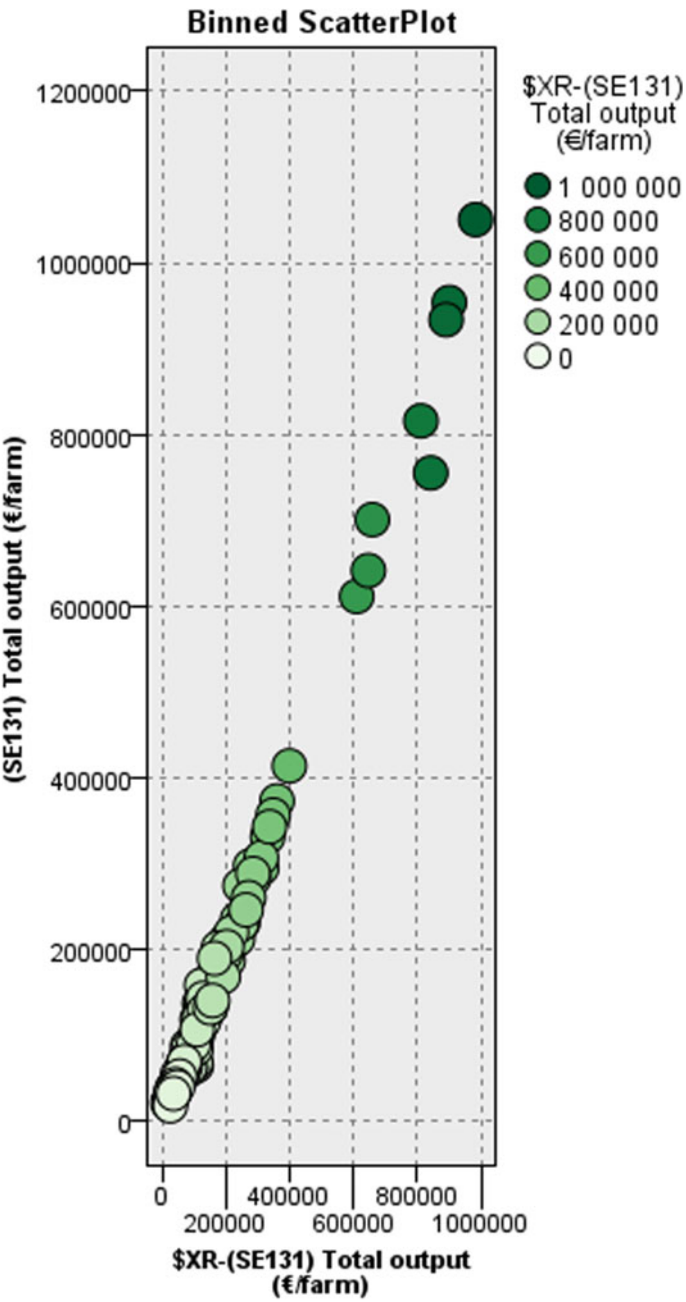

Fig. 1.3 Relationships between the observed values and the predicted ones for the agricultural output of the European Union farming regions, with data at the farm level, for the year 2020

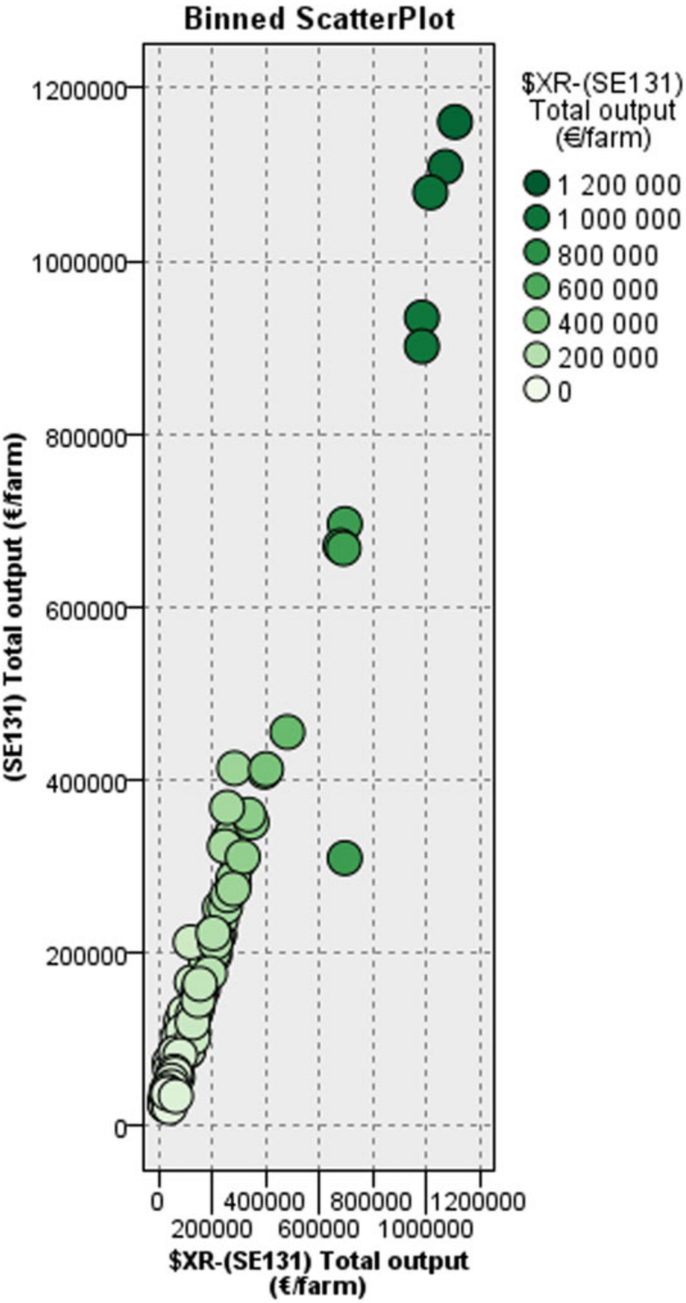

Fig. 1.4 Relationships between the observed values and the predicted ones for the agricultural output of the European Union farming regions, with data at the farm level, for the year 2021

Table 1.4 Importance of the predictors for the agricultural output of the European Union farming regions, with data at the farm level, for the year 2018

Nodes	Importance
Net worth (€)	0.0072
Milk yield cattle dairy cows (kg/cow)	0.0073
Gross farm income (€)	0.0073
Milk yield (kg/cow)	0.0077
Yield of maize (q/ha)	0.0078
Industrial crops (€/farm)	0.0079
Feed for grazing livestock (€)	0.0087
Total output livestock and livestock products (€/farm)	0.0090
Total inputs (€)	0.0120
Economic size (€'000)	0.0244

Table 1.5 Models with the highest accuracy (the lowest relative error) for the agricultural output of the European Union farming regions, with data at the farm level, for the year 2019

Model	Build time	Correlation	No. fields	Relative error
Linear	1	1.000	1	0.000
CHAID	1	1.000	1	0.000
Neural net	1	1.000	171	0.001
Random forest	1	0.993	178	0.017
Random trees	1	0.980	178	0.045

Table 1.6 Importance of the predictors for the agricultural output of the European Union farming regions, with data at the farm level, for the year 2019

Nodes	Importance
Total assets, opening valuation (€)	0.0074
Other livestock specific costs (incl. veterinary expenses) (€/farm)	0.0080
Specific crop costs (€/ha)	0.0083
Vegetables and flowers (€/farm)	0.0083
Contract work (€)	0.0087
Other crop specific costs (€)	0.0095
Forestry specific costs (€)	0.0097
Total output livestock and livestock products (€/farm)	0.0116
Milk yield (kg/cow)	0.0285
Yield of wheat (q/ha)	0.0456

Table 1.7 Models with the highest accuracy (the lowest relative error) for the agricultural output of the European Union farming regions, with data at the farm level, for the year 2020

Model	Build time	Correlation	No. fields	Relative error
Linear	1	1.000	11	0.000
Neural net	1	1.000	169	0.001
C&R tree	1	0.993	32	0.015
Random forest	1	0.994	178	0.020
Random trees	1	0.984	178	0.039

Table 1.8 Importance of the predictors for the agricultural output of the European Union farming regions, with data at the farm level, for the year 2020

Nodes	Importance
Breeding livestock (€)	0.0070
Dairy cows (LU)	0.0070
Total intermediate consumption (€)	0.0070
Other crop specific costs (€)	0.0071
Total assets, opening valuation (€)	0.0072
Other rural development payments (€)	0.0074
Yield of wheat (q/ha)	0.0074
Total output crops and crop production (€/farm)	0.0079
Total output livestock and livestock products (€/farm)	0.0083
Total inputs (€)	0.0095

Table 1.9 Models with the highest accuracy (the lowest relative error) for the agricultural output of the European Union farming regions, with data at the farm level, for the year 2021

Model	Build time	Correlation	No. fields	Relative error
CHAID	< 1	1.000	3	0.000
Linear	< 1	1.000	12	0.000
Neural net	< 1	1.000	169	0.001
Random forest	< 1	0.994	178	0.012
Random trees	< 1	0.989	178	0.027

Table 1.10 Importance of the predictors for the agricultural output of the European Union farming regions, with data at the farm level, for the year 2021

Nodes	Importance
Forestry and wood processing (€)	0.0077
Vegetables and flowers (ha)	0.0077
Seeds and plants home-grown (€)	0.0077
Intangible assets (€/farm)	0.0078
Other crop specific costs (€)	0.0083
Other crop output (€/farm)	0.0083
Other rural development payments (€)	0.0084
Vegetables and flowers (€/farm)	0.0091
Total inputs (€)	0.0104
Yield of wheat (q/ha)	0.0107

1.4 Discussion and Conclusions

Predicting the total output of the European Union farms is fundamental to support the decisions of the farmers, the design of policy instruments by the policymakers and the implementation of plans adopted by the governments. The new technologies associated with the digital era may contribute significantly to these frameworks, namely the solutions related to the machine learning approaches. From this perspective, this study aimed to apply the new methodologies from era 4.0 to identify the more accurate models and the most important indicators to predict the total output of the European Union farms. For that, the procedures proposed by the software IBM SPSS Modeler were followed and statistical information from the Farm Accountancy Data Network database was considered for the period 2018–2021. The information available in this database is microeconomic data for representative farms of each country and agricultural region, for example.

The literature review about these topics highlights the diversity of factors that impact agricultural output and the importance of artificial intelligence in dealing with these particularities of agriculture, particularly in more vulnerable contexts in terms of food security. The new technologies open new opportunities for assessment approaches and to collect information through alternative solutions. The current challenges of agriculture claim new methodologies that promote a more sustainable development, specifically to make compatible the environmental and economic dimensions.

The data analysis shows the difficulties in predicting farming indicators over different years, because of the vulnerability of the sector to internal (biological conditions of some crucial production factors, such as plants and animals) and external factors (market dynamics and climate, for example) that affect significantly the agricultural output. In any case, the countries with contexts of higher values for the total output are, for instance, Belgium, Denmark, Germany, Netherlands and Slovakia. Croatia, Greece, Poland and Romania are between the countries with examples of lower revenues for the farmers.

The results, from the application of machine learning approaches to identify the most accurate models and the most important predictors, reveal that there are some similarities in the findings obtained for the four years considered (2018, 2019, 2020 and 2021), but there are also, in some circumstances, relevant differences in the outcomes identified, claiming for more research in these fields to bring more insights for the stakeholders.

In terms of practical implications, the findings of this study highlight the importance of the new technologies to predict farming indicators in contexts of some annual variability in the variables of the farms. For policy recommendation, it is suggested to implement instruments and measures that promote the application of these digital approaches in the sector to better support farmers, policymakers and public institutions. For future research, it would be interesting to consider the effect of the time in these assessments, with panel data econometric methodologies, for example.

Acknowledgements This work is funded by National Funds through the FCT—Foundation for Science and Technology, I.P., within the scope of the project Refª UIDB/00681/2020 (https://doi.org/10.54499/UIDP/00681/2020). This research is also funded by the Enovo company. This study was carried out under the international project "Agriculture 4.0: Current reality, potentialities and policy proposals" (CERNAS-IPV/2022/008). Furthermore, we would like to thank the CERNAS Research Centre and the Polytechnic Institute of Viseu for their support. This work is too co-financed by the PRR—Plano de Recuperação e Resiliência (República Portuguesa) and the European Next Generation EU Funds (https://recuperarportugal.gov.pt) through application PRR-C05-i03-I-000030—"Carb2Soil—Reforçar a Complementaridade entre agricultura e pecuária para aumentar a fertilidade dos solos e a sua capacidade de sequestro de carbono".

References

1. J. Rajamanickam, S.D. Mani, Kullback chi square and Gustafson Kessel probabilistic neural network based soil fertility prediction. Concurr. Comput. Pract. Exp. **33**, e6460 (2021)
2. A. Soltabayeva, A. Ongaltay, J.O. Omondi, S. Srivastava, Morphological, physiological and molecular markers for salt-stressed plants. Plants Basel **10**, 243 (2021)
3. X. Xu, C. Du, F. Ma, Z. Qiu, J. Zhou, A framework for high-resolution mapping of soil organic matter (SOM) by the integration of fourier mid-infrared attenuation total reflectance spectroscopy (FTIR-ATR), sentinel-2 images, and DEM derivatives. Remote Sens. **15**, 1072 (2023)
4. A. Aldrees, M.F. Javed, A.T.B. Taha, A.M. Mohamed, M. Jasinski, M. Gono, Evolutionary and ensemble machine learning predictive models for evaluation of water quality. J. Hydrol. Reg. Stud. **46**, 101331 (2023)
5. J. Burkhardt, N.W. Chan, B. Bollinger, K.T. Gillingham, Conformity and conservation: evidence from home landscaping and water conservation JEL codes. Am. J. Agr. Econ. **104**, 228 (2022)
6. S. Li, M. Abdelkareem, N. Al-Arifi, Mapping groundwater prospective areas using remote sensing and GIS-based data driven frequency ratio techniques and detecting land cover changes in the Yellow River Basin. China. Land **12**, 771 (2023)
7. H. Bedair et al., Global warming status in the African continent: sources, challenges, policies, and future direction. Int. J. Environ. Res. **17**, 45 (2023)
8. C.-H. Kuan, Y. Leu, C.-P. Lee, A strategy of assessing climate factors influence for agriculture output. KSII Trans. Internet Inf. Syst. **16**, 1414 (2022)

9. C.-H. Kuan, Y. Leu, W.-S. Lin, C.-P. Lee, The estimation of the long-term agricultural output with a robust machine learning prediction model. Agricult. Basel **12**, 1075 (2022)
10. M. Qayyum, Y. Zhang, M. Wang, Y. Yu, S. Li, W. Ahmad, S.N. Maodaa, S.R.M. Sayed, J. Gan, Advancements in technology and innovation for sustainable agriculture: understanding and mitigating greenhouse gas emissions from agricultural soils. J. Environ. Manag. **347**, 119147 (2023)
11. M. Bhagat, D. Kumar, Efficient feature selection using BoWs and SURF method for leaf disease identification. Multimed. Tools Appl. **82**, 28187 (2023)
12. M.K.I. Rahmani, H.M.A. Ghanimi, S.F. Jilani, M. Aslam, M. Alharbi, R. Alroobaea, S. Sengan, Early pathogen prediction in crops using nano biosensors and neural network-based feature extraction and classification. Big Data Res. **34**, 100412 (2023)
13. Z. Tarek, M. Elhoseny, M.I. Alghamdi, I.M. El-Hasnony, Leveraging three-tier deep learning model for environmental cleaner plants production. Sci. Rep. **13**, 19499 (2023)
14. E. Khudyakova, M. Nikanorov, I. Bystrenina, T. Cherevatova, I. Sycheva, Forecasting the production of gross output in agricultural sector of the Ryazan Oblast. Estud. Econ. Apl. **39**, 5171 (2021)
15. R. Chew, J. Rineer, R. Beach, M. O'Neil, N. Ujeneza, D. Lapidus, T. Miano, M. Hegarty-Craver, J. Polly, D.S. Temple, Deep neural networks and transfer learning for food crop identification in UAV images. Drones-Basel **4**, 7 (2020)
16. S. Levi, Living standards shape individual attitudes on genetically modified food around the world. Food Qual. Prefer. **95**, 104371 (2022)
17. FADN, *Several Statistics*. https://agriculture.ec.europa.eu/data-and-analysis/farm-structures-and-economics/fadn_en
18. IBM SPSS Modeler, *SPSS Modeler: Overview*. https://www.ibm.com/products/spss-modeler

Chapter 2
Applying Artificial Intelligence to Predict Crop Output

Abstract The agricultural output has several parts, and depending on the characteristics of the farms, one of these parcels is related to crop production. Including in the crop output, the sources of these incomes are diverse. In any case, crop production has a fundamental role in the sustainability of the farms and society, as a source of income for the farmers and food for the population. In this context, it is important to understand the main factors that may support the stakeholders in predicting the crop output in the European Union farms. The main objective of this research is to identify the most adjusted models and the most important variables to predict crop income in the European Union context. For that, data from the Farm Accountancy Data Network were considered, as well as approaches associated with artificial intelligence. The main findings provide relevant insights and knowledge, namely for farmers and policymakers that may be considered in the processes of agricultural planning, management and policy design.

Keywords Machine learning · European Union databases · Agricultural sector

2.1 Introduction

Several studies have considered artificial intelligence for predictions about crop output and related dimensions worldwide. Some of them focus on the following issues: rainfall prediction in Ethiopian context [1]; organic potato yield estimation [2]; predictors of potato yield [3]; increase in agricultural yields [4]; soil properties analysis in arid regions [5]; greenhouse production with remote control [6]; pests classification [7]; diseases identification [8]; environmental impact analysis of productions (tomato and cucumber) in greenhouses [9]; plant factory requirements [10]; environmental impacts assessment [11]; intelligent greenhouses [12]; decisions on water, land and food nexus [13] and energy usage estimation [14].

Artificial intelligence allows the implementation of smart farming approaches, where the machine and deep learning techniques have their relevance. Machine learning methodologies are considered for crop selection and management and deep

V. J. P. D. Martinho, *Machine Learning Approaches for Evaluating Statistical Information in the Agricultural Sector*, SpringerBriefs in Applied Sciences and Technology, https://doi.org/10.1007/978-3-031-54608-2_2

learning technologies are taken into account for crop production prediction [15]. In these frameworks of smart farming, the Internet of Things (IoT) systems and Big Data are crucial to improve the accuracy of the crop output assessment [16]. Artificial intelligence and IoT approaches have been used, for instance, to identify fruit diseases [17], or fertiliser recommendations [18].

The several dimensions associated with artificial intelligence may contribute to a more sustainable agricultural sector, following the international deals for greener farming production, such as defined by the European Union in the Green Deal strategy, for example [19]. The weed management without, or reduced application, of herbicides is an illustration of the digital approaches contribute to more sustainable agriculture [20].

The Climate-Smart Agriculture concept is more a case where the new technologies associated with the digital approaches may contribute to improve the efficiency of the farms, with possibilities of increasing agricultural production and food security in a way compatible with the efforts to preserve the environment [21].

The digital transition and other innovative technologies and practices bring new potentialities for agriculture [22]. Between these innovations appear, for instance, the new approaches related to vertical farming and biotechnology.

Taking into account these frameworks, this research aims to bring more insights into the application of artificial intelligence to predict crop output in the European Union farms, considering data from the Farm Accountancy Data Network [23] and following the procedures suggested by the software IBM SPSS Modeler [24].

2.2 Data Assessment

If the pandemic had an impact on the farming indicators, the year 2021 was a period of some recovery in the dynamics of the European Union farms. This is visible in Table 2.1 for the average values, for example, of the crop output [total output crops and crop production (€/farm)]. However, the effect of prices on these results should be noted. In any case, the aim here is to bring more insights into the dynamics of the farms over different years.

The normalised values (considering $(x_i - x_{\text{minimum}})/(x_{\text{maximum}} - x_{\text{minimum}})$) presented in Table 2.2 reveal the importance of the output of the crops in the farms from Czechia, Denmark, Netherlands, Slovakia and some regions of France and Germany. The lowest values for the variable total output crops and crop production (€/farm) appear in Ireland and some regions, for example, from Greece, Poland, Portugal, Romania and Spain.

Table 2.1 Growth rate (%) results for the crop output of the European Union countries, with data at the farm level, over the period 2019–2021

Member state	Year	
	2020	2021
Austria	6.233	20.825
Belgium	8.770	7.098
Bulgaria	– 3.753	**45.241**
Croatia	4.228	22.384
Cyprus	*– 4.600*	*6.332*
Czechia	**25.992**	16.057
Denmark	**16.735**	9.338
Estonia	1.083	*5.407*
Finland	2.344	18.294
France	*– 4.931*	19.622
Germany	1.175	18.992
Greece	0.465	17.002
Hungary	9.093	18.396
Ireland	*– 6.250*	**23.296**
Italy	2.503	7.995
Latvia	**13.057**	*– 0.229*
Lithuania	**20.948**	*4.772*
Luxembourg	**14.343**	15.570
Netherlands	3.282	6.897
Poland	4.466	**28.602**
Portugal	*– 13.139*	16.017
Romania	*– 12.824*	**42.433**
Slovakia	0.428	22.628
Slovenia	– 2.164	*– 0.052*
Spain	5.096	7.413
Sweden	5.151	**35.072**
Average	3.759	16.746

Note Bold corresponds to the highest values and italic to the lowest

Table 2.2 Normalised values for the crop output of the European Union farming regions, with data at the farm level, over the period 2019–2021

Member state	Region	Year		
		2019	2020	2021
Austria	Austria	0.056	0.056	0.051
Belgium	Vlaanderen	0.295	0.308	0.243
Belgium	Wallonie	0.171	0.174	0.138
Bulgaria	Severen tsentralen	0.198	0.174	0.187
Bulgaria	Severoiztochen	0.217	0.148	0.214
Bulgaria	Severozapaden	0.206	0.235	0.217
Bulgaria	Yugoiztochen	0.112	0.094	0.120
Bulgaria	Yugozapaden	0.043	0.043	0.031
Bulgaria	Yuzhen tsentralen	0.048	0.051	0.042
Croatia	Jadranska Hrvatska	0.028	0.024	*0.020*
Croatia	Kontinentalna Hrvatska	0.029	0.030	0.029
Cyprus	Cyprus	0.034	0.030	0.023
Czechia	Czechia	0.362	**0.434**	**0.373**
Denmark	Denmark	0.437	**0.484**	**0.392**
Estonia	Estonia	0.171	0.164	0.127
Finland	Etelä-Suomi	0.142	0.138	0.118
Finland	Pohjanmaa	0.157	0.144	0.139
Finland	Pohjois-Suomi	0.123	0.134	0.109
Finland	Sisä-Suomi	0.105	0.101	0.090
France	Alsace	0.286	0.260	0.196
France	Aquitaine	0.274	0.234	0.192
France	Auvergne	0.049	0.042	0.043
France	Basse-Normandie	0.129	0.112	0.115
France	Bourgogne	0.342	0.327	0.309
France	Bretagne	0.191	0.180	0.155
France	Centre	0.352	0.316	0.325
France	Champagne-Ardenne	0.416	0.339	0.301
France	Corse	0.208	0.183	0.157
France	Franche-Comté	0.132	0.108	0.095
France	Guadeloupe	0.103	0.093	0.062
France	Haute-Normandie	**0.452**	0.358	0.337
France	Île-de-France	**0.566**	**0.499**	**0.476**
France	La Réunion	0.166	0.159	0.118
France	Languedoc-Roussillon	0.269	0.277	0.201
France	Limousin	0.045	0.033	0.029

(continued)

Table 2.2 (continued)

Member state	Region	Year		
		2019	2020	2021
France	Lorraine	0.176	0.164	0.171
France	Midi-Pyrénées	0.147	0.130	0.121
France	Nord-Pas-de-Calais	0.345	0.324	0.271
France	Pays de la Loire	0.237	0.188	0.190
France	Picardie	**0.495**	0.383	0.365
France	Poitou–Charentes	0.371	0.376	0.348
France	Provence-Alpes-Côte d'Azur	0.405	0.397	0.287
France	Rhône-Alpes	0.190	0.182	0.150
Germany	Baden-Württemberg	0.178	0.147	0.125
Germany	Bayern	0.134	0.128	0.112
Germany	Brandenburg	**0.952**	**0.946**	**0.815**
Germany	Hessen	0.161	0.148	0.129
Germany	Mecklenburg-Vorpommern	**1.000**	**1.000**	**1.000**
Germany	Niedersachsen	0.247	0.238	0.199
Germany	Nordrhein-Westfalen	0.224	0.226	0.217
Germany	Rheinland-Pfalz	0.320	0.304	0.248
Germany	Saarland	0.120	0.118	0.105
Germany	Sachsen	**0.674**	**0.725**	**0.571**
Germany	Sachsen-Anhalt	**0.807**	**0.813**	**0.782**
Germany	Schleswig–Holstein/Hamburg	0.307	0.238	0.225
Germany	Thüringen	**0.962**	**0.991**	**0.873**
Greece	Ipiros-Peloponissos-Nissi Ioniou	0.026	*0.022*	*0.020*
Greece	Makedonia-Thraki	0.037	0.034	0.029
Greece	Sterea Ellas-Nissi Egaeou-Kriti	*0.022*	0.023	0.021
Greece	Thessalia	0.031	0.029	0.025
Hungary	Alföld	0.097	0.108	0.092
Hungary	Dunántúl	0.136	0.132	0.120
Hungary	Észak-Magyarország	0.122	0.097	0.095
Ireland	Ireland	*0.022*	*0.019*	*0.018*
Italy	Abruzzo	0.066	0.062	0.051
Italy	Alto Adige	0.087	0.103	0.072
Italy	Basilicata	0.082	0.071	0.052
Italy	Calabria	0.054	0.040	0.047
Italy	Campania	0.081	0.082	0.062

(continued)

Table 2.2 (continued)

Member state	Region	Year		
		2019	2020	2021
Italy	Emilia-Romagna	0.152	0.141	0.114
Italy	Friuli-Venezia Giulia	0.142	0.116	0.105
Italy	Lazio	0.102	0.108	0.072
Italy	Liguria	0.112	0.107	0.094
Italy	Lombardia	0.124	0.134	0.116
Italy	Marche	0.071	0.080	0.061
Italy	Molise	0.054	0.050	0.045
Italy	Piemonte	0.138	0.138	0.111
Italy	Puglia	0.084	0.076	0.063
Italy	Sardegna	0.043	0.048	0.035
Italy	Sicilia	0.065	0.062	0.051
Italy	Toscana	0.131	0.135	0.084
Italy	Trentino	0.093	0.094	0.078
Italy	Umbria	0.079	0.075	0.055
Italy	Valle d'Aosta	0.067	0.059	0.046
Italy	Veneto	0.164	0.153	0.122
Latvia	Latvia	0.085	0.092	0.067
Lithuania	Lithuania	0.061	0.071	0.055
Luxembourg	Luxembourg	0.086	0.094	0.080
Netherlands	The Netherlands	**0.603**	**0.590**	**0.467**
Poland	Malopolska i Pogórze	*0.022*	*0.019*	0.021
Poland	Mazowsze i Podlasie	*0.021*	*0.021*	0.021
Poland	Pomorze i Mazury	0.062	0.061	0.058
Poland	Wielkopolska and Slask	0.050	0.052	0.049
Portugal	Açores e Madeira	*0.012*	*0.014*	*0.008*
Portugal	Alentejo e Algarve	0.083	0.043	0.054
Portugal	Norte e Centro	0.030	0.030	0.022
Portugal	Ribatejo e Oeste	0.080	0.071	0.060
Romania	Bucuresti-Ilfov	0.114	0.062	0.045
Romania	Centru	0.027	0.025	0.025
Romania	Nord-Est	0.032	0.024	0.032
Romania	Nord-Vest	*0.021*	*0.021*	*0.017*
Romania	Sud-Est	0.068	0.041	0.059
Romania	Sud-Muntenia	0.056	0.040	0.049

(continued)

Table 2.2 (continued)

Member state	Region	Year		
		2019	2020	2021
Romania	Sud-Vest-Oltenia	*0.021*	*0.014*	*0.017*
Romania	Vest	0.043	0.047	0.040
Slovakia	Slovakia	**0.690**	**0.657**	**0.597**
Slovenia	Slovenia	0.027	0.024	*0.017*
Spain	Andalucía	0.159	0.137	0.107
Spain	Aragón	0.125	0.155	0.126
Spain	Asturias	*0.021*	*0.018*	*0.017*
Spain	Canarias	0.241	0.222	0.182
Spain	Cantabria	*0.000*	*0.000*	*0.000*
Spain	Castilla y León	0.091	0.100	0.073
Spain	Castilla-La Mancha	0.103	0.105	0.088
Spain	Cataluña	0.127	0.127	0.102
Spain	Comunidad Valenciana	0.082	0.101	0.085
Spain	Extremadura	0.075	0.077	0.057
Spain	Galicia	*0.025*	*0.020*	*0.017*
Spain	Islas Baleares	0.066	0.080	0.061
Spain	La Rioja	0.161	0.237	0.193
Spain	Madrid	0.054	0.052	0.053
Spain	Murcia	0.191	0.217	0.167
Spain	Navarra	0.123	0.113	0.099
Spain	País Vasco	0.096	0.080	0.067
Sweden	Län i norra Sverige	0.105	0.078	0.089
Sweden	Skogsoch mellanbygdslän	0.120	0.113	0.109
Sweden	Slättbyggdslän	0.270	0.275	0.272

Note Bold corresponds to the highest values and italic to the lowest

2.3 Main Findings

The findings presented in Tables 2.3, 2.5 and 2.7 reveal that the models with the highest accuracy for the training set are, with minor differences, the same for the three years taken into account (2019, 2020 and 2021). The linear, regression, CHAID, random forest, neural net and random trees are the most accurate models.

Table 2.3 Models with the highest accuracy (the lowest relative error) for the crop output of the European Union farming regions, with data at the farm level, for the year 2019

Model	Build time	Correlation	No. fields	Relative error
Linear	1	1.000	1	0.000
CHAID	1	1.000	1	0.000
Neural net	1	0.998	171	0.004
Random forest	1	0.985	178	0.030
Random trees	1	0.969	178	0.066

Table 2.4 Importance of the predictors for the crop output of the European Union farming regions, with data at the farm level, for the year 2019

Nodes	Importance
Farmhouse consumption (€)	0.0072
Farm net value added (€)	0.0072
Gross farm income (€)	0.0074
Vegetables and flowers (ha)	0.0078
Forestry and wood processing (€)	0.0083
Specific crop costs (€/ha)	0.0085
Seeds and plants (€)	0.0092
Vegetables and flowers (€/farm)	0.0093
Milk yield cattle dairy cows (kg/cow)	0.0328
Yield of wheat (q/ha)	0.0555

Table 2.5 Models with the highest accuracy (the lowest relative error) for the crop output of the European Union farming regions, with data at the farm level, for the year 2020

Model	Build time	Correlation	No. fields	Relative error
CHAID	< 1	1.000	3	0.000
Regression	< 1	1.000	60	0.000
Linear	< 1	1.000	14	0.000
Neural net	< 1	1.000	169	0.001
Random trees	< 1	0.981	178	0.047

Table 2.6 Importance of the predictors for the crop output of the European Union farming regions, with data at the farm level, for the year 2020

Nodes	Importance
Forestry and wood processing (€)	0.0087
Total crops output (€/ha)	0.0088
Specific crop costs (€/ha)	0.0092
Pigs (LU)	0.0099
Unpaid labour input (hrs)	0.0108
Forest land including standing timber (€/farm)	0.0115
LFA subsidies (€)	0.0117
Vineyards (ha)	0.0137
Intangible assets (€/farm)	0.0144
Energy crops (ha)	0.0152

Table 2.7 Models with the highest accuracy (the lowest relative error) for the crop output of the European Union farming regions, with data at the farm level, for the year 2021

Model	Build time	Correlation	No. fields	Relative error
CHAID	< 1	1.000	3	0.000
Linear	< 1	1.000	18	0.000
Neural net	< 1	1.000	169	0.001
Random forest	< 1	0.987	178	0.027
Random trees	< 1	0.985	178	0.034

Table 2.8 Importance of the predictors for the crop output of the European Union farming regions, with data at the farm level, for the year 2021

Nodes	Importance
Vegetables and flowers (ha)	0.0064
Labour input (h)	0.0064
Poultry (LU)	0.0064
Other output (€/farm)	0.0065
Agritourism (€)	0.0068
Unpaid labour input (AWU)	0.0073
Cereals (ha)	0.0082
Yield of wheat (q/ha)	0.0111
Fertiliser N (q)	0.0156
Seeds and plants (€)	0.0248

The accuracy of these models, for the total output crops and crop production (€/farm), is highlighted in the relationships among the observed values and the predicted ones shown in Figs. 2.1, 2.2 and 2.3.

In general, the most important predictors are different for the three years considered (Tables 2.4, 2.6 and 2.8), confirming the challenges of predicting farming indicators with internal variables. Nonetheless, the specific crop costs and variables associated with crop productivity may be taken into account to support the stakeholders in the total output crops and crop production (€/farm) prediction.

2.4 Discussion and Conclusions

Crop production is crucial for the sustainability of the agricultural sector and plays a fundamental role in food security worldwide. In this way, the prediction of the crop output is important, namely in the current context of increased demand for food, because of the growth of the world's inhabitants. This prediction is also essential to support the design and implementation of adjusted practices, namely by the national and international decision-makers and, in this perspective, to better deal with the environmental problems that challenge presently the farming sector. Considering these motivations, this research intended to suggest models with high accuracy and important variables to predict the crop output in the European Union countries. To achieve these objectives, microeconomic data, at the farm level, were considered from the Farm Accountancy Data Network for the period 2019–2021. This statistical information was analysed through artificial intelligence approaches, following the procedures proposed by the software IBM SPSS Modeler.

The consideration of artificial intelligence to predict crop output worldwide is already highlighted in scientific documents, namely for yield estimations and assessments of variables that may affect the crop production, such as soil, water, energy, fertilisers, diseases and environmental changes. The Climate-Smart Agriculture concept is an example recognised internationally where the digital approaches may, indeed, bring new opportunities for the agricultural sector and the world context related to agriculture.

The data analysis reveals the impacts of the pandemic on the dynamics of the European Union farming sector, with some recovery in 2021. In any case, the crop productions have importance in countries such as Czechia, Denmark, Netherlands, Slovakia, France and Germany and less pertinence in frameworks from Ireland, Greece, Poland, Portugal, Romania and Spain, for example.

The findings obtained with the application of artificial intelligence approaches show the importance of models such as linear, regression, CHAID, random forest, neural net and random trees to predict the crop output in the European Union contexts. These results highlight the relevance of predictors such as specific crop costs and variables associated with crop productivity.

In terms of practical implications, predictive models of the crop output such as linear, regression, CHAID, random forest, neural net and random trees may provide

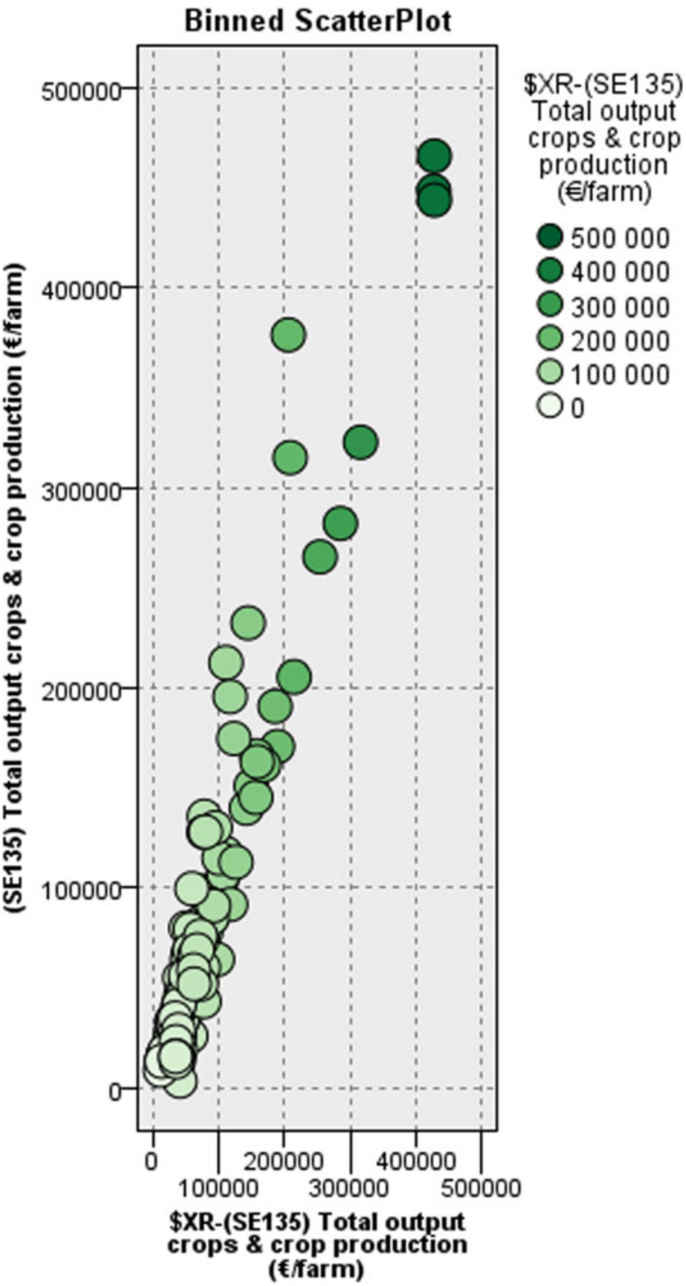

Fig. 2.1 Relationships between the observed values and the predicted ones for the crop output of the European Union farming regions, with data at the farm level, for the year 2019

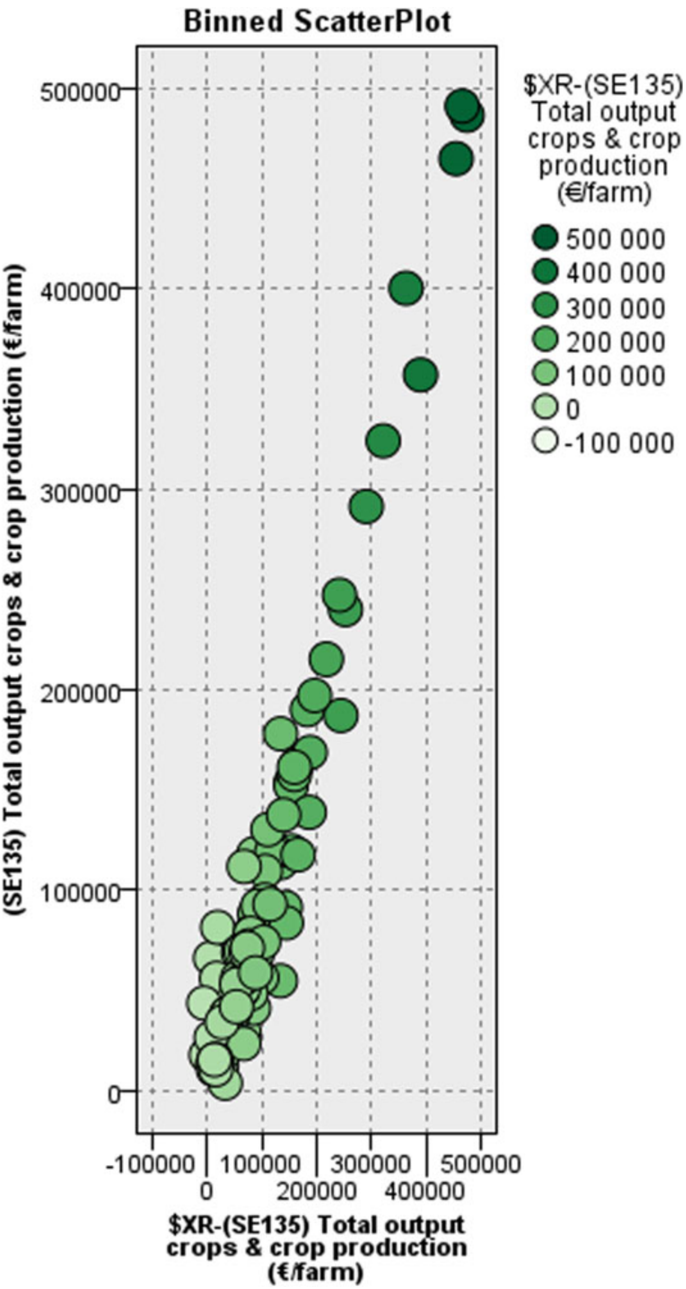

Fig. 2.2 Relationships between the observed values and the predicted ones for the crop output of the European Union farming regions, with data at the farm level, for the year 2020

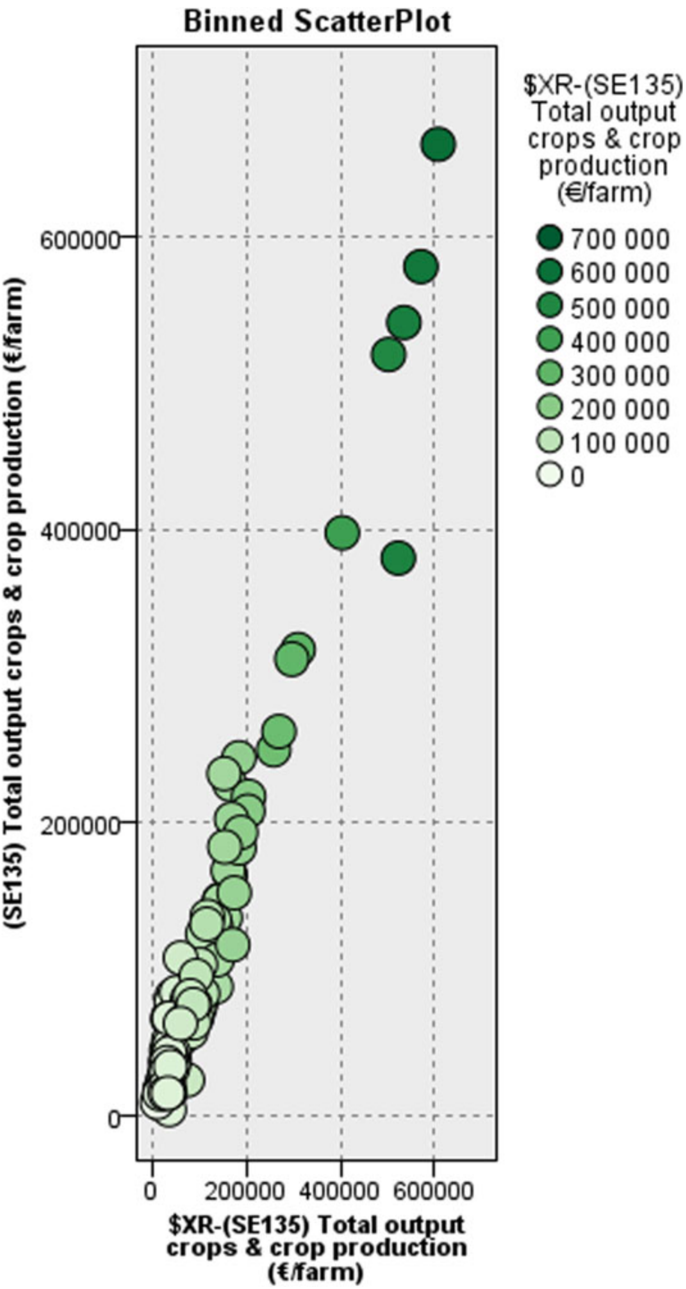

Fig. 2.3 Relationships between the observed values and the predicted ones for the crop output of the European Union farming regions, with data at the farm level, for the year 2021

relevant insights for the stakeholders related to the European Union contexts. Specifically, when variables associated with specific costs and crop productivity will be considered. For policy recommendations, it is suggested to identify practices and approaches that allow improvement of the crop output and productivity with fewer resources. For future research, it could be interesting to consider other databases and other approaches to compare with the findings identified in this research.

Acknowledgements This work is funded by National Funds through the FCT—Foundation for Science and Technology, I.P., within the scope of the project Refª UIDB/00681/2020 (https://doi.org/https://doi.org/10.54499/UIDP/00681/2020). This research is also funded by the Enovo company. This study was carried out under the international project "Agriculture 4.0: Current reality, potentialities and policy proposals" (CERNAS-IPV/2022/008). Furthermore, we would like to thank the CERNAS Research Centre and the Polytechnic Institute of Viseu for their support. This work is too co-financed by the PRR—Plano de Recuperação e Resiliência (República Portuguesa) and the European Next Generation EU Funds (https://recuperarportugal.gov.pt) through application PRR-C05-i03-I-000030—"Carb2Soil—Reforçar a Complementaridade entre agricultura e pecuária para aumentar a fertilidade dos solos e a sua capacidade de sequestro de carbono".

References

1. W.T. Abebe, D. Endalie, Artificial intelligence models for prediction of monthly rainfall without climatic data for meteorological stations in Ethiopia. J. Big Data **10**, 2 (2023)
2. K. Abrougui, K. Gabsi, B. Mercatoris, C. Khemis, R. Amami, S. Chehaibi, Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). Soil Tillage Res. **190**, 202 (2019)
3. Y. Lin, S. Li, S. Duan, Y. Ye, B. Li, G. Li, D. Lyv, L. Jin, C. Bian, J. Liu, Methodological evolution of potato yield prediction: a comprehensive review. Front. Plant Sci. **14**, 1214006 (2023)
4. K. Ennouri, S. Smaoui, Y. Gharbi, M. Cheffi, O.B. Braiek, M. Ennouri, M.A. Triki, Usage of artificial intelligence and remote sensing as efficient devices to increase agricultural system yields. J. Food Qual. **2021**, 6242288 (2021)
5. M.A. El-Sayed, A.H. Abd-Elazem, A.R.A. Moursy, E.S. Mohamed, D.E. Kucher, M.E. Fadl, Integration vis-NIR spectroscopy and artificial intelligence to predict some soil parameters in arid region: a case study of Wadi Elkobaneyya, South Egypt. Agron. Basel **13**, 935 (2023)
6. S. Hemming, F. de Zwart, A. Elings, I. Righini, A. Petropoulou, Remote control of greenhouse vegetable production with artificial intelligence greenhouse climate, irrigation, and crop production. Sensors **19**, 1807 (2019)
7. M.E. Karar, F. Alsunaydi, S. Albusaymi, S. Alotaibi, A new mobile application of agricultural pests recognition using deep learning in cloud computing system. Alex. Eng. J. **60**, 4423 (2021)
8. D.-E.A. Sanchez, A.A. Rodriguez, I.F.B. Martinez, C.S. Gonzalez, C.B. Garcia, Use of trained convolutional neural networks for analysis of symptoms caused by botrytis Fabae Sard. Rev. Cienc. Agric. **40**, e1198 (2023)
9. B. Khoshnevisan, S. Rafiee, M. Omid, H. Mousazadeh, S. Clark, Environmental impact assessment of tomato and cucumber cultivation in greenhouses using life cycle assessment and adaptive neuro-fuzzy inference system. J. Clean. Prod. **73**, 183 (2014)
10. Y. Liu, S. Mousavi, Z. Pang, Z. Ni, M. Karlsson, S. Gong, Plant factory: a new playground of industrial communication and computing. Sensors **22**, 147 (2022)
11. A. Nabavi-Pelesaraei, S. Rafiee, S.S. Mohtasebi, H. Hosseinzadeh-Bandbafha, K. Chau, Integration of artificial intelligence methods and life cycle assessment to predict energy output and environmental impacts of paddy production. Sci. Total. Environ. **631–632**, 1279 (2018)

12. A.S. Petropoulou, B. van Marrewijk, F. de Zwart, A. Elings, M. Bijlaard, T. van Daalen, G. Jansen, S. Hemming, Lettuce production in intelligent greenhouses-3D imaging and computer vision for plant spacing decisions. Sensors **23**, 2929 (2023)
13. W. Shao, Y. Ding, J. Wen, P. Zhu, L. Ou, Optimal decision-making in the water, land and food nexus using artificial intelligence and extreme machine learning. Water Supply **23**, 4166 (2023)
14. S. Venkatesan, J. Lim, H. Ko, Y. Cho, A machine learning based model for energy usage peak prediction in smart farms. Electronics **11**, 218 (2022)
15. Y. Akkem, S.K. Biswas, A. Varanasi, Smart farming using artificial intelligence: a review. Eng. Appl. Artif. Intell. **120**, 105899 (2023)
16. T. Alahmad, M. Nemenyi, A. Nyeki, Applying IoT sensors and big data to improve precision crop production: a review. Agron. Basel **13**, 2603 (2023)
17. S.K. Kabilesh, D. Mohanapriya, P. Suseendhar, J. Indra, T. Gunasekar, N. Senthilvel, Research on artificial intelligence based fruit disease identification system (AI-FDIS) with the internet of things (IoT). J. Intell. Fuzzy Syst. **44**, 6593 (2023)
18. B. Swaminathan, S. Palani, S. Vairavasundaram, K. Kotecha, V. Kumar, IoT-driven artificial intelligence technique for fertilizer recommendation model. IEEE Consum. Electron. Mag. **12**, 109 (2023)
19. V. Balaska, Z. Adamidou, Z. Vryzas, A. Gasteratos, Sustainable crop protection via robotics and artificial intelligence solutions. Machines **11**, 774 (2023)
20. D.J. Bloomer, K.C. Harrington, H. Ghanizadeh, T.K. James, Robots and shocks: emerging non-herbicide weed control options for vegetable and arable cropping. N. Z. J. Agric. Res. **67**, 81–103 (2023)
21. V.J.P.D. Martinho, R.P.F. Guine, Integrated-smart agriculture: contexts and assumptions for a broader concept. Agron. Basel **11**, 1568 (2021)
22. K. Hardy, T. Orridge, X. Heynes, S. Gunasena, S. Grundy, C. Lu, Farming the future: contemporary innovations enhancing sustainability in the agri-sector. Ann. Plant Rev. **4**, 263 (2021)
23. FADN, *Several Statistics*. https://agriculture.ec.europa.eu/data-and-analysis/farm-structures-and-economics/fadn_en
24. IBM SPSS Modeler, *SPSS Modeler: Overview*. https://www.ibm.com/products/spss-modeler

Chapter 3
Predictive Machine Learning Models for Livestock Output

Abstract Agricultural planning always had an important role in the performance of agriculture, but in our days this component of agricultural management seems to have an increased responsibility, because of the challenges imposed by the current contexts, specifically those related to the sustainability of the associated activities and processes. In fact, currently, it is important to reduce the environmental impacts of the farming dynamics and raise production to deal with the increased demand for food worldwide. The livestock activities are particularly complex and call for adjusted plans and management decisions. The new technologies associated with the digital transition may bring relevant added value, namely to predict outputs. This chapter aims to suggest models and predictors to support the farmers and other stakeholders to better design policies and farm plans. Statistical information from the European Union databases was considered. The results found are useful tools to improve the performance of the European Union farms, particularly those specialised in livestock production.

Keywords Accuracy · Artificial intelligence · Characteristics of farms in the European Union

3.1 Introduction

In the current contexts of opportunities open by the digital transition, namely to deal with Big Data and complex frameworks, the consideration of new approaches for agricultural planning and management has spread worldwide with enormous potentialities to improve the dynamics and performance of the farms, with benefits for the farmers (more income) and the populations (more food security).

For example, to forecast beef carcass weight, in the Brazilian context, the following approaches were considered [1]: generalised linear regression; random forests and multilayer neural networks. Of referring, in addition, the use of the following methods: multiple linear regression and random forest were used to predict

V. J. P. D. Martinho, *Machine Learning Approaches for Evaluating Statistical Information in the Agricultural Sector*, SpringerBriefs in Applied Sciences and Technology, https://doi.org/10.1007/978-3-031-54608-2_3

the herbage mass [2]; network-based fuzzy inference system and multilayer perceptron to estimate food production [3]; artificial neural network model to diagnose the incidence of intrauterine growth restriction in sheep [4]; random forest classifier to characterise of faecal microbiota in livestock activities [5]; random forest, support vector machine and naïve bayes classifier to estimate beef cattle grazing behaviours [6]. New methodologies were also considered to analyse the antimicrobial resistance in Chinese chicken farms [7].

For the application of new techniques, the collection of information is crucial and here the data obtained through images obtained with unmanned aerial vehicles (UAV) [8], or from other sources (Global Navigation Satellite System [9]; 3D digital images [10]; Sentinel-1 and Sentinel-2 [11]), may bring relevant added value for the assessments.

Applications of machine learning methodologies were still considered in other contexts related direct, or indirectly, with the livestock activities, such as the following: mapping the wooded vegetation in the Australian arid land [12]; identification of vegetation modifications and environmental impacts on rangeland farms [13]; analysing grazing and rumination periods using acoustic information in grazing cattle [14, 15]; determination of cattle and poultry manure properties [16]; land use estimates for insect meat activities [17]; classification of beef cattle producing municipalities [18]; prediction of beef cattle production [19]; prediction nitrogen fertilisation impacts on plant productivity [20]; classifying variables related to productivity in a silvopastoral context [21]; forecast temperature and humidity inside ventilated duck systems [22]; livestock performance prediction [23]; influenza virus analysis in domestic pigs [24]; identification of liquid manure use in Carolina [25] and sustainable animals characterisation [26].

The different perspectives presented before suggest the pertinence of identifying models with better accuracy to predict the livestock output in the European Union farms, considering statistical information from European databases with data at the farm level [27] and approaches related to the new technologies associated with the digital transition [28].

3.2 Data Evaluation

The data considered in this study were found on the Farm Accountancy Data Network database for the European Union agricultural regions and member states. The microeconomic statistical information available in this database is presented for the representative farms of these regions/countries.

Lithuania, Germany, Sweden, Ireland, Slovenia, Greece and Portugal are some of the European Union countries with the highest growth rates for livestock output [total output livestock and livestock products (€/farm)] between 2020 and 2021 (Fig. 3.1).

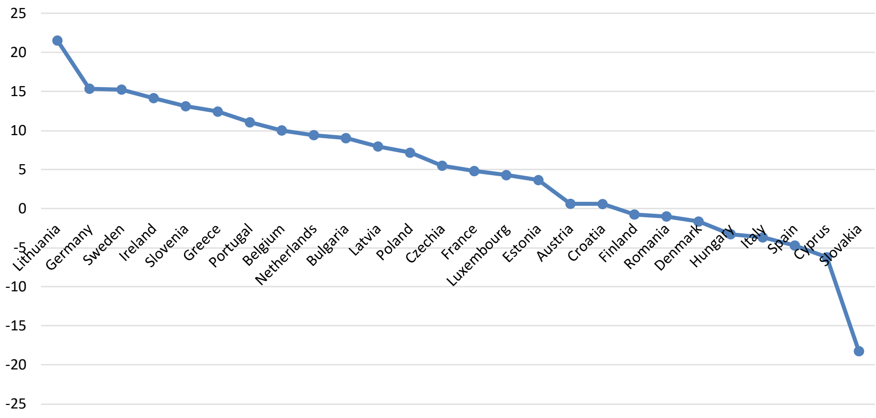

Fig. 3.1 Growth rate (%) results for the livestock output of the European Union countries, with data at the farm level, over the period 2020–2021

Table 3.1, with the normalised values $((x_i - x_{\text{minimum}})/(x_{\text{maximum}} - x_{\text{minimum}}))$, confirms the importance of the agricultural sector in countries, such as Denmark, Netherlands, Germany and France, for example. Greece, Italy, Poland, Portugal and Romania have the agricultural regions where the farms have the lowest values for the total output livestock and livestock products (€/farm).

Table 3.1 Normalised values for the livestock output of the European Union farming regions, with data at the farm level, over the period 2020–2021

Member state	Region	Year	
		2020	2021
Austria	Austria	0.135	0.132
Belgium	Vlaanderen	0.503	0.543
Belgium	Wallonie	0.283	0.290
Bulgaria	Severen tsentralen	0.048	0.070
Bulgaria	Severoiztochen	0.085	0.097
Bulgaria	Severozapaden	0.038	0.044
Bulgaria	Yugoiztochen	0.072	0.065
Bulgaria	Yugozapaden	0.030	0.024
Bulgaria	Yuzhen tsentralen	0.031	0.031
Croatia	Jadranska Hrvatska	0.014	0.013
Croatia	Kontinentalna Hrvatska	0.023	0.023
Cyprus	Cyprus	0.088	0.080

(continued)

Table 3.1 (continued)

Member state	Region	Year	
		2020	2021
Czechia	Czechia	0.367	0.376
Denmark	Denmark	**0.961**	**0.916**
Estonia	Estonia	0.145	0.146
Finland	Etelä-Suomi	0.112	0.113
Finland	Pohjanmaa	0.257	0.244
Finland	Pohjois-Suomi	0.298	0.296
Finland	Sisä-Suomi	0.253	0.231
France	Alsace	0.120	0.119
France	Aquitaine	0.142	0.140
France	Auvergne	0.243	0.246
France	Basse-Normandie	0.482	0.512
France	Bourgogne	0.146	0.159
France	Bretagne	**0.627**	**0.630**
France	Centre	0.109	0.111
France	Champagne-Ardenne	0.077	0.084
France	Corse	0.105	0.093
France	Franche-Comté	0.430	0.442
France	Guadeloupe	0.030	0.025
France	Haute-Normandie	0.230	0.206
France	Île-de-France	0.030	0.030
France	La Réunion	0.067	0.069
France	Languedoc-Roussillon	0.026	0.027
France	Limousin	0.220	0.223
France	Lorraine	0.329	0.319
France	Midi-Pyrénées	0.135	0.134
France	Nord-Pas-de-Calais	0.289	0.283
France	Pays de la Loire	**0.514**	0.542
France	Picardie	0.148	0.144
France	Poitou–Charentes	0.177	0.185
France	Provence-Alpes-Côte d'Azur	0.021	0.023
France	Rhône-Alpes	0.207	0.206
Germany	Baden-Württemberg	0.224	0.238
Germany	Bayern	0.256	0.288
Germany	Brandenburg	**0.872**	**1.000**
Germany	Hessen	0.243	0.285

(continued)

Table 3.1 (continued)

Member state	Region	Year	
		2020	2021
Germany	Mecklenburg-Vorpommern	**0.631**	**0.843**
Germany	Niedersachsen	**0.545**	**0.624**
Germany	Nordrhein-Westfalen	0.405	0.431
Germany	Rheinland-Pfalz	0.109	0.128
Germany	Saarland	0.212	0.254
Germany	Sachsen	**1.000**	**0.960**
Germany	Sachsen-Anhalt	**0.628**	**0.726**
Germany	Schleswig–Holstein/Hamburg	0.492	**0.597**
Germany	Thüringen	**0.897**	**0.918**
Greece	Ipiros-Peloponissos-Nissi Ioniou	*0.012*	0.014
Greece	Makedonia-Thraki	*0.011*	*0.012*
Greece	Sterea Ellas-Nissi Egaeou-Kriti	0.014	0.016
Greece	Thessalia	0.017	0.022
Hungary	Alföld	0.047	0.046
Hungary	Dunántúl	0.083	0.075
Hungary	Észak-Magyarország	0.038	0.031
Ireland	Ireland	0.174	0.193
Italy	Abruzzo	0.017	0.016
Italy	Alto Adige	0.056	0.054
Italy	Basilicata	0.023	0.021
Italy	Calabria	*0.000*	*0.000*
Italy	Campania	0.041	0.041
Italy	Emilia-Romagna	0.108	0.109
Italy	Friuli-Venezia Giulia	0.060	0.058
Italy	Lazio	0.042	0.040
Italy	Liguria	*0.005*	*0.004*
Italy	Lombardia	0.305	0.287
Italy	Marche	0.012	*0.009*
Italy	Molise	0.028	0.027
Italy	Piemonte	0.098	0.088
Italy	Puglia	*0.010*	*0.012*
Italy	Sardegna	0.062	0.067
Italy	Sicilia	*0.007*	*0.006*
Italy	Toscana	0.014	0.014
Italy	Trentino	0.019	0.020
Italy	Umbria	0.034	0.038
Italy	Valle d'Aosta	0.107	0.106

(continued)

Table 3.1 (continued)

Member state	Region	Year	
		2020	2021
Italy	Veneto	0.107	0.075
Latvia	Latvia	0.068	0.072
Lithuania	Lithuania	0.032	0.039
Luxembourg	Luxembourg	0.465	0.470
Netherlands	The Netherlands	**0.673**	**0.714**
Poland	Malopolska i Pogórze	*0.012*	*0.011*
Poland	Mazowsze i Podlasie	0.040	0.041
Poland	Pomorze i Mazury	0.081	0.089
Poland	Wielkopolska and Slask	0.054	0.058
Portugal	Açores e Madeira	0.047	0.043
Portugal	Alentejo e Algarve	0.020	0.024
Portugal	Norte e Centro	0.020	0.022
Portugal	Ribatejo e Oeste	*0.000*	*0.002*
Romania	Bucuresti-Ilfov	*0.008*	*0.003*
Romania	Centru	0.032	0.032
Romania	Nord-Est	0.016	0.016
Romania	Nord-Vest	0.013	0.016
Romania	Sud-Est	0.022	0.017
Romania	Sud-Muntenia	0.019	0.016
Romania	Sud-Vest-Oltenia	*0.011*	*0.008*
Romania	Vest	0.018	0.019
Slovakia	Slovakia	0.485	0.383
Slovenia	Slovenia	0.023	0.026
Spain	Andalucía	0.019	0.020
Spain	Aragón	0.164	0.108
Spain	Asturias	0.118	0.121
Spain	Canarias	0.091	0.096
Spain	Cantabria	0.150	0.165
Spain	Castilla y León	0.216	0.218
Spain	Castilla-La Mancha	0.157	0.114
Spain	Cataluña	0.141	0.143
Spain	Comunidad Valenciana	0.065	0.069
Spain	Extremadura	0.149	0.144

(continued)

Table 3.1 (continued)

Member state	Region	Year	
		2020	2021
Spain	Galicia	0.142	0.132
Spain	Islas Baleares	0.053	0.059
Spain	La Rioja	0.025	0.025
Spain	Madrid	0.112	0.118
Spain	Murcia	0.034	0.036
Spain	Navarra	0.151	0.151
Spain	País Vasco	0.100	0.106
Sweden	Län i norra Sverige	0.170	0.195
Sweden	Skogsoch mellanbygdslän	0.271	0.309
Sweden	Slättbyggdslän	0.204	0.229

Note Bold corresponds to the highest values and italic to the lowest

3.3 Results Obtained

Tables 3.2 and 3.4 reveal that the most accurate models (for the training set) to predict the total output livestock and livestock products (€/farm) are the following: CHAID, linear, neural net, random trees and C&R tree. Figures 3.2 and 3.3, for the relationships between the observed values and the predicted ones, confirm the relevant accuracy of these models.

The number of livestock units and some of the costs associated with the livestock activity may be important predictors of the total output livestock and livestock products (€/farm). This is highlighted by the results found in Tables 3.3 and 3.5.

Table 3.2 Models with the highest accuracy (the lowest relative error) for the livestock output of the European Union farming regions, with data at the farm level, for the year 2020

Model	Build time	Correlation	No. fields	Relative error
CHAID	1	1.000	2	0.000
Linear	1	1.000	4	0.000
Neural net	1	1.000	169	0.001
Random trees	1	0.979	178	0.043
C&R tree	1	0.934	39	0.129

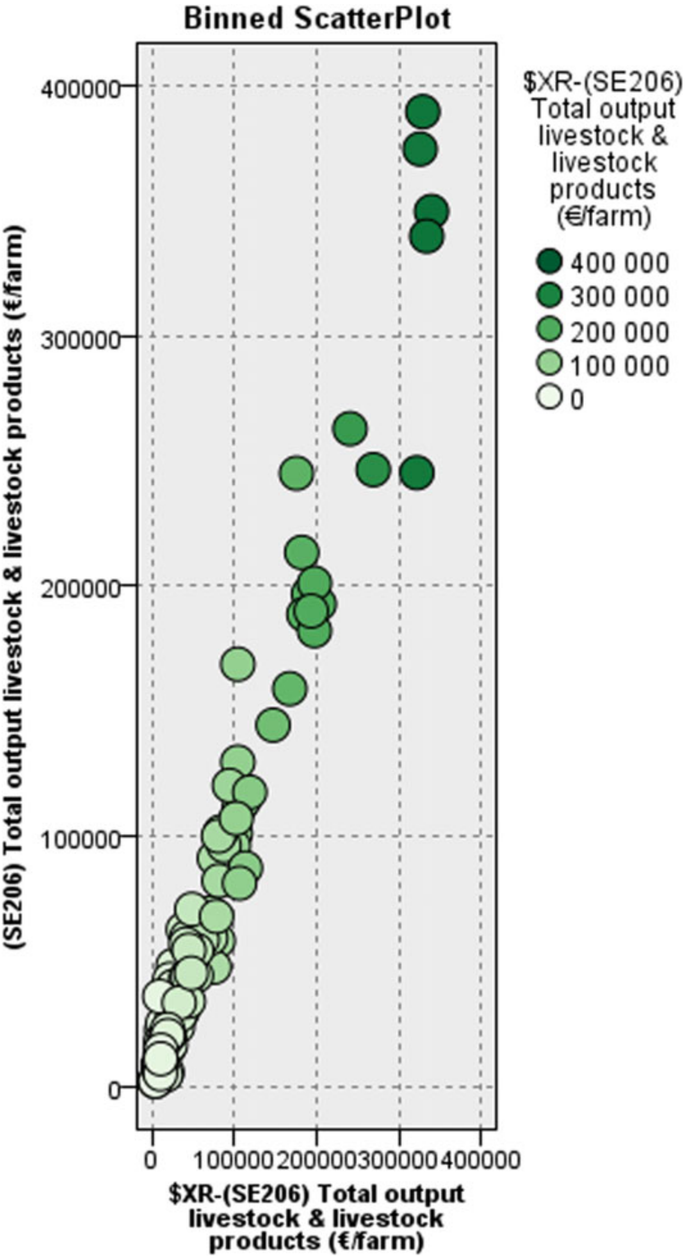

Fig. 3.2 Relationships between the observed values and the predicted ones for the livestock output of the European Union farming regions, with data at the farm level, for the year 2020

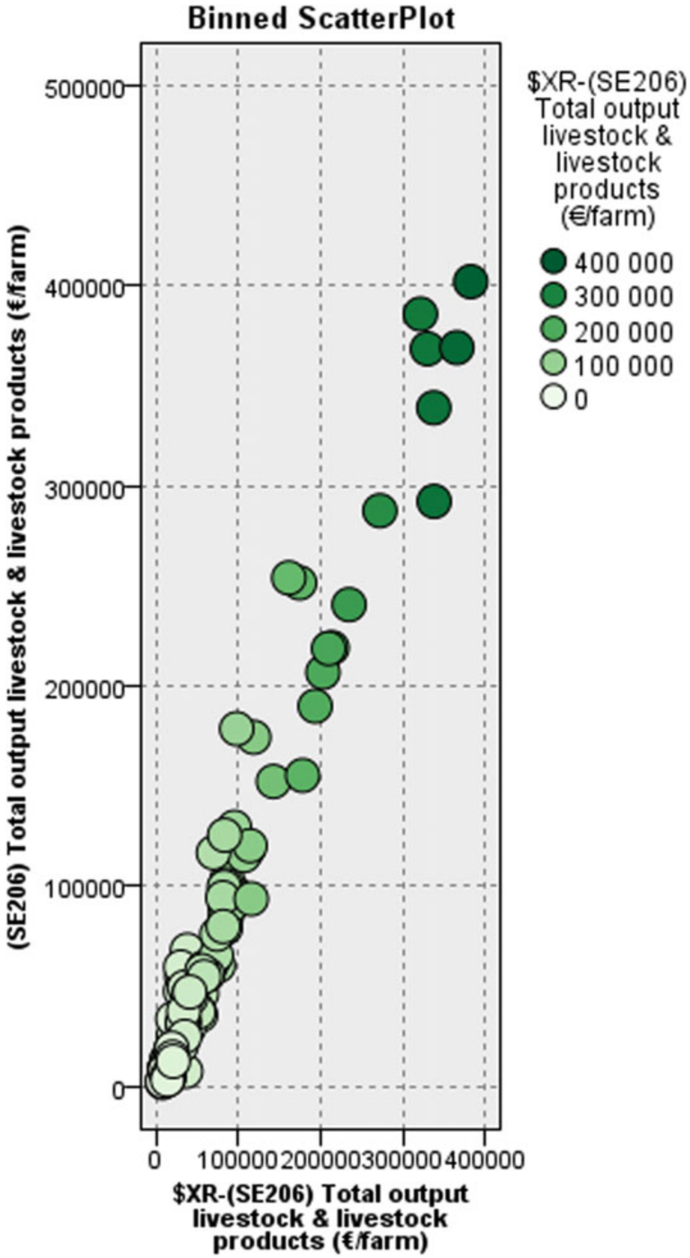

Fig. 3.3 Relationships between the observed values and the predicted ones for the livestock output of the European Union farming regions, with data at the farm level, for the year 2021

Table 3.3 Importance of the predictors for the livestock output of the European Union farming regions, with data at the farm level, for the year 2020

Nodes	Importance
Yield of maize (q/ha)	0.0071
Cattle dairy cows (LU)	0.0075
Cows' milk and milk products (€/farm)	0.0076
Dairy cows (LU)	0.0079
Veterinary expenses (€/farm)	0.0080
Breeding livestock (€)	0.0083
Total specific costs (€)	0.0085
Total inputs (€)	0.0086
Total livestock units (LU)	0.0370
Other livestock specific costs (incl. veterinary expenses) (€/farm)	0.0382

Table 3.4 Models with the highest accuracy (the lowest relative error) for the livestock output of the European Union farming regions, with data at the farm level, for the year 2021

Model	Build time	Correlation	No. fields	Relative error
CHAID	< 1	1.000	4	0.000
Linear	< 1	1.000	3	0.000
Neural net	< 1	1.000	169	0.001
Random trees	< 1	0.989	178	0.027
C&R trees	< 1	0.951	16	0.100

Table 3.5 Importance of the predictors for the livestock output of the European Union farming regions, with data at the farm level, for the year 2021

Nodes	Importance
Dairy cows (LU)	0.0081
Specific crop costs (€/ha)	0.0086
Forestry and wood processing (€)	0.0089
Other crop output (€/farm)	0.0095
(Vegetables and flowers (ha)	0.0096
Non-breeding livestock (€)	0.0110
Total inputs (€)	0.0117
Veterinary expenses (€/farm)	0.0143
Total livestock units (LU)	0.0245
Other livestock specific costs (incl. veterinary expenses) (€/farm)	0.0271

3.4 Discussion and Conclusions

The livestock activities have an economic relevance for farmers specialised in these productions and contribute to food security worldwide. Nonetheless, the relevance of these activities for world sustainability is not consensual between researchers, because of their impacts on the environmental conditions and their implications, in certain circumstances, on human health. In any case, the prediction of the livestock output may bring relevant contributions for both of these frameworks, in some situations to support the mitigation of negative impacts and in other contexts to improve the revenues of the involved farmers in these specific productions. In this scenario, this chapter proposed to bring more insights about accurate models and important variables to predict the livestock output in the European Union farms. Considering these objectives, machine learning approaches were considered, following the procedures proposed by new solutions. Microeconomic data, at the farm level, were also taken into account from European Union databases.

In the current times, the information is fundamental to support better management and planning decisions. The point here is to adopt adjusted approaches to collect and assess this information. The new digital solutions may play here a relevant role. Several smart methodologies have been considered by the scientific community, such as the following: generalised linear regression; random forest; multilayer neural networks; multiple linear regression; network-based fuzzy inference system; multilayer perceptron and support vector machine.

The data assessment highlights, for example, Lithuania, Germany, Sweden, Ireland, Slovenia, Greece and Portugal as some of the European Union countries with the highest growth rates (at current prices) for livestock output among 2020 and 2021. However, Denmark, Netherlands, Germany and France, for instance, are countries with contexts that show the importance of livestock activities for the respective farmers. On the other hand, Greece, Italy, Poland, Portugal and Romania have the agricultural frameworks where the sector has the lowest values for the livestock output.

The results revealed the accuracy of models, such as CHAID, linear, neural net, random trees and C&R tree to predict the livestock output and the importance of the following predictors: number of livestock units; and some of the costs associated with the livestock activity.

In terms of practical implications, the number of livestock units may be considered as important predictors of the livestock output, through machine learning approaches. The design of policies, in the framework of the Common Agricultural Policy, for example, to promote more sustainable livestock production and meat consumption may be relevant as a suggestion in terms of policy recommendation, namely to make the output less interconnected with the number of livestock units. For future research, it would be relevant to assess the effect of some variables lagged one or more years.

Acknowledgements This work is funded by National Funds through the FCT—Foundation for Science and Technology, I.P., within the scope of the project Refª UIDB/00681/2020 (https://doi.org/https://doi.org/10.54499/UIDP/00681/2020). This research is also funded by the Enovo company. This study was carried out under the international project "Agriculture 4.0: Current reality, potentialities and policy proposals" (CERNAS-IPV/2022/008). Furthermore, we would like to thank the CERNAS Research Centre and the Polytechnic Institute of Viseu for their support. This work is too co-financed by the PRR—Plano de Recuperação e Resiliência (República Portuguesa) and the European Next Generation EU Funds (https://recuperarportugal.gov.pt) through application PRR-C05-i03-I-000030—"Carb2Soil—Reforçar a Complementaridade entre agricultura e pecuária para aumentar a fertilidade dos solos e a sua capacidade de sequestro de carbono".

References

1. V.C.F. Aiken, A.F.A. Fernandes, T.L. Passafaro, J.S. Acedo, F.G. Dias, J.R.R. Dorea, G.J.M. Rosa, Forecasting beef production and quality using large-scale integrated data from Brazil. J. Anim. Sci. **98**, skaa089 (2020)
2. D.J. Murphy, P. Shine, B.O. Brien, M.O. Donovan, M.D. Murphy, Utilising grassland management and climate data for more accurate prediction of herbage mass using the rising plate meter. Precis. Agric. **22**, 1189 (2021)
3. S. Nosratabadi, S. Ardabili, Z. Lakner, C. Mako, A. Mosavi, Prediction of food production using machine learning algorithms of multilayer perceptron and ANFIS. Agricult. Basel **11**, 408 (2021)
4. W. Wang, S. Chen, L. Qiao, S. Zhang, Q. Liu, K. Yang, Y. Pan, J. Liu, W. Liu, Four markers useful for the distinction of intrauterine growth restriction in sheep. Animals **13**, 3305 (2023)
5. X. Wang, Z. Zhang, B. Li, W. Hao, W. Yin, S. Ai, J. Han, R. Wang, Z. Duan, Depicting fecal microbiota characteristic in Yak, Cattle, Yak-Cattle hybrid and Tibetan sheep in different eco-regions of Qinghai-Tibetan Plateau. Microbiol. Spectr. **10**, e00021–e00022 (2022)
6. R.N. Watanabe, P.A. Bernardes, E.P. Romanzini, L.G. Braga, T.R. Brito, R.W. Teobaldo, R.A. Reis, D.P. Munari, Strategy to predict high and low frequency behaviors using triaxial accelerometers in grazing of beef cattle. Animals **11**, 3438 (2021)
7. M. Baker et al., Machine learning and metagenomics reveal shared antimicrobial resistance profiles across multiple chicken farms and abattoirs in China. Nat. Food **4**, 707 (2023)
8. L. Feng, Z. Zhang, Y. Ma, Y. Sun, Q. Du, P. Williams, J. Drewry, B. Luck, Multitask learning of alfalfa nutritive value from UAV-based hyperspectral images. IEEE Geosci. Remote Sens. Lett. **19**, 1–5 (2022)
9. E.S. Fogarty, D.L. Swain, G.M. Cronin, L.E. Moraes, D.W. Bailey, M. Trotter, Developing a simulated online model that integrates GNSS, accelerometer and weather data to detect parturition events in grazing sheep: a machine learning approach. Animals **11**, 303 (2021)
10. G. Gebreyesus, V. Milkevych, J. Lassen, G. Sahana, Supervised learning techniques for dairy cattle body weight prediction from 3D digital images. Front. Genet. **13**, 947176 (2023)
11. T.G. Morais, M. Jongen, C. Tufik, N.R. Rodrigues, I. Gama, J. Serrano, M.C. Goncalves, R. Mano, T. Domingos, R.F.M. Teixeira, Satellite-based estimation of soil organic carbon in Portuguese grasslands. Front. Environ. Sci. **11**, 1240106 (2023)
12. J. Barnetson, S. Phinn, P. Scarth, Mapping woody vegetation cover across Australia's arid rangelands: utilising a machine-learning classification and low-cost remotely piloted aircraft system. Int. J. Appl. Earth Obs. Geoinf. **83**, 101909 (2019)
13. C.D. Hasselerharm, E. Yanco, J.S. McManus, B.H. Smuts, D. Ramp, Wildlife-friendly farming recouples grazing regimes to stimulate recovery in semi-arid rangelands. Sci. Total. Environ. **788**, 147602 (2021)
14. J.O. Chelotti, S.R. Vanrell, L.S.M. Rau, J.R. Galli, A.M. Planisich, S.A. Utsumi, D.H. Milone, L.L. Giovanini, H.L. Rufiner, An online method for estimating grazing and rumination bouts using acoustic signals in grazing cattle. Comput. Electron. Agric. **173**, 105443 (2020)

15. J.O. Chelotti, S.R. Vanrell, L.S. Martinez-Rau, J.R. Galli, S.A. Utsumi, A.M. Planisich, S.A. Almiron, D.H. Milone, L.L. Giovanini, H.L. Ruffner, Using segment-based features of jaw movements to recognise foraging activities in grazing cattle. Biosyst. Eng. **229**, 69–84 (2023)
16. E. Cobbinah, O. Generalao, S.K. Lageshetty, I. Adrianto, S. Singh, G.G. Dumancas, Using near-infrared spectroscopy and stacked regression for the simultaneous determination of fresh cattle and poultry manure chemical properties. Chemosensors **10**, 410 (2022)
17. H. Doi, R.N. Mulia, Future land use for insect meat production among countries: a global classification. Front. Nutr. **8**, 661056 (2021)
18. M.C.R. Fasiaben, S.R.M. Oliveira, A.S. Moraes, M.M.T.B. Almeida, O.C. de Oliveira, G.S. Eusebio, Differentiation and characterization of beef cattle producing municipalities in the Mata Atlantica biome according to the 2017 Agricultural Census. Rev. Bras. Comput. Apl. **14**, 67 (2022)
19. S. Gritsenko, A. Ruchay, V. Kolpakov, S. Lebedev, H. Guo, A. Pezzuolo, On-barn forecasting beef cattle production based on automated non-contact body measurement system. Animals **13**, 611 (2023)
20. S. He et al., Differences in background environment and fertilization method mediate plant response to nitrogen fertilization in alpine grasslands on the Qinghai-Tibetan Plateau. Sci. Total. Environ. **906**, 167272 (2024)
21. T.P. Kharel, A.J. Ashworth, P.R. Owens, D. Philipp, A.L. Thomas, T.J. Sauer, Teasing apart Silvopasture system components using machine learning for optimization. Soil Syst. **5**, 41 (2021)
22. S. Lee, I. Lee, U. Yeo, J. Kim, R. Kim, Machine learning approach to predict air temperature and relative humidity inside mechanically and naturally ventilated duck houses: application of recurrent neural network. Agricult. Basel **12**, 318 (2022)
23. M. Lozano-Jaramillo, S.W. Alemu, T. Dessie, H. Komen, J.W.M. Bastiaansen, Using phenotypic distribution models to predict livestock performance. Sci. Rep. **9**, 15371 (2019)
24. M. Mito, T. Aoki, K. Mizutani, K. Zempo, N. Wakatsuki, Y. Maeda, N. Takemae, R. Tsunekuni, T. Saito, Frequency analysis of the sneeze caused by swine influenza virus strains: automatic sneeze around-the-clock detection using a support vector machine. Comput. Electron. Agric. **179**, 105789 (2020)
25. K. Shea, D. Schaffer-Smith, R.L. Muenich, Using remote sensing to identify liquid manure applications in eastern North Carolina. J. Environ. Manag. **317**, 115334 (2022)
26. R.M.F. Silveira, D.A.E. Facanha, C.M. Mcmanus, J. Ferreira, I.J.O. da Silva, Machine intelligence applied to sustainability: a systematic methodological proposal to identify sustainable animals. J. Clean. Prod. **420**, 138292 (2023)
27. FADN, *Several Statistics*. https://agriculture.ec.europa.eu/data-and-analysis/farm-structures-and-economics/fadn_en
28. IBM SPSS Modeler, *SPSS Modeler: Overview*. https://www.ibm.com/products/spss-modeler

Chapter 4
Predicting the Total Costs of Production Factors on Farms in the European Union

Abstract The dynamics of the agricultural sector depend on the performance of the farms and their respective profitability. The cost control in the farms is particularly important, considering the reduced profit margins in agriculture. In fact, in some contexts, the level of farm costs is very similar to the amounts of income, calling, in many cases, for financial support for the farmers, justified by the need to guarantee food security and social and environmental sustainability. In this framework, contributions that support policymakers and farmers to make decisions that promote farm cost reduction are fundamental. Considering this scenario, this study intends to consider machine learning approaches and data from the European databases to identify the most adjusted approaches to predict the total costs in the farms. This study brought relevant outputs for the design of adjusted measures, plans and instruments for the European Union agriculture and respective processes and activities.

Keywords Artificial intelligence · Adjusted models and predictors · Agriculture

4.1 Introduction

Often, the costs of production in the farms affect significantly the level of profitability in the agricultural sector and this justifies, in some circumstances the subsidies given to farmers, considering their contributions to economic, social and environmental sustainability. In fact, some farming productions achieve high levels of revenue, but the profitability is low due to the amount of costs associated.

In this perspective, the total costs assessment, which includes different items [1], assumes particular importance in any activity [2] and socioeconomic sector. Of referring, for example, the following contexts, where the costs analysis, based on machine learning approaches, was highlighted as relevant: supplier–buyer interactions [3]; medical practices [4]; healthcare systems [5]; medical therapies [6]; hospital emergency management [7]; hospital surgeries management [8]; hospital pharmacy inventory [9]; postoperative decisions [10, 11]; antibiotic management [12]; urban waste management [13]; computation systems [14]; transportation networks [15];

V. J. P. D. Martinho, *Machine Learning Approaches for Evaluating Statistical Information in the Agricultural Sector*, SpringerBriefs in Applied Sciences and Technology, https://doi.org/10.1007/978-3-031-54608-2_4

barge planning systems [16]; equipment inspection needs [17]; greenhouse requirements [18]; bridge construction [19]; engineering decisions [20]; victims evacuation [21]; inventory planning and management [22]; wind turbine technical problems assessment [23]; software defects prediction [24]; credit card fraud prediction [25]; human–robot partnership [26]; debris removal [27] and manufacturing decisions [28].

The machine learning approaches have allowed to improve the efficiency and the accuracy of the models used to predict the more diverse dimensions and contexts, however, often, lack in these methodologies the capacity to apply the predictive scenarios in the real frameworks. Some research claims the need for approaches that put together learning and planning, as two issues of artificial intelligence [29].

A great part of the studies analysed, related to cost evaluation, taking into account machine learning models, focused on medical practices, healthcare systems management, urban planning, engineering and manufacturing decisions and computation contexts. This indicates that there is still a field to be explored related to the agricultural contexts and the potentialities to better manage the farming costs.

Considering these motivations, this study proposes to predict the total costs associated with the use of production factors in the European Union farms, taking into account statistical information from European databases with microdata [30] and using approaches proposed by solutions [31] that consider machine learning methodologies. This research is based on the previous results emphasised by Martinho [32].

4.2 Data Investigation

Slovakia, Denmark, Netherlands, Czechia, Germany, Luxembourg, Belgium, Sweden and France are some of the European Union countries with the highest values for the total costs [total inputs (€)], over the year 2021 (Fig. 4.1). The lowest values appear for countries such as the following: Portugal; Croatia; Romania and Greece. Of course, we have here the effect of the price levels between the European Union member states, nonetheless, the objective is to analyse the different realities of the countries.

The normalised values $((x_i - x_{\text{minimum}})/(x_{\text{maximum}} - x_{\text{minimum}}))$ presented in Table 4.1 confirm the dimension of the total costs [total inputs (€)] in the farms from Czechia, Denmark, Germany, Netherlands and Slovakia, for example. Agricultural regions from Greece, Poland, Portugal and Romania, for instance, are between the European Union frameworks where the farms present the lowest values for the total inputs (€).

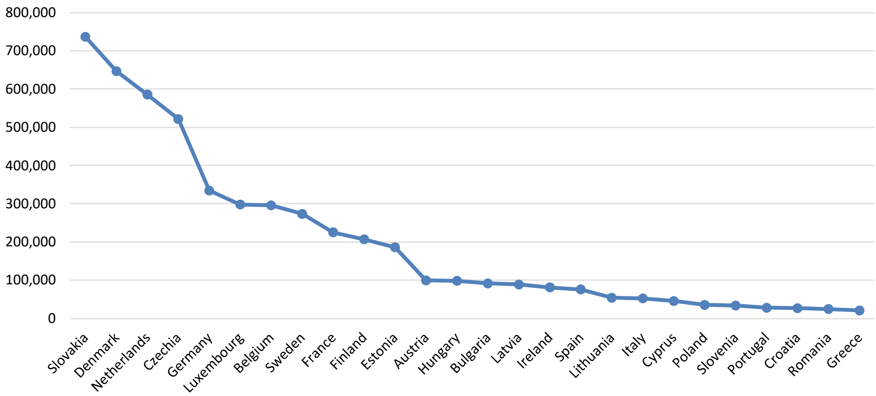

Fig. 4.1 Values in levels for the total costs of the European Union countries, with data at the farm level, for the year 2021

Table 4.1 Normalised values for the total costs of the European Union farming regions, with data at the farm level, for the year 2021

Member state	Region	Year 2021
Austria	Austria	0.066
Belgium	Vlaanderen	0.269
Belgium	Wallonie	0.136
Bulgaria	Severen tsentralen	0.090
Bulgaria	Severoiztochen	0.104
Bulgaria	Severozapaden	0.099
Bulgaria	Yugoiztochen	0.063
Bulgaria	Yugozapaden	0.015
Bulgaria	Yuzhen tsentralen	0.025
Croatia	Jadranska Hrvatska	*0.003*
Croatia	Kontinentalna Hrvatska	0.010
Cyprus	Cyprus	0.023
Czechia	Czechia	**0.396**
Denmark	Denmark	**0.493**
Estonia	Estonia	0.133
Finland	Etelä-Suomi	0.122
Finland	Pohjanmaa	0.183
Finland	Pohjois-Suomi	0.199
Finland	Sisä-Suomi	0.159
France	Alsace	0.121

(continued)

Table 4.1 (continued)

Member state	Region	Year
		2021
France	Aquitaine	0.147
France	Auvergne	0.108
France	Basse-Normandie	0.211
France	Bourgogne	0.183
France	Bretagne	0.249
France	Centre	0.166
France	Champagne-Ardenne	0.150
France	Corse	0.100
France	Franche-Comté	0.172
France	Guadeloupe	0.045
France	Haute-Normandie	0.215
France	Île-de-France	0.208
France	La Réunion	0.064
France	Languedoc-Roussillon	0.102
France	Limousin	0.097
France	Lorraine	0.186
France	Midi-Pyrénées	0.104
France	Nord-Pas-de-Calais	0.207
France	Pays de la Loire	0.243
France	Picardie	0.200
France	Poitou–Charentes	0.195
France	Provence-Alpes-Côte d'Azur	0.134
France	Rhône-Alpes	0.134
Germany	Baden-Württemberg	0.135
Germany	Bayern	0.144
Germany	Brandenburg	**0.923**
Germany	Hessen	0.157
Germany	Mecklenburg-Vorpommern	**0.891**
Germany	Niedersachsen	**0.288**
Germany	Nordrhein-Westfalen	0.243
Germany	Rheinland-Pfalz	0.142
Germany	Saarland	0.135
Germany	Sachsen	**0.761**
Germany	Sachsen-Anhalt	**0.755**
Germany	Schleswig–Holstein/Hamburg	0.279
Germany	Thüringen	**1.000**
Greece	Ipiros-Peloponissos-Nissi Ioniou	*0.001*

(continued)

Table 4.1 (continued)

Member state	Region	Year
		2021
Greece	Makedonia-Thraki	0.007
Greece	Sterea Ellas-Nissi Egaeou-Kriti	*0.003*
Greece	Thessalia	0.008
Hungary	Alföld	0.053
Hungary	Dunántúl	0.101
Hungary	Észak-Magyarország	0.033
Ireland	Ireland	0.051
Italy	Abruzzo	0.012
Italy	Alto Adige	0.035
Italy	Basilicata	0.014
Italy	Calabria	*0.004*
Italy	Campania	0.018
Italy	Emilia-Romagna	0.052
Italy	Friuli-Venezia Giulia	0.043
Italy	Lazio	0.031
Italy	Liguria	0.016
Italy	Lombardia	0.099
Italy	Marche	0.014
Italy	Molise	0.011
Italy	Piemonte	0.046
Italy	Puglia	0.017
Italy	Sardegna	0.018
Italy	Sicilia	0.010
Italy	Toscana	0.030
Italy	Trentino	0.017
Italy	Umbria	0.027
Italy	Valle d'Aosta	0.039
Italy	Veneto	0.050
Latvia	Latvia	0.057
Lithuania	Lithuania	0.030
Luxembourg	Luxembourg	0.221
Netherlands	The Netherlands	**0.445**
Poland	Malopolska i Pogórze	*0.002*
Poland	Mazowsze i Podlasie	0.010
Poland	Pomorze i Mazury	0.037
Poland	Wielkopolska and Slask	0.026
Portugal	Açores e Madeira	0.008

(continued)

Table 4.1 (continued)

Member state	Region	Year
		2021
Portugal	Alentejo e Algarve	0.019
Portugal	Norte e Centro	*0.006*
Portugal	Ribatejo e Oeste	0.017
Romania	Bucuresti-Ilfov	*0.005*
Romania	Centru	0.008
Romania	Nord-Est	*0.007*
Romania	Nord-Vest	*0.001*
Romania	Sud-Est	0.015
Romania	Sud-Muntenia	0.012
Romania	Sud-Vest-Oltenia	*0.000*
Romania	Vest	0.009
Slovakia	Slovakia	**0.563**
Slovenia	Slovenia	0.014
Spain	Andalucía	0.031
Spain	Aragón	0.061
Spain	Asturias	0.035
Spain	Canarias	0.101
Spain	Cantabria	0.042
Spain	Castilla y León	0.074
Spain	Castilla-La Mancha	0.052
Spain	Cataluña	0.079
Spain	Comunidad Valenciana	0.032
Spain	Extremadura	0.045
Spain	Galicia	0.031
Spain	Islas Baleares	0.034
Spain	La Rioja	0.061
Spain	Madrid	0.042
Spain	Murcia	0.051
Spain	Navarra	0.079
Spain	País Vasco	0.046
Sweden	Län i norra Sverige	0.142
Sweden	Skogsoch mellanbygdslän	0.172
Sweden	Slättbyggdslän	0.220

Note Bold corresponds to the highest values and italic to the lowest

4.3 Results Found

The models with the highest accuracy (for the training set), in the year 2021, are the following (Table 4.2): linear (linear regression); CHAID (Chi-squared Automatic Interaction Detection); regression (linear regression); linear-AS (linear regression); neural net (neural network); random trees (multiple decision trees); random forest (considers a tree model as the reference model); C&R tree (Classification and Regression tree) and SVM (support vector machine). The pertinence of these models is confirmed by Fig. 4.2 for the relationships between the observed values and the predicted ones for the total inputs (€).

The most important predictors of the total inputs (€), in the European Union farms, are the following (Table 4.3): unpaid labour input (hrs); subsidies on external factors (€); total intermediate consumption (€); forestry specific costs (€); agritourism (€); other cattle (LU); other output (€/farm); sugar beet (€/farm); energy crops (ha) and yield of wheat (q/ha).

These results show the importance of unpaid labour, subsidies, intermediate consumption, livestock units and sugar beet, for example, to predict the total inputs in the European Union agricultural regions.

Table 4.2 Models with the highest accuracy (the lowest relative error) for the total costs of the European Union farming regions, with data at the farm level, for the year 2021

Model	Build time	Correlation	No. fields	Relative error
Linear	< 1	1.000	6	0.000
CHAID	< 1	1.000	7	0.000
Regression	< 1	1.000	60	0.000
Linear-AS	< 1	1.000	178	0.000
Neural net	< 1	1.000	169	0.001
Random trees	< 1	0.993	178	0.018
Random forest	< 1	0.994	178	0.021
C&R tree	< 1	0.978	30	0.044
SVM	< 1	0.690	169	1.151

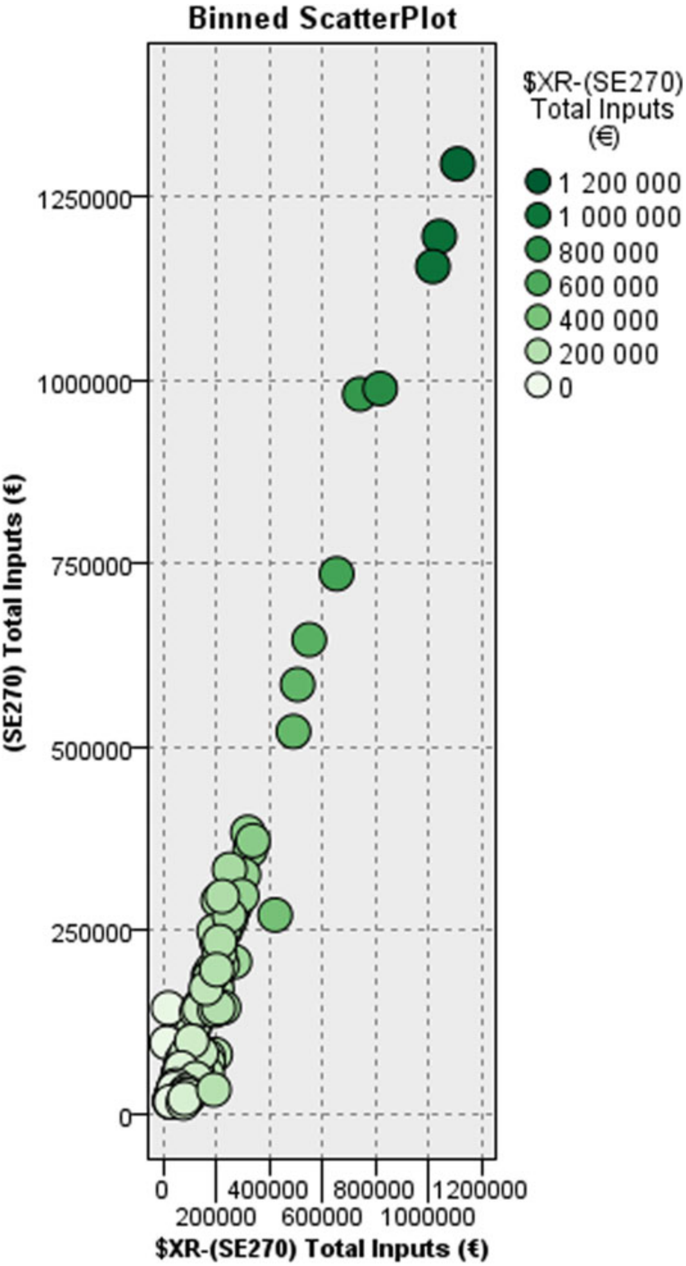

Fig. 4.2 Relationships between the observed values and the predicted ones for the total costs of the European Union farming regions, with data at the farm level, for the year 2021

Table 4.3 Importance of the predictors for the total costs of the European Union farming regions, with data at the farm level, for the year 2021

Nodes	Importance
Unpaid labour input (h)	0.0070
Subsidies on external factors (€)	0.0071
Total intermediate consumption (€)	0.0073
Forestry specific costs (€)	0.0073
Agritourism (€)	0.0078
Other cattle (LU)	0.0078
Other output (€/farm)	0.0083
Sugar beet (€/farm)	0.0101
Energy crops (ha)	0.0115
Yield of wheat (q/ha)	0.0189

4.4 Discussion and Conclusions

The level of the costs in the farms impacts significantly the profitability of the agricultural sector. In fact, in some farming productions, the amount of the revenues is high, but often, in these cases, the profitability is affected due to the dimension of the total costs associated with the different activities developed in the farms. In general, some of these costs may be reduced with more efficient practices and processes. The new solutions from the digital era offer potentialities to increase agricultural outputs with the same, or fewer, resources. This is true for the use of energy and water in the farms, for example, but it is also true for the use of other production factors, such as fertilisers and crop protection products. Taking into account these motivations, this study suggested analysing the accurate models that may support the stakeholders to predict total costs in the European Union representative farms, considering data for the year 2021 from European databases with microeconomic data. These data considered at the farm level were assessed through machine learning solutions following the procedures proposed by new approaches.

The literature review shows that there is still a field to be explored about the application of the new digital technologies to predict the total costs in the agricultural sector worldwide and its importance in supporting the farmers' decisions, as well as the design process of policies implemented in the farming frameworks. The low profitability of agriculture and considering the importance of the agricultural sector for sustainability, specifically of less favoured regions, justify, in some circumstances the implementation of policies to maintain the farmers in the respective activities and regions.

In the European Union agricultural context, Slovakia, Denmark, Netherlands, Czechia, Germany, Luxembourg, Belgium, Sweden and France are the countries with the greatest values for the total costs. Portugal, Croatia, Romania and Greece have the lowest total costs. These contexts are also verified with the data disaggregated for the different European Union agricultural regions.

Linear, CHAID, regression, linear-AS, neural net, random trees, random forest, C&R tree and SVM are the most accurate models identified for the contexts assessed. On the other hand, unpaid labour input, subsidies on external factors, total intermediate consumption, forestry specific costs, agritourism, other cattle, other output, sugar beet, energy crops and yield of wheat are the most important predictors.

In terms of practical implication, there are here relevant insights for the stakeholders, namely to support them in predicting the total costs in the European Union farming sector. For policy recommendation, it is suggested to rethink the interlinkages between the instruments and measures, defined in the framework of the Common Agricultural Policy, and the production factors use, considering the relevance of subsidies on external factors to predict the farming costs. In future investigations, it could be important to assess the real impact of the most important predictors on the total costs. On the other hand, it would be interesting to consider the effects of time on these assessments, considering these findings and those obtained by Martinho [32]. There are some similarities, but also relevant dissimilarities explained by the consideration of data from different years.

Acknowledgements This work is funded by National Funds through the FCT—Foundation for Science and Technology, I.P., within the scope of the project Refª UIDB/00681/2020 (https://doi.org/https://doi.org/10.54499/UIDP/00681/2020). This research is also funded by the Enovo company. This study was carried out under the international project "Agriculture 4.0: Current reality, potentialities and policy proposals" (CERNAS-IPV/2022/008). Furthermore, we would like to thank the CERNAS Research Centre and the Polytechnic Institute of Viseu for their support. This work is too co-financed by the PRR—Plano de Recuperação e Resiliência (República Portuguesa) and the European Next Generation EU Funds (https://recuperarportugal.gov.pt) through application PRR-C05-i03-I-000030—"Carb2Soil—Reforçar a Complementaridade entre agricultura e pecuária para aumentar a fertilidade dos solos e a sua capacidade de sequestro de carbono".

References

1. C. Cheng, L. Tang, Robust policies for a multi-stage production/inventory problem with switching costs an uncertain demand. Int. J. Prod. Res. **56**, 4264 (2018)
2. S.M. Dastjerdi, Z.M. Mosammam, P. Ahmadi, E. Houshfar, Transient analysis and optimization of an off-grid hydrogen and electric vehicle charging station with temporary residences. Sust. Cities Soc. **97**, 104742 (2023)
3. F. Bodendorf, Q. Xie, P. Merkl, J. Franke, A multi-perspective approach to support collaborative cost management in supplier-buyer dyads. Int. J. Prod. Econ. **245**, 108380 (2022)
4. A.K. Chan et al., Does state malpractice environment affect outcomes following spinal fusions? A robust statistical and machine learning analysis of 549,775 discharges following spinal fusion surgery in the United States. Neurosurg. Focus **49**, E18 (2020)
5. O. Ericson, J. Hjelmgren, F. Sjovall, J. Soderberg, I. Persson, The potential cost and cost-effectiveness impact of using a machine learning algorithm for early detection of sepsis in intensive care units in Sweden. J. Health Econ. Outcome. Res. **9**, 101 (2022)
6. L.S. Fernandes de Carvalho, S. Gioppato, M.D. Fernandez, B.C. Trindade, J.C.Q.E. Silva, R.G. Sergio Miranda, J.R. Matos de Souza, W. Nadruz, S.E. Fontes Avila, A.C. Sposito, Machine learning improves the identi fication of individuals with higher morbidity and avoidable health costs after acute coronary syndromes. Value Health **23**, 1570 (2020)

7. D.W. Frost, S. Vembu, J. Wang, K. Tu, Q. Morris, H.B. Abrams, Using the electronic medical record to identify patients at high risk for frequent emergency department visits and high system costs. Am. J. Med. **130**, e17 (2017)
8. D. Goyal, J. Guttag, Z. Syed, R. Mehta, Z. Elahi, M. Saeed, Comparing precision machine learning with consumer, quality, and volume metrics for ranking orthopedic surgery hospitals: retrospective study. J. Med. Internet Res. **22**, e22765 (2020)
9. F. Rojas, V. Leiva, M. Huerta, C. Martin-Barreiro, Lot-size models with uncertain demand considering its skewness/kurtosis and stochastic programming applied to hospital pharmacy with sensor-related COVID-19 data. Sensors **21**, 5198 (2021)
10. T.J. Loftus, M.M. Ruppert, T. Ozrazgat-Baslanti, J.A. Balch, P.A. Efron, P.J. Tighe, W.R. Hogan, P. Rashidi, G.R. Upchurch, A. Bihorac, Association of postoperative undertriage to hospital wards with mortality and morbidity. JAMA Netw. Open **4**, e2131669 (2021)
11. T.M. Loftus et al., Postoperative overtriage to an intensive care unit is associated with low value of care. Ann. Surg. **277**, 179 (2023)
12. A.M. Voermans, J.C. Mewes, M.R. Broyles, L.M.G. Steuten, Cost-effectiveness analysis of a procalcitonin-guided decision algorithm for antibiotic stewardship using real-world US hospital data. OMICS **23**, 508 (2019)
13. J. Ferrer, E. Alba, BIN-CT: urban waste collection based on predicting the container fill level. Biosystems **186**, 103962 (2019)
14. S. Ghose, A. Boroumand, J.S. Kim, J. Gomez-Luna, O. Mutlu, Processing-in-memory: a workload-driven perspective. IBM J. Res. Dev. **63**, 3 (2019)
15. E. Gocmen, R. Erol, Transportation problems for intermodal networks: mathematical models, exact and heuristic algorithms, and machine learning. Exp. Syst. Appl. **135**, 374 (2019)
16. V. Gumuskaya, W. van Jaarsveld, R. Dijkman, P. Grefen, A. Veenstra, Integrating stochastic programs and decision trees in capacitated barge planning with uncertain container arrivals. Transp. Res. Pt. C-Emerg. Technol. **132**, 103383 (2021)
17. Y. Javid, A Bi-objective mathematical model to determine risk-based inspection programs. Process Saf. Environ. Protect. **146**, 893 (2021)
18. N. Khani, M.H.K. Manesh, V.C. Onishi, Optimal 6E design of an integrated solar energy-driven polygeneration and CO_2 capture system: a machine learning approach. Therm. Sci. Eng. Prog. **38**, 101669 (2023)
19. M. Kovacevic, F. Antoniou, Machine-learning-based consumption estimation of prestressed steel for prestressed concrete bridge construction. Build. Basel **13**, 1187 (2023)
20. C. Liu, Y. Hu, T. Yu, Q. Xu, C. Liu, X. Li, C. Shen, Optimizing the water treatment design and management of the artificial lake with water quality modeling and surrogate-based approach. Water **11**, 391 (2019)
21. S.M. Nabavi, B. Vahdani, B.A. Nadjafi, M.A. Adibi, Synchronizing victim evacuation and debris removal: a data-driven robust prediction approach. Eur. J. Oper. Res. **300**, 689 (2022)
22. P. Raghuram, S. Bhupesh, R. Manivannan, P.S. Anand, V.R. Sreedharan, Modeling and analyzing the inventory level for demand uncertainty in the VUCA world: evidence from biomedical manufacturer. IEEE Trans. Eng. Manag. **70**, 2944 (2023)
23. A. Santolamazza, D. Dadi, V. Introna, A data-mining approach for wind turbine fault detection based on SCADA data analysis using artificial neural networks. Energies **14**, 1845 (2021)
24. V.S. Sheng, B. Gu, W. Fang, J. Wu, Cost-sensitive learning for defect escalation. Knowl. Based Syst. **66**, 146 (2014)
25. A. Singh, A. Jain, An efficient credit card fraud detection approach using cost-sensitive weak learner with imbalanced dataset. Comput. Intell. **38**, 2035 (2022)
26. T. Smith, P. Benardos, D. Branson, Assessing worker performance using dynamic cost functions in human robot collaborative tasks. Proc. Inst. Mech. Eng. C J. Eng. Mech. Eng. Sci. **234**, 289 (2020)
27. B. Vahdani, A flexible framework to coordinate debris clearance and relief distribution operations: a robust machine learning approach. Exp. Syst. Appl. **229**, 120512 (2023)
28. F. Zheng, Z. Wang, Y. Xu, M. Liu, Single machine scheduling with uncertain processing times and carbon emission constraint in the shared manufacturing environment. Ann. Oper. Res. **12**, 1–31 (2023)

29. Q. Lv, Y. Chen, Z. Li, Z. Cui, L. Chen, X. Zhang, H. Shen, Achieving data-driven actionability by combining learning and planning. Front. Comput. Sci. **12**, 939 (2018)
30. FADN, Several Statistics. https://agriculture.ec.europa.eu/data-and-analysis/farm-structures-and-economics/fadn_en
31. IBM SPSS Modeler, SPSS Modeler: Overview. https://www.ibm.com/products/spss-modeler
32. V.J.P.D. Martinho, Economic growth, sustainability assessment and artificial intelligence: combinations among these three dimensions, in *Economic Growth: Advances in Analysis Methodologies and Technologies* (Springer, Cham, 2023), pp. 133–144

Chapter 5
The Most Important Predictors of Fertiliser Costs

Abstract The control of the fertiliser costs in the agricultural sector is fundamental for the profitability of the farms and to mitigate environmental impacts. Indeed, the fertiliser costs have, at least, two components, one related to the fertiliser prices and the other associated with the amount of fertiliser applied in the farming processes. The fertiliser application in agricultural activities has a relevant impact on soil health and water quality. The efficiency of the processes linked with the fertiliser application in the farms is crucial to avoid disruptions in the sustainable development required for agriculture worldwide. In these frameworks, it is important to bring more insights about the predictors of the fertiliser costs in the European Union farms. Taking into account these motivations, this chapter considered artificial intelligence approaches and data from the European Union databases to identify the most adjusted models. The findings of this research contribute to the understanding of the most important variables to promote more sustainability in the European Union farming sector.

Keywords European Union agriculture · Machine learning models · Farming indicators

5.1 Introduction

Between the different agricultural costs, those related to fertilisers assume special relevance, considering their impacts on the budget of the farms and the environmental consequences, namely in the soil and water characteristics and quality. In this way, a continuous assessment of these costs is important and the new technologies and methodologies associated with the smart transition may bring here significant contributions [1] to the automation and digitalisation of agriculture [2]. The current challenges claim new approaches in the agricultural sector that ensure a more sustainable development in agricultural activities [3] and adjusted fertiliser recommendations [4] and manure applications [5].

The main contribution, of the digital transformation in the context of Agriculture 4.0, for more sustainable fertiliser applications, is associated with the possibility of

V. J. P. D. Martinho, *Machine Learning Approaches for Evaluating Statistical Information in the Agricultural Sector*, SpringerBriefs in Applied Sciences and Technology, https://doi.org/10.1007/978-3-031-54608-2_5

improving the efficiency of the related processes in the farms [6]. These approaches allow for reducing the respective costs [7], namely in times of higher prices [8], with potentialities to increase the respective productions and profitability of the farmers [9].

The Internet of Things, robotics technology, machine learning, artificial approaches and Big Data are among the main technologies to implement smart farming to deal with the current challenges [10]. These new realities brought the need to make compatible, in the agricultural sector, productivity growth and guaranteeing food security without compromising sustainability and environmental quality [11], where the losses reduction assume special relevance [12]. The question here is the adjustment of the farming processes, with a better organisation [13], specifically in the fertilisers use [14] and management [15], for better soil quality [16] and carbon sequestration, in agreement with the international proposals [17], namely in the greenhouses structures [18], for example.

The soil characteristics indicators analysis, such as phosphorus [19] and nitrogen, is particularly important to support the farmers' decisions and better control the respective costs and the implications on sustainability [20]. The application of fertilisers is interrelated with other production factors, namely energy, machinery, crop protection products [21] and water [22]. Smart farming solutions may bring important added value for the fertiliser costs prediction [23], but also for the processes of application, using, for instance, unmanned aerial vehicles [24].

Considering the context described before, it seems important to identify adjusted models and important predictors to estimate the fertiliser costs in the European Union farms, using data from the European farming databases [25], taking into account the procedures proposed by some solutions for the machine learning approaches [26] and following the results obtained by Martinho [27].

5.2 Data Exploration

The growth rates obtained over the period 2019–2020 for the fertiliser costs [fertilisers (€)] in the European Union representative farms are presented in Fig. 5.1. The highest values with current prices (the idea is to show the differences in the dynamics of the European Union member states) were obtained for the following countries: Czechia, Denmark, Slovenia, Lithuania, Latvia, Sweden, Spain and Estonia.

Table 5.1, with the normalised values ($(x_i - x_{\text{minimum}})/(x_{\text{maximum}} - x_{\text{minimum}})$), reveals that the farms from Czechia and Slovakia are those with the highest values for fertiliser costs. The same happens for some regions of France and Germany. Some regions from Croatia, Greece, Portugal, Romania and Spain, for example, are where the representative farms have the lowest costs with fertilisers.

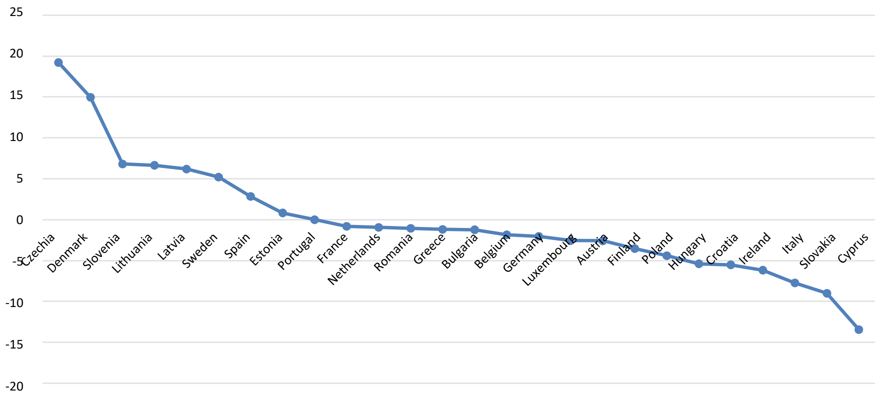

Fig. 5.1 Growth rate (%) results for the fertiliser costs of the European Union countries, with data at the farm level, over the period 2019–2020

Table 5.1 Normalised values for the fertiliser costs of the European Union farming regions, with data at the farm level, over the period 2019–2020

Member state	Region	Year	
		2019	2020
Austria	Austria	0.033	0.035
Belgium	Vlaanderen	0.127	0.131
Belgium	Wallonie	0.156	0.153
Bulgaria	Severen tsentralen	0.188	0.185
Bulgaria	Severoiztochen	0.186	0.162
Bulgaria	Severozapaden	0.185	0.212
Bulgaria	Yugoiztochen	0.111	0.122
Bulgaria	Yugozapaden	0.032	0.032
Bulgaria	Yuzhen tsentralen	0.042	0.045
Croatia	Jadranska Hrvatska	*0.006*	*0.006*
Croatia	Kontinentalna Hrvatska	0.032	0.033
Cyprus	Cyprus	0.021	0.019
Czechia	Czechia	0.302	**0.370**
Denmark	Denmark	0.249	0.295
Estonia	Estonia	0.190	0.197
Finland	Etelä-Suomi	0.123	0.123
Finland	Pohjanmaa	0.146	0.150
Finland	Pohjois-Suomi	0.154	0.157
Finland	Sisä-Suomi	0.117	0.107
France	Alsace	0.154	0.154

(continued)

Table 5.1 (continued)

Member state	Region	Year	
		2019	2020
France	Aquitaine	0.147	0.151
France	Auvergne	0.115	0.109
France	Basse-Normandie	0.182	0.180
France	Bourgogne	0.235	0.235
France	Bretagne	0.119	0.114
France	Centre	**0.358**	**0.369**
France	Champagne-Ardenne	0.270	0.272
France	Corse	0.056	0.053
France	Franche-Comté	0.199	0.199
France	Guadeloupe	0.067	0.081
France	Haute-Normandie	**0.335**	0.337
France	Île-de-France	**0.438**	**0.487**
France	La Réunion	0.111	0.154
France	Languedoc-Roussillon	0.064	0.081
France	Limousin	0.107	0.103
France	Lorraine	0.265	0.268
France	Midi-Pyrénées	0.138	0.138
France	Nord-Pas-de-Calais	0.263	0.270
France	Pays de la Loire	0.169	0.173
France	Picardie	**0.415**	**0.420**
France	Poitou–Charentes	0.246	0.229
France	Provence-Alpes-Côte d'Azur	0.108	0.109
France	Rhône-Alpes	0.115	0.127
Germany	Baden-Württemberg	0.099	0.101
Germany	Bayern	0.087	0.082
Germany	Brandenburg	**0.598**	**0.599**
Germany	Hessen	0.126	0.131
Germany	Mecklenburg-Vorpommern	**1.000**	**1.000**
Germany	Niedersachsen	0.182	0.177
Germany	Nordrhein-Westfalen	0.104	0.107
Germany	Rheinland-Pfalz	0.102	0.108
Germany	Saarland	0.151	0.155

(continued)

Table 5.1 (continued)

Member state	Region	Year	
		2019	2020
Germany	Sachsen	**0.516**	**0.569**
Germany	Sachsen-Anhalt	**0.631**	**0.650**
Germany	Schleswig–Holstein/Hamburg	0.248	0.246
Germany	Thüringen	**0.671**	**0.749**
Greece	Ipiros-Peloponissos-Nissi Ioniou	*0.009*	*0.009*
Greece	Makedonia-Thraki	0.025	0.027
Greece	Sterea Ellas-Nissi Egaeou-Kriti	*0.013*	0.016
Greece	Thessalia	0.019	0.022
Hungary	Alföld	0.053	0.056
Hungary	Dunántúl	0.124	0.116
Hungary	Észak-Magyarország	0.075	0.069
Ireland	Ireland	0.090	0.087
Italy	Abruzzo	0.033	0.030
Italy	Alto Adige	0.015	*0.010*
Italy	Basilicata	0.046	0.040
Italy	Calabria	0.019	0.021
Italy	Campania	0.041	0.048
Italy	Emilia-Romagna	0.093	0.091
Italy	Friuli-Venezia Giulia	0.090	0.058
Italy	Lazio	0.039	0.043
Italy	Liguria	0.071	0.072
Italy	Lombardia	0.094	0.079
Italy	Marche	0.044	0.047
Italy	Molise	0.033	0.032
Italy	Piemonte	0.075	0.061
Italy	Puglia	0.040	0.048
Italy	Sardegna	0.031	0.026
Italy	Sicilia	0.031	0.029
Italy	Toscana	0.051	0.057
Italy	Trentino	0.019	0.021
Italy	Umbria	0.027	0.028
Italy	Valle d'Aosta	0.025	*0.001*
Italy	Veneto	0.058	0.054
Latvia	Latvia	0.101	0.112

(continued)

Table 5.1 (continued)

Member state	Region	Year	
		2019	2020
Lithuania	Lithuania	0.091	0.101
Luxembourg	Luxembourg	0.145	0.145
Netherlands	The Netherlands	0.108	0.111
Poland	Malopolska i Pogórze	0.020	0.023
Poland	Mazowsze i Podlasie	0.031	0.032
Poland	Pomorze i Mazury	0.098	0.097
Poland	Wielkopolska and Slask	0.076	0.074
Portugal	Açores e Madeira	0.027	0.032
Portugal	Alentejo e Algarve	0.031	0.029
Portugal	Norte e Centro	*0.006*	*0.008*
Portugal	Ribatejo e Oeste	0.034	0.046
Romania	Bucuresti-Ilfov	0.041	0.028
Romania	Centru	*0.011*	*0.014*
Romania	Nord-Est	0.024	0.024
Romania	Nord-Vest	*0.011*	0.015
Romania	Sud-Est	0.049	0.048
Romania	Sud-Muntenia	0.041	0.044
Romania	Sud-Vest-Oltenia	*0.014*	*0.015*
Romania	Vest	0.026	0.031
Slovakia	Slovakia	**0.600**	**0.557**
Slovenia	Slovenia	*0.008*	*0.010*
Spain	Andalucía	0.077	0.075
Spain	Aragón	0.098	0.098
Spain	Asturias	*0.003*	*0.003*
Spain	Canarias	0.099	0.101
Spain	Cantabria	*0.000*	*0.000*
Spain	Castilla y León	0.102	0.118
Spain	Castilla-La Mancha	0.057	0.057
Spain	Cataluña	0.054	0.059
Spain	Comunidad Valenciana	0.037	0.055
Spain	Extremadura	0.045	0.053

(continued)

Table 5.1 (continued)

Member state	Region	Year	
		2019	2020
Spain	Galicia	0.016	0.018
Spain	Islas Baleares	0.035	0.046
Spain	La Rioja	0.084	0.110
Spain	Madrid	0.030	0.033
Spain	Murcia	0.065	0.085
Spain	Navarra	0.109	0.108
Spain	País Vasco	0.046	0.048
Sweden	Län i norra Sverige	0.108	0.070
Sweden	Skogsoch mellanbygdslän	0.107	0.106
Sweden	Slättbyggdslän	0.191	0.217

Note Bold corresponds to the highest values and italic to the lowest

5.3 Findings Obtained

Linear, CHAID, neural network, random trees and C&R tree are the models with the highest accuracy for the training set, over the period 2019–2020 (Tables 5.2 and 5.4). The pertinence of these models to predict the fertiliser costs is also highlighted by the relationships among the observed values and the predicted ones presented in Figs. 5.2 and 5.3.

Cereals output and area, as well as seeds and plants costs, appear in the two years considered (2019 and 2020) between the most important predictors of the fertiliser costs in the representative farms of the European Union agricultural regions (Tables 5.3 and 5.5).

Table 5.2 Models with the highest accuracy (the lowest relative error) for the fertiliser costs of the European Union farming regions, with data at the farm level, for the year 2019

Model	Build time	Correlation	No. fields	Relative error
Linear	< 1	1.000	1	0.000
CHAID	< 1	1.000	1	0.000
Neural net	< 1	1.000	171	0.001
Random trees	< 1	0.960	178	0.084
C&R tree	< 1	0.952	29	0.101

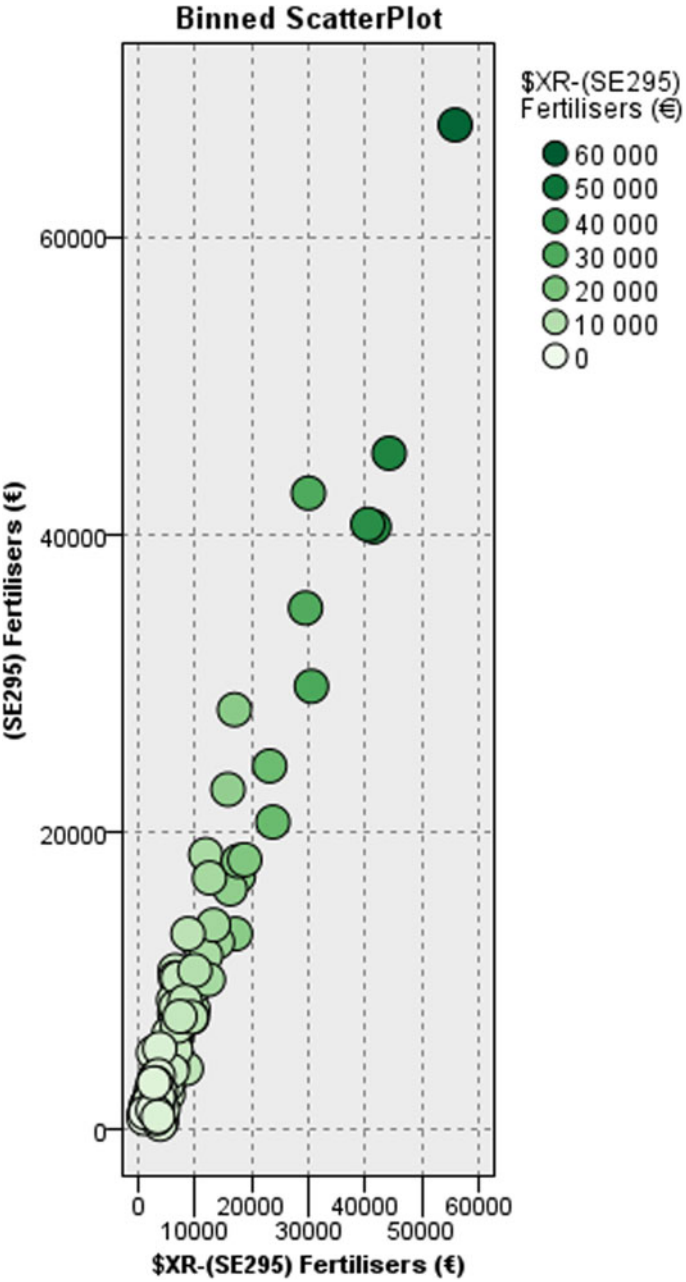

Fig. 5.2 Relationships between the observed values and the predicted ones for the fertiliser costs of the European Union farming regions, with data at the farm level, for the year 2019

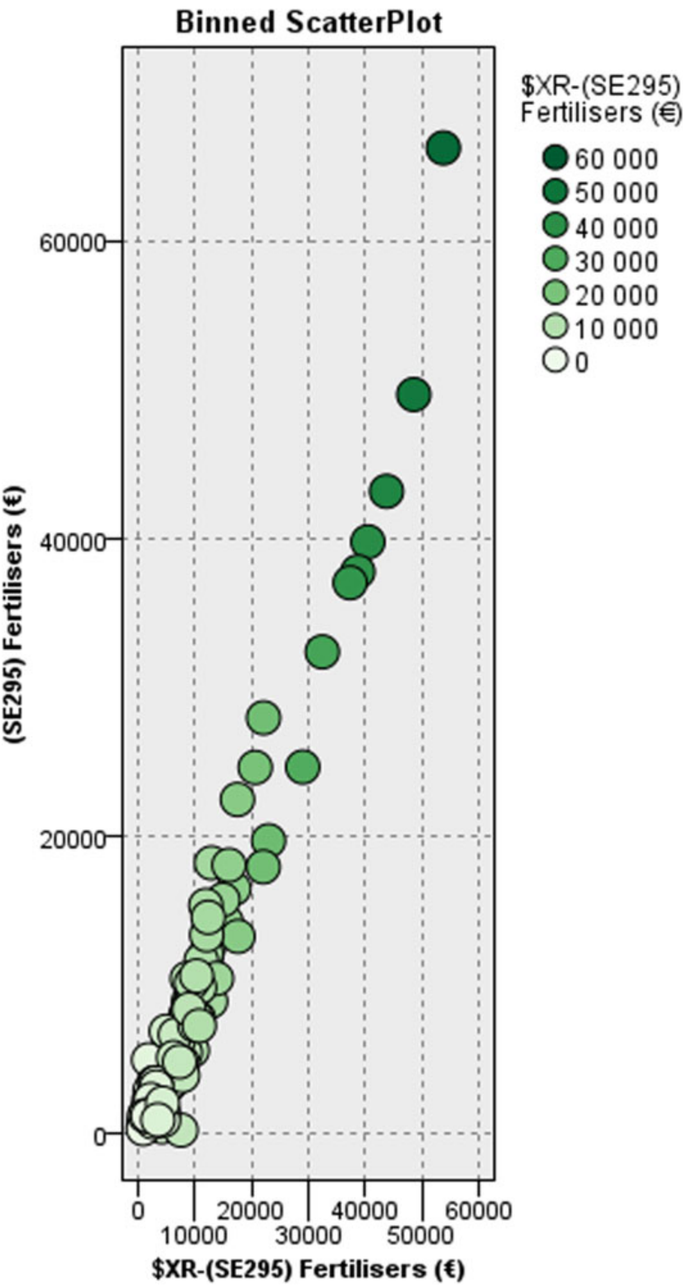

Fig. 5.3 Relationships between the observed values and the predicted ones for the fertiliser costs of the European Union farming regions, with data at the farm level, for the year 2020

Table 5.3 Importance of the predictors for the fertiliser costs of the European Union farming regions, with data at the farm level, for the year 2019

Nodes	Importance
Fruit (excl. citrus and grapes) (€/farm)	0.0063
Seeds and plants (€)	0.0064
Cereals (€/farm)	0.0065
Total farming overheads (€)	0.0065
Cereals (ha)	0.0065
Vegetables and flowers (ha)	0.0068
Total utilised agricultural area (ha)	0.0073
Cattle dairy cows (LU)	0.0084
Milk yield (kg/cow)	0.0322
Yield of wheat (q/ha)	0.0528

Table 5.4 Models with the highest accuracy (the lowest relative error) for the fertiliser costs of the European Union farming regions, with data at the farm level, for the year 2020

Model	Build time	Correlation	No. fields	Relative error
CHAID	< 1	1.000	4	0.000
Neural net	< 1	1.000	169	0.001
Linear	< 1	0.998	22	0.004
Random trees	< 1	0.959	178	0.086
C&R tree	< 1	0.907	34	0.178

Table 5.5 Importance of the predictors for the fertiliser costs of the European Union farming regions, with data at the farm level, for the year 2020

Nodes	Importance
Other field crops (ha)	0.0068
Cereals (ha)	0.0069
Stock of agricultural products (€)	0.0070
VAT balance excluding on investments (€)	0.0072
Seeds and plants (€)	0.0072
Vegetables and flowers (€/farm)	0.0074
Subsidies on intermediate consumption (€)	0.0080
Pigmeat (€/farm)	0.0081
Cereals (€/farm)	0.0128
Permanent crops (ha)	0.0364

5.4 Discussion and Conclusions

Between the different production factors used in the agricultural sector, fertilisers appear among the most critical, considering their impacts on the farming costs and their repercussions on the environmental and social conditions. In fact, the application of fertilisers in agricultural production has implications on the water quality, soil characteristics and consequently on human health. The efficient use of this production factor is fundamental to reduce the costs in the farms and to mitigate, particularly, the environmental impacts. These aspects are principally important in the current contexts of climate change and global warming. The solutions associated with artificial intelligence may bring relevant contributions to better deal with these challenges. In this way, this chapter taken into account artificial intelligence approaches to predict fertiliser costs in the European Union farming sector, following the procedures proposed by innovative solutions. For that, data from European Union databases, with microeconomic statistical information, were considered for the period 2019–2020.

The new methodologies may contribute significantly to improve the sustainability of the use of fertilisers and manure application, namely with more efficient practices and processes. Improvements in the efficiency of the practices associated with the use of fertilisers in the farms are especially important, considering the interrelationships among the fertilisers' application and the use of other production factors, such as energy, water, machinery and crop protection products.

Some European Union countries/regions, such as Czechia, Slovakia and some regions from France and Germany are between the contexts with higher values for fertiliser costs. Some regions from Croatia, Greece, Portugal, Romania and Spain, for example, are among the frameworks with the lowest costs of fertilisers.

Some of the most accurate models to predict the fertiliser costs in the European Union farms are, for example, the following: linear; CHAID; neural network; random trees and C&R tree. The most important predictors are the following: cereals output and area; seeds and plants costs.

In terms of practical implications, the findings highlight that the choices of the farmers about the agricultural productions to implement in the farms, as well as their dimensions, are interrelated with the prediction of the fertiliser costs. For policy recommendation, it may be important to improve the interlinkages of the policy instruments with the fertiliser costs, because in the research here carried out the policy measures designed in the framework of the Common Agricultural Policy do not appear among the most important predictors. For future research, it is suggested to assess the real impact of the farm dimension on the level of fertiliser costs.

Acknowledgements This work is funded by National Funds through the FCT—Foundation for Science and Technology, I.P., within the scope of the project Refª UIDB/00681/2020 (https://doi.org/https://doi.org/10.54499/UIDP/00681/2020). This research is also funded by the Enovo company. This study was carried out under the international project "Agriculture 4.0: Current reality, potentialities and policy proposals" (CERNAS-IPV/2022/008). Furthermore, we would like to thank the CERNAS Research Centre and the Polytechnic Institute of Viseu for their support. This work is too co-financed by the PRR—Plano de Recuperação e Resiliência (República Portuguesa) and the European Next Generation EU Funds (https://recuperarportugal.gov.pt) through application

PRR-C05-i03-I-000030—"Carb2Soil—Reforçar a Complementaridade entre agricultura e pecuária para aumentar a fertilidade dos solos e a sua capacidade de sequestro de carbono".

References

1. R.P. Sishodia, R.L. Ray, S.K. Singh, Applications of remote sensing in precision agriculture: a review. Remote Sens. **12**, 3136 (2020)
2. A. Subeesh, C.R. Mehta, Automation and digitization of agriculture using artificial intelligence and internet of things. Artif. Intell. Agric. **5**, 278 (2021)
3. G.F. Albaaji, S.S.V. Chandra, Artificial intelligence SoS framework for sustainable agricultural production. Comput. Electron. Agric. **213**, 108182 (2023)
4. B. Swaminathan, S. Palani, S. Vairavasundaram, K. Kotecha, V. Kumar, IoT-driven artificial intelligence technique for fertilizer recommendation model. IEEE Consum. Electron. Mag. **12**, 109 (2023)
5. D. Ather, S. Madan, M. Nayak, R. Tripathi, R. Kant, S.S. Kshatri, R. Jain, Selection of smart manure composition for smart farming using artificial intelligence technique. J. Food Qual. **2022**, 4351825 (2022)
6. M.A. Dayioglu, U. Turker, Digital transformation for sustainable future—agriculture 4.0: a review. J. Agric. Sci. Tarim. Bilim. Derg. **27**, 373 (2021)
7. M.K. Senapaty, A. Ray, N. Padhy, IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture. Computers **12**, 61 (2023)
8. K. Demestichas, E. Daskalakis, Data lifecycle management in precision agriculture supported by information and communication technology. Agron. Basel **10**, 1648 (2020)
9. E. Elbasi et al., Artificial intelligence technology in the agricultural sector: a systematic literature review. IEEE Access **11**, 171 (2023)
10. P. Wang, K. You, Y.H. Ong, J.N. Yeoh, J.P.Q. Ong, A.T.L. Truong, A. Blasiak, E.K.-H. Chow, D. Ho, WisDM green: harnessing artificial intelligence to design and prioritize compound combinations in peat moss for sustainable farming applications. Adv. Intell. Syst. **4**, 2200095 (2022)
11. D.F. Yepez-Ponce, J.V. Salcedo, P.D. Rosero-Montalvo, J. Sanchis, Mobile robotics in smart farming: current trends and applications. Front. Artif. Intell. **6**, 1213330 (2023)
12. M.A. Wani, A. Din, I.T. Nazki, T.U.U. Rehman, J.M.M. Al-Khayri, S.M. Jain, R.A. Lone, Z.A. Bhat, M. Mushtaq, Navigating the future: exploring technological advancements and emerging trends in the sustainable ornamental industry. Front. Environ. Sci. **11**, 1188643 (2023)
13. R. Singh, R. Singh, A. Gehlot, S.V. Akram, N. Priyadarshi, B. Twala, Horticulture 4.0: adoption of industry 4.0 technologies in horticulture for meeting sustainable farming. Appl. Sci. Basel **12**, 12557 (2022)
14. C. Ganeshkumar, S.K. Jena, A. Sivakumar, T. Nambirajan, Artificial intelligence in agricultural value chain: review and future directions. J. Agribus. Dev. Emerg. Econ. **13**, 379 (2023)
15. Y. Zhang, X. Wang, T. Liu, R. Wang, Y. Li, Q. Xue, P. Yang, Sustainable fertilisation management via tensor multi-task learning using multi-dimensional agricultural data. J. Ind. Inf. Integr. **34**, 100461 (2023)
16. S.I. Hassan, M.M. Alam, U. Illahi, M.A. Al-Ghamdi, S.H. Almotiri, M.M. Su'ud, A systematic review on monitoring and advanced control strategies in smart agriculture. IEEE Access **9**, 32517 (2021)
17. R. Lal, The role of industry and the private sector in promoting the "4 per 1000" initiative and other negative emission technologies. Geoderma **378**, 114613 (2020)
18. C. Maraveas, C.-S. Karavas, D. Loukatos, T. Bartzanas, K.G. Arvanitis, E. Symeonaki, Agricultural greenhouses: resource management technologies and perspectives for zero greenhouse gas emissions. Agricult. Basel **13**, 1464 (2023)

19. A. Siedliska, P. Baranowski, J. Pastuszka-Wozniak, M. Zubik, J. Krzyszczak, Identification of plant leaf phosphorus content at different growth stages based on hyperspectral reflectance. BMC Plant Biol. **21**, 28 (2021)
20. M.A. Hossen, P.K. Diwakar, S. Ragi, Total nitrogen estimation in agricultural soils via aerial multispectral imaging and LIBS. Sci. Rep. **11**, 12693 (2021)
21. A. Kaab, M. Sharifi, H. Mobli, A. Nabavi-Pelesaraei, K. Chau, Combined life cycle assessment and artificial intelligence for prediction of output energy and environmental impacts of sugarcane production. Sci. Total. Environ. **664**, 1005 (2019)
22. G. Sajith, R. Srinivas, A. Golberg, J. Magner, Bio-inspired and artificial intelligence enabled hydro-economic model for diversified agricultural management. Agric. Water Manag. **269**, 107638 (2022)
23. V.J.P.D. Martinho, Fertiliser cost prediction in European union farms: machine-learning approaches through artificial neural networks. Open Agricult. **8**, 20220191 (2023)
24. J. Kim, S. Kim, C. Ju, H.I. Son, Unmanned aerial vehicles in agriculture: a review of perspective of platform, control, and applications. IEEE Access **7**, 105100 (2019)
25. FADN, Several Statistics. https://agriculture.ec.europa.eu/data-and-analysis/farm-structures-and-economics/fadn_en
26. IBM SPSS Modeler, SPSS Modeler: Overview. https://www.ibm.com/products/spss-modeler
27. V.J.P.D. Martinho, Main factors that explain the use of fertilisers on farms in the European Union: contributions to a more sustainable development, in *Impacts of Climate Change and Economic and Health Crises on the Agriculture and Food Sectors* (IGI Global, 2022), pp. 155–184

Chapter 6
Important Indicators for Predicting Crop Protection Costs

Abstract The crop protection costs have economic impacts on the profitability of the farms and environmental consequences due to, in some circumstances, the residues that remain in the soils after the application. The crop protection application may have also direct impacts on human health, because of the residues which remain in the agricultural products, particularly when applied in a non-efficient way. The Common Agricultural Policy in the European Union has already a set of measures to encourage farmers to reduce the level of crop protection application in farming activities. In any case, it is important to bring more insights into these contexts, specifically identifying the most important predictors of crop protection costs in the European Union farms. To achieve these objectives, this study takes into account approaches from the new technologies associated with the digital transition and data from the European Union Farm Accountancy Data Network. The insights obtained allowed us to highlight the most adjusted models and the most important variables to predict crop protection costs in European agriculture.

Keywords Digital transition · European Union farm accountancy data network · Common agricultural policy

6.1 Introduction

Crop protection application in agricultural production is generally a concern for farmers, public institutions, associations and policymakers, because of the potential impacts of the associated products on the environment and human health [1]. The policymakers from national, European and International institutions are concerned with these products and have designed legislation to control the use of these applications in quality and quantity, to mitigate the consequences of these practices.

The efficient use of crop protection products in farms is one way to reduce the impacts of these activities on sustainability [2]. For that, the approaches available in the framework of the digital transformation may contribute significantly to allow increases in the productivity with lower negative effects, particularly supporting more

V. J. P. D. Martinho, *Machine Learning Approaches for Evaluating Statistical Information in the Agricultural Sector*, SpringerBriefs in Applied Sciences and Technology, https://doi.org/10.1007/978-3-031-54608-2_6

adjusted disease diagnosis [3] and management [4]. For an effective application of the new technologies, information collection is a fundamental phase, including through imagery [5]. The same happens with the availability of data [6] and the combination of artificial intelligence methodologies with other emergent approaches [7] in the diverse dimensions of crop protection [8, 9]. Some of the data considered in the analysis based on artificial intelligence techniques are obtained from databases, such as Sentinel-2 [10], or through sensors [11], for example.

The following artificial intelligence techniques have been considered by the scientific community to identify crop biotic stresses [12]: random forest; support vector machine; decision tree; Naive Bayes; convolutional neural network; long short-term memory; deep convolutional neural network and deep belief network. The focus of some research is to reduce the resources and expertise that usually these approaches need [13] for pest detection [14]. The detection of the biotic threats for agricultural production, supported by the new technologies, is relevant for adjusted farm management [15].

The current contexts associated with climate change and the consequent global warming create new challenges for crop protection practices and for the efficacy of the respective products to mitigate the impacts of the different pathogen agents [16]. For an adjusted application of crop protection products, the prediction of plant pathogen agents is another crucial step [17] to support the farmers' decisions [18], including in grapevine [19]. Weed identification and mapping are also important for a more efficient use of the required applications [20]. Early detection of biotic farming threats is fundamental for profitable farm management [21], including resistant weeds [22].

The aim of this study is to present adjusted models and important predictors to assess the crop protection costs in the European Union farms, through data from the Farm Accountancy Data Network [23] and taking into account the procedures proposed by the software IBM SPSS Modeler [24].

6.2 Data Examination

On average over the period 2018–2021, the crop protection costs [crop protection (€)] were higher in the representative farms of the following European Union countries (Fig. 6.1): Slovakia; Czechia; Denmark; Netherlands; France; Germany; Belgium; Estonia; Luxembourg and Sweden.

With the statistical information disaggregated at the level of the European Union agricultural regions, Table 6.1 reveals that the averages for the crop protection costs, over the period 2018–2021, were bigger in Czechia, Slovakia and regions from France and Germany. The lowest averages appear in agricultural regions from Croatia, Greece, Italy, Poland, Portugal, Romania and Spain.

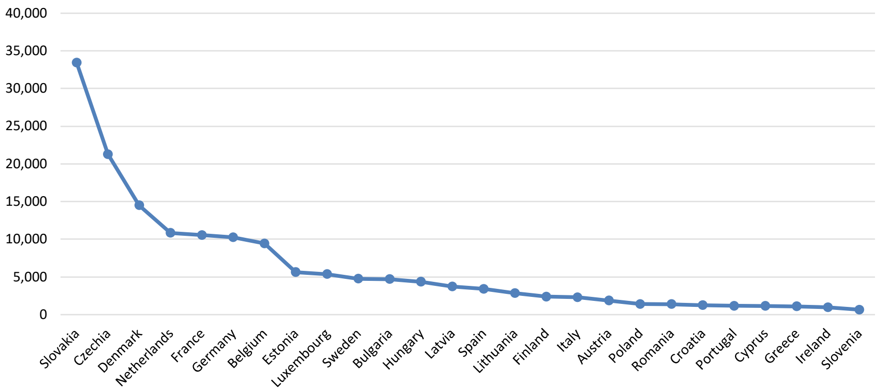

Fig. 6.1 Average values for the crop protection costs of the European Union countries, with data at the farm level, over the period 2018–2021

Table 6.1 Average values for the crop protection costs of the European Union farming regions, with data at the farm level, over the period 2018–2021

Member state	Region	Average
Austria	Austria	1870
Belgium	Vlaanderen	10,489
Belgium	Wallonie	7693
Bulgaria	Severen tsentralen	7901
Bulgaria	Severoiztochen	7235
Bulgaria	Severozapaden	8440
Bulgaria	Yugoiztochen	5051
Bulgaria	Yugozapaden	1224
Bulgaria	Yuzhen tsentralen	1958
Croatia	Jadranska Hrvatska	*619*
Croatia	Kontinentalna Hrvatska	1489
Cyprus	Cyprus	1153
Czechia	Czechia	**21,286**
Denmark	Denmark	14,503
Estonia	Estonia	5634
Finland	Etelä-Suomi	2923
Finland	Pohjanmaa	2816
Finland	Pohjois-Suomi	1050
Finland	Sisä-Suomi	1077
France	Alsace	6276
France	Aquitaine	8159

(continued)

Table 6.1 (continued)

Member state	Region	Average
France	Auvergne	2694
France	Basse-Normandie	8542
France	Bourgogne	11,775
France	Bretagne	6976
France	Centre	20,653
France	Champagne-Ardenne	13,005
France	Corse	3126
France	Franche-Comté	6445
France	Guadeloupe	2095
France	Haute-Normandie	**24,091**
France	Île-de-France	**26,576**
France	La Réunion	1983
France	Languedoc-Roussillon	9958
France	Limousin	2797
France	Lorraine	11,954
France	Midi-Pyrénées	7476
France	Nord-Pas-de-Calais	17,854
France	Pays de la Loire	8959
France	Picardie	**26,645**
France	Poitou–Charentes	14,514
France	Provence-Alpes-Côte d'Azur	7577
France	Rhône-Alpes	5408
Germany	Baden-Württemberg	6723
Germany	Bayern	5140
Germany	Brandenburg	**29,854**
Germany	Hessen	8400
Germany	Mecklenburg-Vorpommern	**56,154**
Germany	Niedersachsen	10,004
Germany	Nordrhein-Westfalen	8260
Germany	Rheinland-Pfalz	7902
Germany	Saarland	5808
Germany	Sachsen	**29,007**
Germany	Sachsen-Anhalt	**41,741**
Germany	Schleswig–Holstein/Hamburg	11,354
Germany	Thüringen	**47,158**
Greece	Ipiros-Peloponissos-Nissi Ioniou	*593*
Greece	Makedonia-Thraki	1816

(continued)

Table 6.1 (continued)

Member state	Region	Average
Greece	Sterea Ellas-Nissi Egaeou-Kriti	*691*
Greece	Thessalia	1682
Hungary	Alföld	3237
Hungary	Dunántúl	6676
Hungary	Észak-Magyarország	4890
Ireland	Ireland	967
Italy	Abruzzo	1815
Italy	Alto Adige	2071
Italy	Basilicata	1524
Italy	Calabria	*520*
Italy	Campania	2254
Italy	Emilia-Romagna	5114
Italy	Friuli-Venezia Giulia	3511
Italy	Lazio	1217
Italy	Liguria	1985
Italy	Lombardia	4412
Italy	Marche	1948
Italy	Molise	1144
Italy	Piemonte	4381
Italy	Puglia	2150
Italy	Sardegna	792
Italy	Sicilia	896
Italy	Toscana	1933
Italy	Trentino	2548
Italy	Umbria	1673
Italy	Valle d'Aosta	*406*
Italy	Veneto	3773
Latvia	Latvia	3728
Lithuania	Lithuania	2854
Luxembourg	Luxembourg	5376
Netherlands	The Netherlands	10,845
Poland	Malopolska i Pogórze	*703*
Poland	Mazowsze i Podlasie	759
Poland	Pomorze i Mazury	3169
Poland	Wielkopolska and Slask	2391
Portugal	Açores e Madeira	*382*
Portugal	Alentejo e Algarve	1120

(continued)

Table 6.1 (continued)

Member state	Region	Average
Portugal	Norte e Centro	811
Portugal	Ribatejo e Oeste	4651
Romania	Bucuresti-Ilfov	2479
Romania	Centru	840
Romania	Nord-Est	1046
Romania	Nord-Vest	750
Romania	Sud-Est	2573
Romania	Sud-Muntenia	2722
Romania	Sud-Vest-Oltenia	1067
Romania	Vest	1218
Slovakia	Slovakia	**33,436**
Slovenia	Slovenia	*660*
Spain	Andalucía	4031
Spain	Aragón	4508
Spain	Asturias	*260*
Spain	Canarias	4994
Spain	Cantabria	*105*
Spain	Castilla y León	3130
Spain	Castilla-La Mancha	1994
Spain	Cataluña	4694
Spain	Comunidad Valenciana	3687
Spain	Extremadura	2530
Spain	Galicia	847
Spain	Islas Baleares	1465
Spain	La Rioja	4810
Spain	Madrid	1417
Spain	Murcia	6809
Spain	Navarra	4405
Spain	País Vasco	2756
Sweden	Län i norra Sverige	1469
Sweden	Skogsoch mellanbygdslän	2272
Sweden	Slättbyggdslän	6127

Note Bold corresponds to the highest values and italic to the lowest

6.3 Findings Identified

The models with the highest accuracy (for the training set) to predict the crop protection costs, on average over the period 2018–2021, are the following (Table 6.2): CHAID (Chi-squared Automatic Interaction Detection); XGBoost linear (advanced application of a gradient boosting algorithm with a linear model as the reference model); linear (linear regression); XGBoost tree (advanced application of a gradient boosting algorithm with a tree model as the reference model); neural net (neural network); random trees (multiple decision trees); C&R tree (Classification and Regression tree); random forest (algorithm with a tree model as the reference model); linear-AS (linear regression) and SVM (support vector machine).

The relationships among the observed values of the crop protection costs and those predicted presented in Fig. 6.2 show the predictive relevance of these models.

The most important predictors of the crop protection costs in the representative farms of the European Union agricultural regions are the following (Table 6.3): fertiliser K_2O (q); sugar beet (€/farm); energy (€); net investment on fixed assets (€); fertiliser N (q); economic size (€'000); seeds and plants (€); paid labour input (AWU); total output crops and crop production (€/farm) and fertilisers (€). These results show that the net investment on fixed assets, the economic size and the paid labour input, for example, may support the stakeholders in predicting the crop protection costs.

Table 6.2 Models with the highest accuracy (the lowest relative error) for the crop protection costs of the European Union farming regions, with data at the farm level on average over the period 2018–2021

Model	Build time	Correlation	No. fields	Relative error
CHAID	3	1.000	15	0.000
XGBoost linear	3	0.997	177	0.005
Linear	3	0.996	19	0.007
XGBoost tree	3	0.998	177	0.014
Neural net	3	0.989	168	0.025
Random trees	3	0.978	177	0.053
C&R tree	3	0.958	34	0.083
Random forest	3	0.944	177	0.123
Linear-AS	3	0.871	177	0.252
SVM	3	0.949	170	1.479

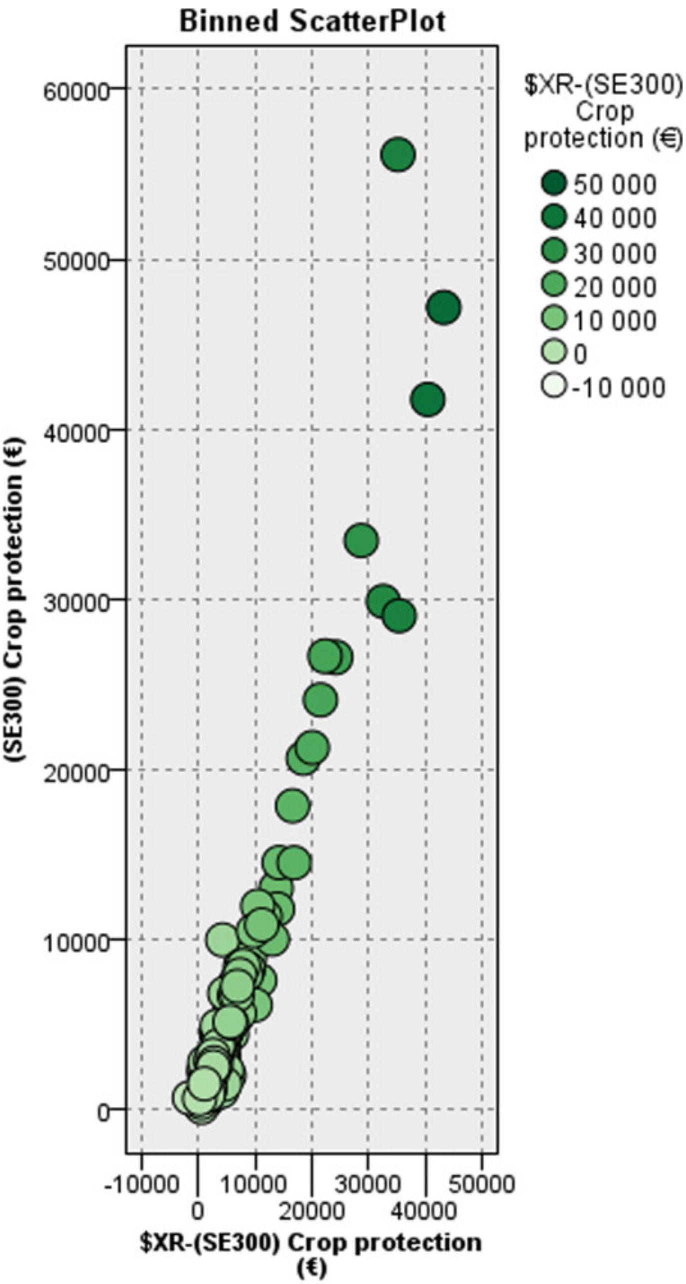

Fig. 6.2 Relationships between the observed values and the predicted ones for the crop protection costs of the European Union farming regions, with data at the farm level on average over the period 2018–2021

Table 6.3 Importance of the predictors for the crop protection costs of the European Union farming regions, with data at the farm level on average over the period 2018–2021

Nodes	Importance
Fertiliser K_2O (q)	0.0073
Sugar beet (€/farm)	0.0081
Energy (€)	0.0081
Net investment on fixed assets (€)	0.0093
Fertiliser N (q)	0.0098
Economic size (€'000)	0.0127
Seeds and plants (€)	0.0139
Paid labour input (AWU)	0.0235
Total output crops and crop production (€/farm)	0.0295
Fertilisers (€)	0.0518

6.4 Discussion and Conclusions

Crop protection products, jointly with fertilisers, are between the production factors used in the farms that more concerns bring to the national and international institutions related to the agricultural sector and the environmental conditions worldwide. In fact, crop protection products brought important contributions to the farming dynamics, namely to deal with biotic stresses, but also brought new challenges because of their potential impacts on the sustainability of agriculture and human health. The machine learning solutions offer new potentialities to improve the planning and management of the farms and, in this context, mitigate the negative implications of crop protection products use. Considering this scenario, the study here presented aimed to find accurate models to predict the costs with the use of crop protection products and identify the most important predictors. To achieve these aims, microeconomic statistical information from the Farm Accountancy Data Network was considered, on average, for the period 2018–2021. These data were analysed considering machine learning solutions, following the procedures proposed by the software IBM SPSS Modeler.

The new approaches available to collect information, namely those associated with the Sentinel-2, sensors and unmanned aerial vehicles, for example. This information collected through alternative solutions and new methodologies available to assess these data are crucial to support early detections of biotic farming stresses and, in this way, to better use and apply the crop protection products.

In the European Union context, the crop protection costs, on average over the period 2018–2021, were higher in the representative farms of Slovakia, Czechia, Denmark, Netherlands, France, Germany, Belgium, Estonia, Luxembourg and Sweden. The lowest averages appear, for example, in agricultural regions from Croatia, Greece, Poland, Portugal and Romania.

CHAID, XGBoost linear, linear, XGBoost tree, neural net, random trees, C&R tree, random forest, linear-AS, and SVM are the most accurate models to predict the crop protection products in the European Union representative farms. Fertiliser K_2O, sugar beet, energy, net investment on fixed assets, fertiliser N, economic size, seeds and plants, paid labour input, total output crops and crop production, and fertilisers are the most important predictors.

In terms of practical implications, the results obtained in this research suggest a set of important predictors that may provide relevant insights for crop protection costs prediction. For policy recommendation, it is suggested to improve the interlinkages of the European Union policy instruments and measures with the crop protection costs. In future investigations, it would be important to quantify the relationships of the most important predictors with the crop protection costs.

Acknowledgements This work is funded by National Funds through the FCT—Foundation for Science and Technology, I.P., within the scope of the project Refª UIDB/00681/2020 (https://doi.org/https://doi.org/10.54499/UIDP/00681/2020). This research is also funded by the Enovo company. This study was carried out under the international project "Agriculture 4.0: Current reality, potentialities and policy proposals" (CERNAS-IPV/2022/008). Furthermore, we would like to thank the CERNAS Research Centre and the Polytechnic Institute of Viseu for their support. This work is too co-financed by the PRR—Plano de Recuperação e Resiliência (República Portuguesa) and the European Next Generation EU Funds (https://recuperarportugal.gov.pt) through application PRR-C05-i03-I-000030—"Carb2Soil—Reforçar a Complementaridade entre agricultura e pecuária para aumentar a fertilidade dos solos e a sua capacidade de sequestro de carbono".

References

1. A. Indu, A.S. Baghel, A. Bhardwaj, W. Ibrahim, Optimization of pesticides spray on crops in agriculture using machine learning. Comput. Intell. Neurosci. **2022**, 9408535 (2022)
2. L. Li, L. Sun, J. Guo, S. Li, P. Jiang, Identification of crop diseases based on improved genetic algorithm and extreme learning machine. CMC-Comput. Mat. Contin. **65**, 761 (2020)
3. M. Kalpana, L. Karthiba, K. Senguttuvan, R. Parimalarangan, Diagnosis of major foliar diseases in black gram (*Vigna mungo* L.) using convolution neural network (CNN). Legume Res. **46**, 940 (2023)
4. L. Luo, Q. Chang, Q. Wang, Y. Huang, Identification and severity monitoring of maize dwarf mosaic virus infection based on hyperspectral measurements. Remote Sens. **13**, 4560 (2021)
5. A. Amrani, F. Sohel, D. Diepeveen, D. Murray, M.G.K. Jones, Insect detection from imagery using YOLOv3-based adaptive feature fusion convolution network. Crop Pasture Sci. **74**, 615 (2023)
6. R.H.L. Ip, L.-M. Ang, K.P. Seng, J.C. Broster, J.E. Pratley, Big data and machine learning for crop protection. Comput. Electron. Agric. **151**, 376 (2018)
7. G. Caballero-Vidal, C. Bouysset, J. Gevar, H. Mbouzid, C. Nara, J. Delaroche, J. Golebiowski, N. Montagne, S. Fiorucci, E. Jacquin-Joly, Reverse chemical ecology in a moth: machine learning on odorant receptors identifies new behaviorally active agonists. Cell. Mol. Life Sci. **78**, 6593 (2021)
8. K. Simmons, J. Kinney, A. Owens, D. Kleier, K. Bloch, D. Argentar, A. Walsh, G. Vaidyanathan, Comparative study of machine-learning and chemometric tools for analysis of in-vivo high-throughput screening data. J. Chem. Inf. Model. **48**, 1663 (2008)
9. B.R. Terlouw et al., MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. Nucl. Acids Res. **51**, D603 (2023)
10. I. Navrozidis, T. Alexandridis, D. Moshou, A. Haugommard, A. Lagopodi, Implementing sentinel-2 data and machine learning to detect plant stress in olive groves. Remote Sens. **14**, 5947 (2022)

11. A. Sterling, J.A. Di Rienzo, Prediction of South American leaf blight and disease-induced photosynthetic changes in rubber tree, using machine learning techniques on leaf hyperspectral reflectance. Plants-Basel **11**, 329 (2022)
12. M. Chithambarathanu, M.K. Jeyakumar, Survey on crop pest detection using deep learning and machine learning approaches. Multimed. Tools Appl. **82**, 42277 (2023)
13. B. Espejo-Garcia, I. Malounas, E. Vali, S. Fountas, Testing the suitability of automated machine learning for weeds identification. AI **2**, 34 (2021)
14. M. Hayashi, K. Tamai, Y. Owashi, K. Miura, Automated machine learning for identification of pest aphid species (Hemiptera: Aphididae). Appl. Entomol. Zool. **54**, 487 (2019)
15. M. Tannous, C. Stefanini, D. Romano, A deep-learning-based detection approach for the identification of insect species of economic importance. Insects **14**, 148 (2023)
16. K. Dietzel, D. Valle, N. Fierer, J.M. U'Ren, A. Barberan, Geographical distribution of fungal plant pathogens in dust across the United States. Front. Ecol. Evol. **7**, 304 (2019)
17. B. Gokulnath, U.G. Devi, A survey on plant disease prediction using machine learning and deep learning techniques. Inteligen. Artif. **23**, 136 (2020)
18. M.G. Hill, P.G. Connolly, P. Reutemann, D. Fletcher, The use of data mining to assist crop protection decisions on kiwifruit in New Zealand. Comput. Electron. Agric. **108**, 250 (2014)
19. I. Volpi, D. Guidotti, M. Mammini, S. Marchi, Predicting symptoms of downy mildew, powdery mildew, and gray mold diseases of grapevine through machine learning. Ital. J. Agrometeorol. Riv. Ital. Agrometeorol. **57**, 69–78 (2021)
20. C. Kaempfer, L. Ulber, C. Wellhausen, M. Pflanz, Weed detection and mapping for automatic application map generation in crop protection. J. Kult. **73**, 121 (2021)
21. R.M. Math, N. Dharwadkar, Early detection and identification of grape diseases using convolutional neural networks. J. Plant Dis. Prot. **129**, 521 (2022)
22. A. Shirzadifar, S. Bajwa, J. Nowatzki, J. Shojaeiarani, Development of spectral indices for identifying glyphosate-resistant weeds. Comput. Electron. Agric. **170**, 105276 (2020)
23. FADN, Several Statistics. https://agriculture.ec.europa.eu/data-and-analysis/farm-structures-and-economics/fadn_en
24. IBM SPSS Modeler, SPSS Modeler: Overview. https://www.ibm.com/products/spss-modeler

Chapter 7
The Most Adjusted Predictive Models for Energy Costs

Abstract Energy is one of the most important production factors in farms, considering its impact on the profitability of the agricultural sector, its relationship with sustainability and the need for a green transition in agriculture to deal with the challenges created by climate change and the consequent global warming. In the green transition, it is important to replace fossil fuel sources with renewable energies and, in these contexts, the agricultural sector may make a double contribution, producing renewable energy and using more sustainable sources for the different processes and activities in the farms. Taking into account these motivations, this chapter proposes to select the models with better accuracy and the most relevant variables to predict the energy costs in the European Union farming sector. For that, machine learning models were considered, as well as statistical information from European Union databases. This chapter presents useful contributions to better understand the contexts associated with energy cost prediction in European farms.

Keywords Digital Era approaches · Predictors · European Union

7.1 Introduction

Energy is between the most critical production factors in the industry [1]. This resource assumes also special relevance in the agricultural sector, because of its importance for the performance of the farms and its relationships with sustainability. In fact, agriculture consumes energy for its different activities, but may be too a source of renewable energies in diverse ways, from the production of energy crops [2] to the supply of biomass from the by-products of food production. In any case, the costs related to energy are usually relevant and influence significantly the profitability of the sector.

In this perspective, the assessment of the energy costs is fundamental and the algorithms associated with artificial intelligence may produce important added value in these frameworks [3], where improvements in efficiency are crucial [4] to reduce these costs [5]. In general, the energy costs management is a concern in any activity,

V. J. P. D. Martinho, *Machine Learning Approaches for Evaluating Statistical Information in the Agricultural Sector*, SpringerBriefs in Applied Sciences and Technology, https://doi.org/10.1007/978-3-031-54608-2_7

including, for example, the following: strategies implemented to deal with the environmental challenges [6]; traffic performance [7]; residential buildings [8]; farmer behaviours [9]; waste management [10]; digital transformation [11]; smart homes [12, 13]; home activities [14, 15]; smart cities [16]; machine learning applications [17]; deep learning approaches [18] and tools [19]; manufacturing management [20]; industrial organisations [21]; irrigation systems [22]; irrigation networks [23]; sewer structures [24]; water supply systems [25] and ship repair firms [26].

The framework described before highlights that there is a field to be addressed regarding the energy costs assessment in agriculture, namely in the European Union context. These analyses will be relevant to produce more insights for the farmers, policymakers, public institutions and other organisations related to the sector. Taking into account these motivations, this chapter proposes to identify the most important variables to predict the energy costs in European farms, considering adjusted models based on artificial intelligence methodologies.

To achieve these objectives, statistical information from European databases [27] was considered, as well as techniques proposed by smart solutions [28]. The data considered were obtained for variables at the farm level and are related to the farming accountancy and characterise the structural and economic dimensions of the sector at a micro level. These variables are associated with, for example, the dimension of the farms, characteristics, outputs, costs, economic results, labour, investment subsidies and income subsidies.

7.2 Data Study

Figure 7.1 presents, on average over the period 2019–2021, the energy costs for the representative farms of the European Union member states. The highest values appear for the following countries: Slovakia; Czechia; Netherlands; Germany; Denmark; Sweden; Belgium; Finland; Estonia; Luxembourg and France.

Table 7.1 reveals the results for the energy costs, on average over the period 2019–2021, for the representative farms of the European Union agricultural regions. These results confirm, in part, the results highlighted before in Fig. 7.1. In fact, Czechia, Netherlands, Slovakia and some regions of Germany have the highest averages for energy costs. Inversely, regions from Croatia, Greece, Poland, Portugal and Romania, for instance, show the lowest values for the energy cost averages.

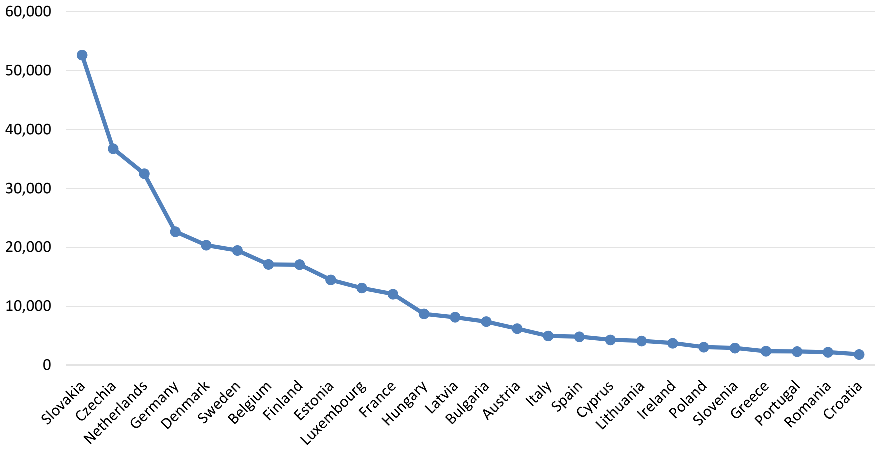

Fig. 7.1 Average values for the energy costs of the European Union countries, with data at the farm level, over the period 2019–2021

Table 7.1 Average values for the energy costs of the European Union farming regions, with data at the farm level, over the period 2019–2021

Member state	Region	Average
Austria	Austria	6190
Belgium	Vlaanderen	22,472
Belgium	Wallonie	8131
Bulgaria	Severen tsentralen	10,446
Bulgaria	Severoiztochen	11,102
Bulgaria	Severozapaden	10,905
Bulgaria	Yugoiztochen	8223
Bulgaria	Yugozapaden	3406
Bulgaria	Yuzhen tsentralen	4121
Croatia	Jadranska Hrvatska	*1239*
Croatia	Kontinentalna Hrvatska	2017
Cyprus	Cyprus	4289
Czechia	Czechia	**36,737**
Denmark	Denmark	20,354
Estonia	Estonia	14,474
Finland	Etelä-Suomi	15,760
Finland	Pohjanmaa	21,231
Finland	Pohjois-Suomi	17,697
Finland	Sisä-Suomi	15,077
France	Alsace	9658
France	Aquitaine	10,883

(continued)

Table 7.1 (continued)

Member state	Region	Average
France	Auvergne	9817
France	Basse-Normandie	14,594
France	Bourgogne	11,545
France	Bretagne	20,889
France	Centre	14,302
France	Champagne-Ardenne	8596
France	Corse	8334
France	Franche-Comté	12,663
France	Guadeloupe	3343
France	Haute-Normandie	15,108
France	Île-de-France	15,220
France	La Réunion	4554
France	Languedoc-Roussillon	6230
France	Limousin	8774
France	Lorraine	15,164
France	Midi-Pyrénées	9708
France	Nord-Pas-de-Calais	14,945
France	Pays de la Loire	16,642
France	Picardie	13,630
France	Poitou–Charentes	11,734
France	Provence-Alpes-Côte d'Azur	8965
France	Rhône-Alpes	10,375
Germany	Baden-Württemberg	13,841
Germany	Bayern	15,216
Germany	Brandenburg	**81,994**
Germany	Hessen	17,167
Germany	Mecklenburg-Vorpommern	**73,406**
Germany	Niedersachsen	**22,670**
Germany	Nordrhein-Westfalen	20,405
Germany	Rheinland-Pfalz	13,173
Germany	Saarland	16,216
Germany	Sachsen	**69,192**
Germany	Sachsen-Anhalt	**65,880**
Germany	Schleswig–Holstein/Hamburg	**25,824**
Germany	Thüringen	**83,953**
Greece	Ipiros-Peloponissos-Nissi Ioniou	*1821*
Greece	Makedonia-Thraki	3022

(continued)

Table 7.1 (continued)

Member state	Region	Average
Greece	Sterea Ellas-Nissi Egaeou-Kriti	2061
Greece	Thessalia	2991
Hungary	Alföld	7715
Hungary	Dunántúl	11,065
Hungary	Észak-Magyarország	8038
Ireland	Ireland	3727
Italy	Abruzzo	4316
Italy	Alto Adige	2473
Italy	Basilicata	4483
Italy	Calabria	2285
Italy	Campania	4391
Italy	Emilia-Romagna	7111
Italy	Friuli-Venezia Giulia	5598
Italy	Lazio	5973
Italy	Liguria	4066
Italy	Lombardia	11,666
Italy	Marche	4457
Italy	Molise	4659
Italy	Piemonte	6328
Italy	Puglia	3894
Italy	Sardegna	3840
Italy	Sicilia	3162
Italy	Toscana	5201
Italy	Trentino	2438
Italy	Umbria	4505
Italy	Valle d'Aosta	3770
Italy	Veneto	6597
Latvia	Latvia	8135
Lithuania	Lithuania	4101
Luxembourg	Luxembourg	13,090
Netherlands	The Netherlands	**32,494**
Poland	Malopolska i Pogórze	*1842*
Poland	Mazowsze i Podlasie	2245
Poland	Pomorze i Mazury	5452
Poland	Wielkopolska and Slask	4521
Portugal	Açores e Madeira	*1791*
Portugal	Alentejo e Algarve	3205

(continued)

Table 7.1 (continued)

Member state	Region	Average
Portugal	Norte e Centro	*1829*
Portugal	Ribatejo e Oeste	4011
Romania	Bucuresti-Ilfov	*1739*
Romania	Centru	*1847*
Romania	Nord-Est	*1475*
Romania	Nord-Vest	*1519*
Romania	Sud-Est	3660
Romania	Sud-Muntenia	3137
Romania	Sud-Vest-Oltenia	*1705*
Romania	Vest	2790
Slovakia	Slovakia	**52,632**
Slovenia	Slovenia	2917
Spain	Andalucía	3792
Spain	Aragón	7784
Spain	Asturias	3298
Spain	Canarias	5911
Spain	Cantabria	3578
Spain	Castilla y León	5878
Spain	Castilla-La Mancha	6586
Spain	Cataluña	6296
Spain	Comunidad Valenciana	2053
Spain	Extremadura	5320
Spain	Galicia	3448
Spain	Islas Baleares	5448
Spain	La Rioja	5052
Spain	Madrid	6988
Spain	Murcia	5128
Spain	Navarra	6470
Spain	País Vasco	4455
Sweden	Län i norra Sverige	17,143
Sweden	Skogsoch mellanbygdslän	16,339
Sweden	Slättbyggdslän	20,882

Note Bold corresponds to the highest values and italic to the lowest

7.3 Results Identified

Table 7.2 exhibits the most accurate models (for the training set) to predict the energy costs (on average over the period 2019–2021). These models are the following: generalised linear (enlarges the general linear model); Chi-squared Automatic Interaction Detection (CHAID); linear (linear regression); XGBoost linear (advanced application of a gradient boosting algorithm with a linear model as the reference model); XGBoost tree (advanced application of a gradient boosting algorithm with a tree model as the reference model); Classification and Regression tree (C&R tree); neural network (neural net); random forest (algorithm with a tree model as the reference); random trees (multiple decision trees) and linear-AS (linear regression). Figure 7.2 confirms the accuracy of these models in predicting the energy costs in the farms of the European Union agricultural regions.

The most important predictors of the energy costs, on average over the period 2019–2021, are the following (Table 7.3): poultry (LU); environmental subsidies (€); labour input (h); cereals (€/farm); vegetables and flowers (€/farm); machinery and building current costs (€); fertiliser P_2O_5 (q); total utilised agricultural area (ha); total labour input (AWU) and fertiliser N (q). The findings highlighted in Table 7.3 reveal the importance of the total area and labour input, for example, to predict the energy costs.

Table 7.2 Models with the highest accuracy (the lowest relative error) for the energy costs of the European Union farming regions, with data at the farm level on average over the period 2019–2021

Model	Build time	Correlation	No. fields	Relative error
Generalised linear	1	1.000	177	0.000
CHAID	1	1.000	13	0.000
Linear	1	0.999	24	0.002
XGBoost linear	1	0.999	177	0.002
XGBoost tree	1	0.999	177	0.012
C&R tree	1	0.985	47	0.033
Neural net	1	0.972	168	0.074
Random forest	1	0.968	177	0.111
Random trees	1	0.915	177	0.179
Linear-AS	1	0.946	177	7.667

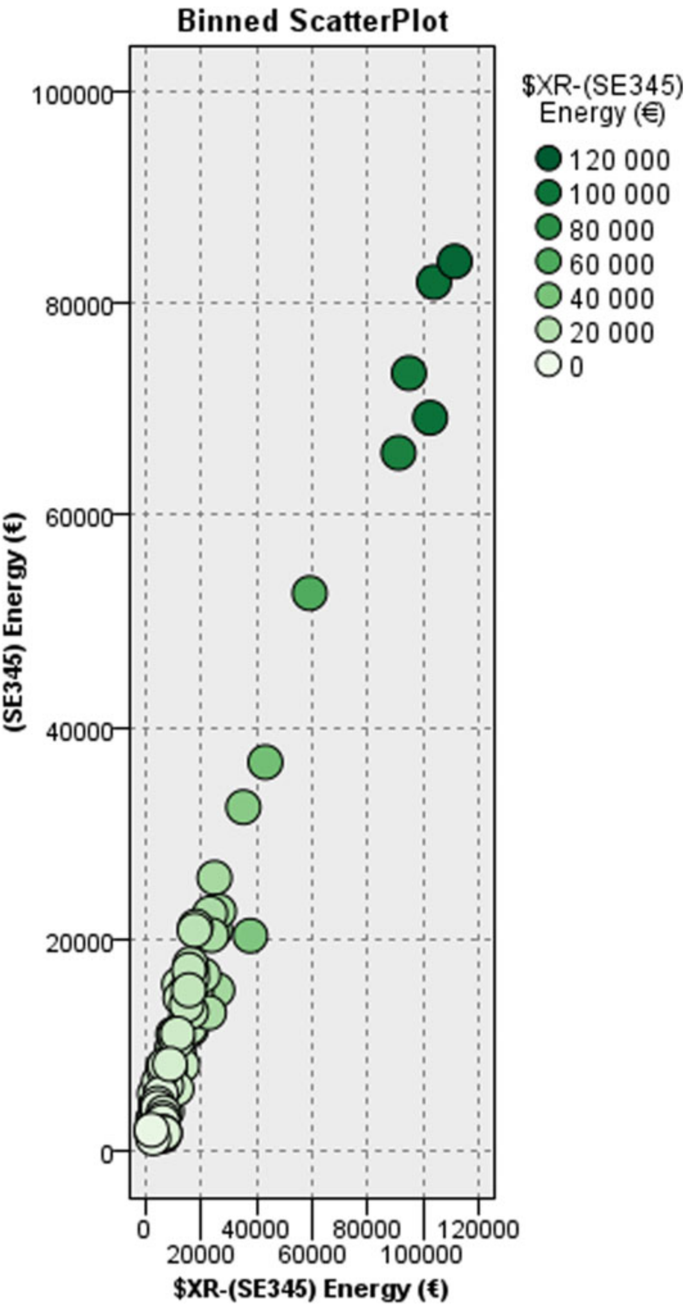

Fig. 7.2 Relationships between the observed values and the predicted ones for the energy costs of the European Union farming regions, with data at the farm level on average over the period 2019–2021

Table 7.3 Importance of the predictors for the energy costs of the European Union farming regions, with data at the farm level on average over the period 2019–2021

Nodes	Importance
Poultry (LU)	0.0071
Environmental subsidies (€)	0.0072
Labour input (h)	0.0074
Cereals (€/farm)	0.0075
Vegetables and flowers (€/farm)	0.0082
Machinery and building current costs (€)	0.0089
Fertiliser P_2O_5 (q)	0.0118
Total utilised agricultural area (ha)	0.0153
Total labour input (AWU)	0.0172
Fertiliser N (q)	0.0285

7.4 Discussion and Conclusions

The agricultural sector contributes to the energy context through consumption and production. The production of energy by farming activities is particularly important for sustainable development worldwide, considering the potentialities of the sector to have renewable sources. In fact, the agricultural sector may contribute to the production of energy through, for instance, energy crops (the sustainability here is questionable, due to the competition with food production) and biomass from farming waste and residues. In any case, energy use in agriculture is always a motive of concern for the farmers (because of the costs), for the policymakers and the decision-makers (due to the potential impacts on sustainability). In this framework, this chapter aimed to assess accurate models to predict the energy costs in the representative farms of the European Union countries and agricultural regions. Machine learning approaches proposed by the IBM SPSS Modeller were taken into account, as well as microeconomic data from European Union databases with data at the farm level for several indicators.

Energy is one of the most important production factors in the economic sectors and this is not an exception in agriculture, because of the associated costs, the potential impacts on the environment and their implications on the dynamics of economic activities. Artificial intelligence may support the farmers to use energy sources more efficiently, for example, in the irrigation systems, irrigation networks, water supply systems and environmental control in greenhouse production.

Slovakia, Czechia, Netherlands, Germany, Denmark, Sweden, Belgium, Finland, Estonia, Luxembourg and France are among the European Union countries with the highest values for energy costs, on average over the period 2019–2021. Contrariwise, farms from Croatia, Greece, Poland, Portugal and Romania reveal the lowest values for energy costs.

Generalised linear, CHAID, linear, XGBoost linear, XGBoost tree, C&R tree, neural net, random forest, random trees and linear-AS are between the models with the most predictive capacity to estimate the energy costs in the European

Union frameworks. Poultry, environmental subsidies, labour input, cereals, vegetables and flowers, machinery and building current costs, fertiliser P_2O_5, total utilised agricultural area, total labour input and fertiliser N.

In terms of practical implications, the findings here obtained suggest interrelationships between environmental subsidies, labour input, fertiliser use and total utilised agricultural area and energy costs. For policy recommendation, it would be important to better assess the interlinkages among environmental subsidies and energy costs. In future studies, it could be important to quantify the relationships among the most important predictors and energy costs.

Acknowledgements This work is funded by National Funds through the FCT–Foundation for Science and Technology, I.P., within the scope of the project Refª UIDB/00681/2020 (https://doi.org/https://doi.org/10.54499/UIDP/00681/2020). This research is also funded by the Enovo company. This study was carried out under the international project "Agriculture 4.0: Current reality, potentialities and policy proposals" (CERNAS-IPV/2022/008). Furthermore we would like to thank the CERNAS Research Centre and the Polytechnic Institute of Viseu for their support. This work is too co-financed by the PRR–Plano de Recuperação e Resiliência (República Portuguesa) and the European NextGeneration EU Funds (https://recuperarportugal.gov.pt) through application PRR-C05-i03-I-000030—"Carb2Soil—Reforçar a Complementaridade entre agricultura e pecuária para aumentar a fertilidade dos solos e a sua capacidade de sequestro de carbono."

References

1. X. Huang, S.H. Hong, M. Yu, Y. Ding, J. Jiang, Demand response management for industrial facilities: a deep reinforcement learning approach. IEEE Access **7**, 82194 (2019)
2. V.J.P.D. Martinho, Energy crops: assessments in the European Union agricultural regions through machine learning approaches. Reg. Sci. Inquiry **15**, 29 (2023)
3. N. Aksoy, I. Genc, Energy storage management for microgrids using n-step bootstrapping Gestion Du Stockage de l'energie Pour Les Micro-Reseaux a l'aide d'un Bootstrapping En n Etapes. IEEE Can. J. Electr. Comp. Eng. **46**, 107 (2023)
4. F. Bellizio, W. Xu, D. Qiu, Y. Ye, D. Papadaskalopoulos, J.L. Cremer, F. Teng, G. Strbac, Transition to digitalized paradigms for security control and decentralized electricity market. Proc. IEEE **111**, 744 (2023)
5. M.A. Hossain, A.R. Hossain, N. Ansari, AI in 6G: energy-efficient distributed machine learning for multilayer heterogeneous networks. IEEE Netw. **36**, 84 (2022)
6. K. Aliyon, F. Rajaee, J. Ritvanen, Use of artificial intelligence in reducing energy costs of a post-combustion carbon capture plant. Energy **278**, 127834 (2023)
7. B. Benjdira, A. Koubaa, A.T. Azar, Z. Khan, A. Ammar, W. Boulila, TAU: a framework for video-based traffic analytics leveraging artificial intelligence and unmanned aerial systems. Eng. Appl. Artif. Intell. **114**, 105095 (2022)
8. J.-S. Chou, T. Duc-Son, Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. Energy **165**, 709 (2018)
9. R. Gonzalez Perea, E. Camacho Poyato, J.A. Rodriguez Diaz, Forecasting of applied irrigation depths at farm level for energy tariff periods using coactive neuro-genetic fuzzy system. Agric. Water Manag. **256**, 107068 (2021)
10. J. Huang, D.D. Koroteev, Artificial intelligence for planning of energy and waste management. Sustain. Energy Technol. Assess. **47**, 101426 (2021)
11. H. Hyden, AI, norms, big data, and the law. Asian J. Law Soc. **7**, 409 (2020)

12. S.E.L. Jaouhari, E.J. Palacios-Garcia, A. Anvari-Moghaddam, A. Bouabdallah, Integrated management of energy, wellbeing and health in the next generation of smart homes. Sensors **19**, 481 (2019)
13. M. Nasir, K. Muhammad, A. Ullah, J. Ahmad, S.W. Baik, M. Sajjad, Enabling automation and edge intelligence over resource constraint IoT devices for smart home. Neurocomputing **491**, 494 (2022)
14. A. Kumari, S. Tanwar, A reinforcement-learning-based secure demand response scheme for smart grid system. IEEE Internet Things J. **9**, 2180 (2022)
15. A. Kumari, S. Tanwar, Multiagent-based secure energy management for multimedia grid communication using Q-learning. Multimed. Tools Appl. **81**, 36645 (2022)
16. J.R. Vazquez-Canteli, S. Ulyanin, J. Kampf, Z. Nagy, Fusing tensorflow with building energy simulation for intelligent energy management in smart cities. Sust. Cit. Soc. **45**, 243 (2019)
17. J. Kwon, D. Park, Hardware/software co-design for TinyML voice-recognition application on resource frugal edge devices. Appl. Sci. Basel **11**, 11073 (2021)
18. M. Li, Y. Wang, An energy-efficient silicon photonic-assisted deep learning accelerator for big data. Wirel. Commun. Mob. Comput. **2020**, 6661022 (2020)
19. T. Wang, S.-Y. Ma, L.G. Wright, T. Onodera, B.C. Richard, P.L. McMahon, An optical neural network using less than 1 photon per multiplication. Nat. Commun. **13**, 123 (2022)
20. T.-K. Liu, S.-S. Lin, P.-W. Hsueh, Optimal design for transport and logistics of steel mill by-product based on double-layer genetic algorithms. J. Low Freq. Noise Vib. Act. Control **40**, 555 (2021)
21. Z. Xu, G. Han, L. Liu, M. Martinez-Garcia, Z. Wang, Multi-energy scheduling of an industrial integrated energy system by reinforcement learning-based differential evolution. IEEE Trans. Green Commun. Netw. **5**, 1077 (2021)
22. M. Mohammed, H. Hamdoun, A. Sagheer, Toward sustainable farming: implementing artificial intelligence to predict optimum water and energy requirements for sensor-based micro irrigation systems powered by solar PV. Agron. Basel **13**, 1081 (2023)
23. R.G. Perea, I.F. Garcia, E.C. Poyato, J.A.R. Diaz, New memory-based hybrid model for middle-term water demand forecasting in irrigated areas. Agric. Water Manag. **284**, 108367 (2023)
24. S. Ostojin, S.R. Mounce, J.B. Boxall, An artificial intelligence approach for optimizing pumping in sewer systems. J. Hydroinform. **13**, 295 (2011)
25. J. Studzinski, A. Ziolkowski, Control of pumps of water supply network under hydraulic and energy optimisation using artificial intelligence. Entropy **22**, 1014 (2020)
26. X. Ren, Q. Jiang, J. Jiang, Z. He, B. Ouyang, B. Peng, Evaluation of cabin energy consumption based on combination weighting and grey fuzzy comprehensive model. EURASIP J. Adv. Signal Process. **2022**, 36 (2022)
27. FADN, Several Statistics. https://agriculture.ec.europa.eu/data-and-analysis/farm-structures-and-economics/fadn_en
28. IBM SPSS Modeler, SPSS Modeler: Overview. https://www.ibm.com/products/spss-modeler

Chapter 8
Machine Learning Methodologies, Wages Paid and the Most Relevant Predictors

Abstract The agricultural sector worldwide has an economic dimension related to the remuneration of the production factors applied in the sector, an environmental contribution associated with the sustainability of rural places and a social dimension related to the employment creation and the consequent level of remuneration of the labour. The question here is about the level of wages paid in the agricultural sector across the European Union countries and about the main factors that may be taken into account to predict the level of these wages paid to agricultural workers. This research intends to select the models with better precision to predict the wages paid in the European Union agriculture and to suggest important predictors from the enormous number of indicators that may be identified in the farms. The findings obtained may be considered relevant support for the design of social and agricultural policies in the European framework.

Keywords Artificial intelligence · Farm accountancy data network · European Union

8.1 Introduction

Agriculture is expected to absorb a small percentage of the total employment in developed countries. This is true, because, in these contexts, the farming sector is predictable to apply new agronomic practises (machinery, for example) and new technologies which allow to release of workers [1] to the other economic sectors, reducing the labour requirements [2].

In this way, agriculture is not an economic sector that contributes significantly to the levels of employment in economies more developed. Another question is about the level of the salaries paid on the farms. A relevant part of the workers who develop their activities in the agricultural sector are not qualified and this reality is reflected in the remuneration paid to this labour. In any case, there is already a problem of workforce scarcity for the diverse farming tasks in many situations [3], due to the increasing of old-age population [4].

V. J. P. D. Martinho, *Machine Learning Approaches for Evaluating Statistical Information in the Agricultural Sector*, SpringerBriefs in Applied Sciences and Technology, https://doi.org/10.1007/978-3-031-54608-2_8

It is important to have a real perception of the levels of employment in the agricultural sector and of the labour dynamics in related activities, where artificial intelligence solutions may contribute significantly [5]. Particularly, to allow for accurate predictions of diverse variables related to these issues, including yield estimations of farming productions [6], to solve the limitations of human labour [7], specifically the shortage of young adults in the workforce [8], and to better manage the human resources [9].

The artificial intelligence techniques were taken into account to address challenges related to farming labour in the following contexts, for example: soil moisture prediction [10]; meat characterisation [11]; harvesting in horticultural productions [12]; harvest time in tomato farms [13]; fruit identification [14]; rural e-commerce [15]; agro-food chains [16]; phenological characterisation of vegetable crops [17]; biotic stresses of grape seedlings [18]; aquaculture farm management [19]; plant production structures [20]; identification of ginger leaf diseases [21]; picking activities [22] and soil hydraulic conductivity [23].

The scarcity of workers for agricultural activities is a reality in many contexts and smart farming solutions bring interesting alternatives [24], namely in the associated labour-intensive practises [25]. Artificial intelligence may contribute to improve the levels of farming productivity and solving the problems of labour availability, but requires specific skills to manage these new technologies [26]. This may reinforce current inequalities in the dynamics of the farming sector [27].

The specific conditions of the human capital have implications on the dynamics of the sectors, including agriculture [28]. Artificial intelligence solutions may reintroduce new concerns about human health and this requires special attention from the national and international institutions with decisions on labour legislation [29].

The perspectives presented previously seem to show that the studies reviewed are more focused on the contributions of the new technologies to address the farming labour scarcity and management than on the remuneration of employment. Following these highlights, this research intends to bring more insights into the models and the variables adjusted to predict the wages paid in the European Union farms. For that, data from the Farm Accountancy Data Network [30] were considered, as well as the procedures proposed by smart solutions [31] and the results found by Martinho [32].

8.2 Data Analysis

The wages paid, on average over the period 2020–2021, were higher in the representative farms of European Union member states, such as Slovakia, Czechia, Denmark, Netherlands, Germany, Estonia, Sweden, France, Belgium, Spain, Finland, Luxembourg, Hungary, Bulgaria and Latvia. Ireland, Romania, Croatia, Poland, Greece and Slovenia frameworks are where the wages paid were lower, on average, over the period taken into account.

The results disaggregated at the agricultural region level and presented in Table 8.1, generally, confirm the contexts described before for the data aggregated for the European Union member states (Fig. 8.1).

Table 8.1 Average values for the wages paid in the European Union farming regions, with data at the farm level, over the period 2020–2021

Member state	Region	Average
Austria	Austria	3386
Belgium	Vlaanderen	20,867
Belgium	Wallonie	4942
Bulgaria	Severen tsentralen	17,292
Bulgaria	Severoiztochen	19,335
Bulgaria	Severozapaden	15,170
Bulgaria	Yugoiztochen	13,466
Bulgaria	Yugozapaden	4170
Bulgaria	Yuzhen tsentralen	6260
Croatia	Jadranska Hrvatska	2849
Croatia	Kontinentalna Hrvatska	*1674*
Cyprus	Cyprus	4045
Czechia	Czechia	**96,351**
Denmark	Denmark	**75,968**
Estonia	Estonia	23,342
Finland	Etelä-Suomi	11,978
Finland	Pohjanmaa	15,521
Finland	Pohjois-Suomi	14,907
Finland	Sisä-Suomi	11,148
France	Alsace	16,255
France	Aquitaine	21,493
France	Auvergne	3645
France	Basse-Normandie	10,895
France	Bourgogne	22,912
France	Bretagne	22,336
France	Centre	12,059
France	Champagne-Ardenne	18,454
France	Corse	24,515
France	Franche-Comté	7678
France	Guadeloupe	9265
France	Haute-Normandie	16,484
France	Île-de-France	23,302
France	La Réunion	12,896

(continued)

Table 8.1 (continued)

Member state	Region	Average
France	Languedoc-Roussillon	21,509
France	Limousin	4896
France	Lorraine	7799
France	Midi-Pyrénées	8790
France	Nord-Pas-de-Calais	17,342
France	Pays de la Loire	21,830
France	Picardie	11,104
France	Poitou–Charentes	14,627
France	Provence-Alpes-Côte d'Azur	38,040
France	Rhône-Alpes	17,067
Germany	Baden-Württemberg	13,625
Germany	Bayern	9424
Germany	Brandenburg	**208,967**
Germany	Hessen	10,865
Germany	Mecklenburg-Vorpommern	**171,144**
Germany	Niedersachsen	23,138
Germany	Nordrhein-Westfalen	20,319
Germany	Rheinland-Pfalz	20,876
Germany	Saarland	10,139
Germany	Sachsen	**192,063**
Germany	Sachsen-Anhalt	**144,769**
Germany	Schleswig–Holstein/Hamburg	22,003
Germany	Thüringen	**257,077**
Greece	Ipiros-Peloponissos-Nissi Ioniou	*1847*
Greece	Makedonia-Thraki	*1857*
Greece	Sterea Ellas-Nissi Egaeou-Kriti	1982
Greece	Thessalia	*1633*
Hungary	Alföld	11,591
Hungary	Dunántúl	16,605
Hungary	Észak-Magyarország	9476
Ireland	Ireland	2662
Italy	Abruzzo	3708
Italy	Alto Adige	7846
Italy	Basilicata	6523
Italy	Calabria	6689
Italy	Campania	6308
Italy	Emilia-Romagna	10,278

(continued)

Table 8.1 (continued)

Member state	Region	Average
Italy	Friuli-Venezia Giulia	6125
Italy	Lazio	10,049
Italy	Liguria	5518
Italy	Lombardia	10,451
Italy	Marche	2192
Italy	Molise	4074
Italy	Piemonte	5641
Italy	Puglia	8035
Italy	Sardegna	3228
Italy	Sicilia	7624
Italy	Toscana	9690
Italy	Trentino	3901
Italy	Umbria	6282
Italy	Valle d'Aosta	8110
Italy	Veneto	8550
Latvia	Latvia	10,266
Lithuania	Lithuania	4595
Luxembourg	Luxembourg	12,848
Netherlands	The Netherlands	**65,145**
Poland	Malopolska i Pogórze	*1003*
Poland	Mazowsze i Podlasie	*1061*
Poland	Pomorze i Mazury	3845
Poland	Wielkopolska and Slask	3390
Portugal	Açores e Madeira	2194
Portugal	Alentejo e Algarve	10,139
Portugal	Norte e Centro	2636
Portugal	Ribatejo e Oeste	4114
Romania	Bucuresti-Ilfov	*1885*
Romania	Centru	2646
Romania	Nord-Est	2541
Romania	Nord-Vest	1935
Romania	Sud-Est	3590
Romania	Sud-Muntenia	2858
Romania	Sud-Vest-Oltenia	*1332*
Romania	Vest	2287
Slovakia	Slovakia	**144,318**

(continued)

Table 8.1 (continued)

Member state	Region	Average
Slovenia	Slovenia	*426*
Spain	Andalucía	14,918
Spain	Aragón	23,105
Spain	Asturias	2328
Spain	Canarias	**48,317**
Spain	Cantabria	*1806*
Spain	Castilla y León	9593
Spain	Castilla-La Mancha	14,059
Spain	Cataluña	18,288
Spain	Comunidad Valenciana	12,331
Spain	Extremadura	14,188
Spain	Galicia	2982
Spain	Islas Baleares	11,224
Spain	La Rioja	26,209
Spain	Madrid	11,963
Spain	Murcia	26,824
Spain	Navarra	8769
Spain	País Vasco	6199
Sweden	Län i norra Sverige	13,313
Sweden	Skogsoch mellanbygdslän	12,836
Sweden	Slättbyggdslän	26,291

Note Bold corresponds to the highest values and italic to the lowest

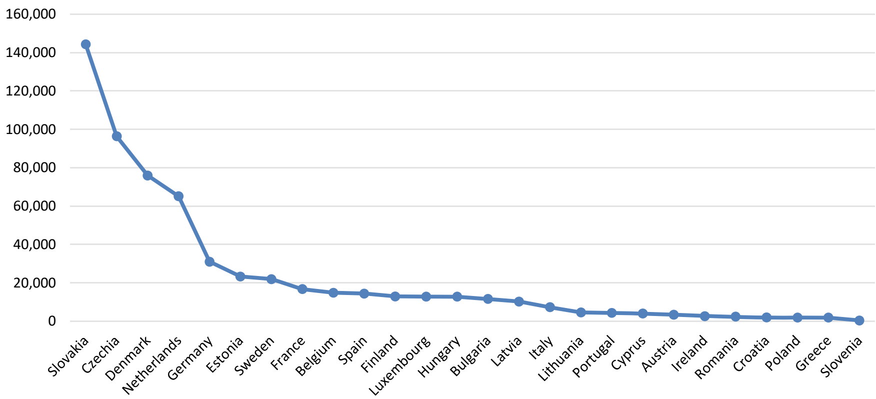

Fig. 8.1 Average values for the wages paid in the European Union countries, with data at the farm level, over the period 2020–2021

8.3 Core Results

The most accurate models (for the training set) to predict the wages paid (on average over the period 2020–2021) are displayed in Table 8.2. These models are, for example, the following: generalised linear (enlarges the general linear model); linear (linear regression); Chi-squared Automatic Interaction Detection (CHAID); XGBoost linear (advanced application of a gradient boosting algorithm with a linear model as the reference model) and XGBoost tree (advanced application of a gradient boosting algorithm with a tree model as the reference model).

The relationships between the observed values of the wage paid in the European Union agricultural regions and the predicted values for this variable, presented in Fig. 8.2, reveal the predictive pertinence of these models.

The most important predictors of the wage paid (on average over the period 2020–2021), in the representative farms of the European Union agricultural regions, are, in decreasing order, for example, the following (Table 8.3): forage crops (€/farm); economic size (€'000); farm net value added (€); total labour input (AWU); paid labour input (AWU) and net Investment on fixed assets (€).

Table 8.2 Models with the highest accuracy (the lowest relative error) for the wages paid in the European Union farming regions, with data at the farm level on average over the period 2020–2021

Model	Build time	Correlation	No. fields	Relative error
Generalised linear	1	1.000	177	0.000
Linear	1	1.000	32	0.000
CHAID	1	1.000	12	0.000
XGBoost linear	1	0.999	177	0.002
XGBoost tree	1	0.996	177	0.019
C&R tree	1	0.968	27	0.065
Random forest	1	0.976	177	0.072
Random trees	1	0.935	177	0.135
Linear-AS	1	0.968	177	0.805
SVM	1	0.891	168	1.024

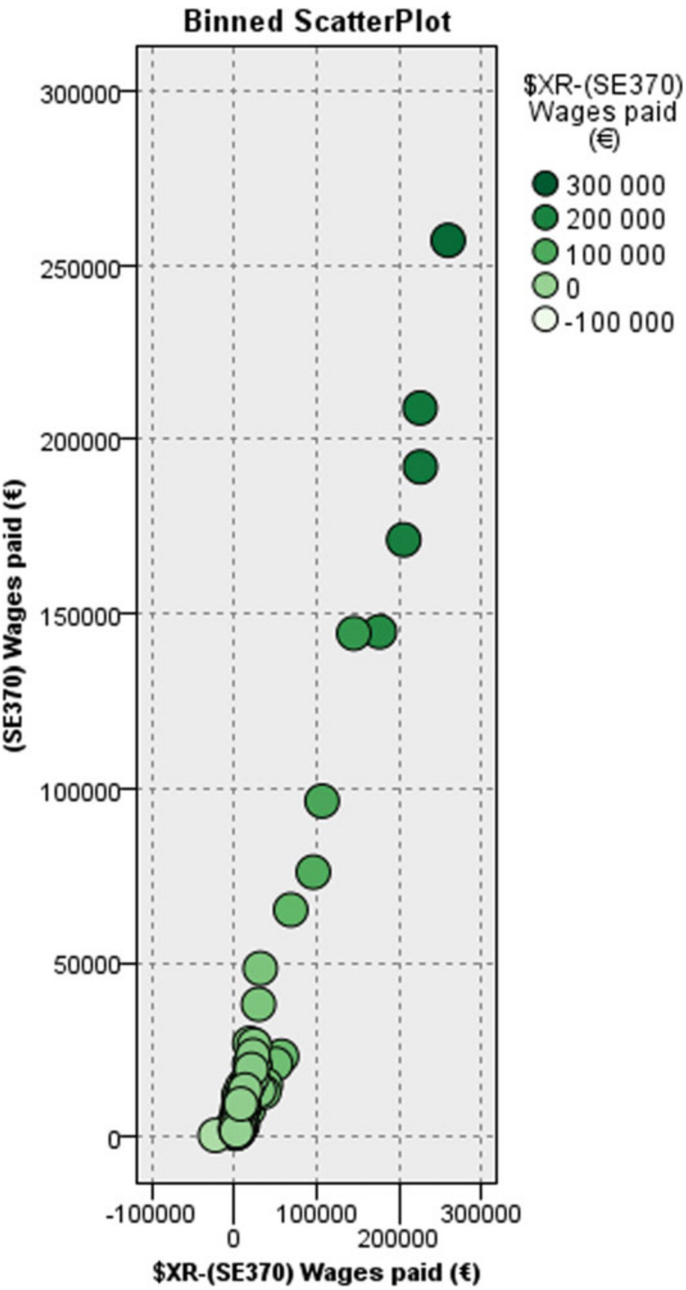

Fig. 8.2 Relationships between the observed values and the predicted ones for the wages paid in the European Union farming regions, with data at the farm level on average over the period 2020–2021

Table 8.3 Importance of the predictors for the wages paid in the European Union farming regions, with data at the farm level on average over the period 2020–2021

Nodes	Importance
Oil-seed crops (€/farm)	0.0089
Rented UAA (ha)	0.0097
Fertiliser N (q)	0.0103
Cash flow/farm total capital (€)	0.0108
Net investment on fixed assets (€)	0.0144
Paid labour input (AWU)	0.0151
Total labour input (AWU)	0.0170
Farm net value added (€)	0.0327
Economic size (€'000)	0.0334
Forage crops (€/farm)	0.0458

8.4 Discussion and Conclusions

The social dimension of agriculture is unquestionable worldwide, namely as a source of employment creation, despite the modernisation of the sector and the consequent release of labour to other sectors. The farming sector appears as an alternative source of employment, in some circumstances, in times of economic crises and unemployment increases, especially in the other sectors. The agricultural sector has also an important contribution in less favoured regions, where the economic dynamics are weaker and the opportunities for employment are scarcer. Nonetheless, the implicit question in these frameworks is about the level of wages paid by the farmers to their workers. Some of the labour used in the farms is unskilled and in these cases, the wages paid are expected to be lower. Considering these perspectives, this research intended to identify models with the highest accuracy to predict the wages paid by the representative farms from the European Union countries and agricultural regions, taking into account microeconomic data from the Farm Accountancy Data Network with data at the farm level. This statistical information was assessed through machine learning approaches and following the procedures proposed by the new solutions.

Digital smart approaches are expected to improve the efficiency of agricultural practises and processes and in this way increase farming profitability. These new contexts may contribute to pay better wages to agricultural workers and attracting more qualified labour. On the other hand, these alternative approaches may support the farmers to deal with the scarcity of labour for some of the farming tasks, particularly in cases where the old-age population increased and the young adults prefer other jobs.

On average, over the period 2020–2021, the wages paid were higher in the farms of European Union countries such as Slovakia, Czechia, Denmark, Netherlands, Germany, Estonia, Sweden, France, Belgium, Spain, Finland, Luxembourg, Hungary, Bulgaria and Latvia. The wages paid had the lowest values in the following member states: Ireland; Romania; Croatia; Poland; Greece and Slovenia.

Generalised linear, linear, CHAID, XGBoost linear and XGBoost tree are between the most accurate models to predict the wages paid in the European Union representative farms. Forage crops, economic size, farm net value added, total labour input, paid labour input and net Investment on fixed assets are the most important indicators to predict the wages paid.

In terms of practical implications, the economic size of the farms, economic results (such as farm net value added) and the level of investment are interrelated with the wages paid in the farms of the European Union. For policy recommendation, it would be important to better interlink the policy instruments with the agricultural workers' salaries. In future research, it could be important to assess how the sustainability indicators are interrelated with the wages paid in the European Union agricultural frameworks.

Acknowledgements This work is funded by National Funds through the FCT–Foundation for Science and Technology, I.P., within the scope of the project Refª UIDB/00681/2020 (https://doi.org/https://doi.org/10.54499/UIDP/00681/2020). This research is also funded by the Enovo company. This study was carried out under the international project "Agriculture 4.0: Current reality, potentialities and policy proposals" (CERNAS-IPV/2022/008). Furthermore we would like to thank the CERNAS Research Centre and the Polytechnic Institute of Viseu for their support. This work is too co-financed by the PRR–Plano de Recuperação e Resiliência (República Portuguesa) and the European NextGeneration EU Funds (https://recuperarportugal.gov.pt) through application PRR-C05-i03-I-000030—"Carb2Soil—Reforçar a Complementaridade entre agricultura e pecuária para aumentar a fertilidade dos solos e a sua capacidade de sequestro de carbono."

References

1. J. Arshad, T.A. Siddiqui, M.I. Sheikh, M.S. Waseem, M.A.B. Nawaz, E.T. Eldin, A.U. Rehman, Deployment of an intelligent and secure cattle health monitoring system. Egypt. Inform. J. **24**, 265 (2023)
2. P. Guadagna, M. Fernandes, F. Chen, A. Santamaria, T. Teng, T. Frioni, D.G. Caldwell, S. Poni, C. Semini, M. Gatti, Using deep learning for pruning region detection and plant organ segmentation in dormant spur-pruned grapevines. Precis. Agric. **24**, 1547 (2023)
3. A. Kalyanaraman, M. Burnett, A. Fern, L. Khot, J. Viers, Special report: the AgAID AI institute for transforming workforce and decision support in agriculture. Comput. Electron. Agric. **197**, 106944 (2022)
4. A. Ravankar, A. Rawankar, A.A. Ravankar, Real-time monitoring of elderly people through computer vision. Artif. Life Robot. **28**, 496 (2023)
5. A.F. Amalia, H.S.P. Rahayu, Y.P. Rahardjo, L. Hutahaean, E.S. Rohaeni, C. Indrawanto, R.A. Saptati, V. Siagian, A. Waris, Artificial intelligence for small hydroponics farms employing fuzzy logic systems and economic analysis. Rev. Bras. Eng. Agric. Ambient. **27**, 690 (2023)
6. F. Cheng, J. Wei, S. Jiang, Q. Chen, Y. Ru, H. Zhou, Feature enhancement guided network for yield estimation of high-density jujube. Plant Methods **19**, 85 (2023)
7. M.M. Anuar, A.A. Halin, T. Perumal, B. Kalantar, Aerial imagery paddy seedlings inspection using deep learning. Remote Sens. **14**, 274 (2022)
8. M.-C. Chiu, W.-M. Yan, S.A. Bhat, N.-F. Huang, Development of smart aquaculture farm management system using IoT and AI-based surrogate models. J. Agric. Food Res. **9**, 100357 (2022)

9. M. Hamano, S. Shiozawa, S. Yamamoto, N. Suzuki, Y. Kitaki, O. Watanabe, Development of a method for detecting the planting and ridge areas in paddy fields using AI, GIS, and precise DEM. Precis. Agric. **24**, 1862 (2023)
10. E.H. Hegazi, A.A. Samak, L. Yang, R. Huang, J. Huang, Prediction of soil moisture content from sentinel-2 images using convolutional neural network (CNN). Agron. Basel **13**, 656 (2023)
11. O. Jarkas, J. Hall, S. Smith, R. Mahmud, P. Khojasteh, J. Scarsbrook, R.K.L. Ko, ResNet and Yolov5-enabled non-invasive meat identification for high-accuracy box label verification. Eng. Appl. Artif. Intell. **125**, 106679 (2023)
12. B. Kaur et al., Insights into the harvesting tools and equipment's for horticultural crops: from then to now. J. Agric. Food Res. **14**, 100814 (2023)
13. J. Kim, I.D. Ha, S. Kwon, I. Jang, M.H. Na, A smart farm DNN survival model considering tomato farm effect. Agricult. Basel **13**, 1782 (2023)
14. T. Li, W. Fang, G. Zhao, F. Gao, Z. Wu, R. Li, L. Fu, J. Dhupia, An improved binocular localization method for apple based on fruit detection using deep learning. Inform. Process. Agricult. **10**, 276 (2023)
15. Y. Lin, C. Li, Does rural E-commerce agglomeration help expand family farms' size? Evidence from Taobao villages in China's Yangtze River Delta. Electron. Commer. Res. **23**, 1731 (2023)
16. D. Mishra, K. Muduli, R. Raut, B.E. Narkhede, H. Shee, S.K. Jana, Challenges facing artificial intelligence adoption during COVID-19 pandemic: an investigation into the agriculture and agri-food supply chain in India. Sustainability **15**, 6377 (2023)
17. L. Rodrigues, S.A. Magalhaes, D.Q. da Silva, F.N. dos Santos, M. Cunha, Computer vision and deep learning as tools for leveraging dynamic phenological classification in vegetable crops. Agron. Basel **13**, 463 (2023)
18. M. Rudenko, Y. Plugatar, V. Korzin, A. Kazak, N. Gallini, N. Gorbunova, The use of computer vision to improve the affinity of rootstock-graft combinations and identify diseases of grape seedlings. Inventions **8**, 92 (2023)
19. N.A. Ubina, S.-C. Cheng, A review of unmanned system technologies with its application to aquaculture farm monitoring and management. Drones-Basel **6**, 12 (2022)
20. D.D. Uyeh, R. Mallipeddi, T. Park, S. Woo, Y. Ha, Technological advancements and economics in plant production systems: how to retrofit? Front. Plant Sci. **13**, 929672 (2022)
21. H. Waheed, W. Akram, S. Ul-Islam, A. Hadi, J. Boudjadar, N. Zafar, A mobile-based system for detecting ginger leaf disorders using deep learning. Fut. Internet **15**, 86 (2023)
22. Z. Wu, H. Du, Artificial intelligence in agricultural picking robot displacement trajectory tracking control algorithm. Wirel. Commun. Mob. Comput. **2022**, 3105909 (2022)
23. S.S. Yamac, H. Negis, C. Seker, A.M.M. Memon, B. Kurtulus, M. Todorovic, G. Alomair, Saturated hydraulic conductivity estimation using artificial intelligence techniques: a case study for calcareous alluvial soils in a semi-arid region. Water **14**, 3875 (2022)
24. L.-B. Chen, X.-R. Huang, W.-H. Chen, Design and implementation of an artificial intelligence of things-based autonomous mobile robot system for pitaya harvesting. IEEE Sens. J. **23**, 13220 (2023)
25. M.-C. Chen, Y.-T. Cheng, C.-Y. Liu, Implementation of a fruit quality classification application using an artificial intelligence algorithm. Sens. Mater. **34**, 151 (2022)
26. S. Figiel, Development of artificial intelligence and potential impact of its applications in agriculture on labor use and productivity. Zag. Ekon. Rol. **373**, 5 (2022)
27. K. Legun, K.A. Burch, L. Klerkx, Can a robot be an expert? The social meaning of skill and its expression through the prospect of autonomous AgTech. Agric. Hum. Values **40**, 501 (2023)
28. N. Akseer et al., Economic costs of childhood stunting to the private sector in low- and middle-income countries. EClinicalMedicine **45**, 101320 (2022)
29. J. Walusiak-Skorupa, P. Kaczmarek, M. Wiszniewska, Artificial intelligence and employee's health: new challenges. Med. Pr. **74**, 227 (2023)

30. FADN, Several Statistics. https://agriculture.ec.europa.eu/data-and-analysis/farm-structures-and-economics/fadn_en
31. IBM SPSS Modeler, SPSS Modeler: Overview. https://www.ibm.com/products/spss-modeler
32. V.J.P.D. Martinho, Labour drivers in the agricultural sector of the European Union: the social role of farms, in *Impacts of Climate Change and Economic and Health Crises on the Agriculture and Food Sectors* (IGI Global, 2022), pp. 134–154

Chapter 9
Predictors of Interest Paid in the European Union's Agricultural Sector

Abstract In general, the interest paid does not assume a relevant dimension in the overall costs present in the European Union farms. In fact, considering the agricultural sector characteristics, the Common Agricultural Policy measures and the dynamics of the banking sector in the European Union, the interest paid is a small part of the costs supported by the farmers. In any case, banking loans are fundamental for farming investments and in this way, it is important to understand their respective context. Considering these motivations, this research proposes to consider artificial intelligence approaches and data from the Farm Accountancy Data Network to identify the models with higher accuracy and the most important indicators to predict the interest paid by the farms of the European Union. The contributions of this research bring relevant insights into the dynamics of the bank loans for the European Union agricultural sector and the respective measures inside the Common Agricultural Policy framework.

Keywords Machine learning models · European Union statistics · Common agricultural policy

9.1 Introduction

The conditions of bank credit impact the dynamics and performance of any economic sector, particularly in the rural regions [1], and specifically agriculture [2]. These scenarios occur because of the requirements associated with the investment frameworks and the working capital management.

In addition, the complements of capital needed to use the financial support provided by the national governments and international institutions justify the importance of credit for the economic agents. Nonetheless, the levels of interest paid by the European Union farmers, for example, seem to represent a small part of the total farming costs [3].

For these contexts and others related to the agricultural sector, the artificial intelligence methodologies represent important solutions [4], namely to predict situations

V. J. P. D. Martinho, *Machine Learning Approaches for Evaluating Statistical Information in the Agricultural Sector*, SpringerBriefs in Applied Sciences and Technology, https://doi.org/10.1007/978-3-031-54608-2_9

of financial stress [5], modelling credit risks [6], assess new credit demand [7] and analyse the causes of farming credit demand [8]. However, agriculture 4.0 brings also new credit risks [9]. Agricultural cooperatives have a crucial role in supporting farmers in their diverse challenges and tasks [10], including the need for credit. The same happens with the credit cooperatives for the financial inclusion of the less favoured population [11].

A relevant part of the studies carried out considering artificial intelligence in agriculture is related to sustainability, namely the environmental component and the respective needs of carbon sequestration [12] to mitigate the implications of the global warming associated with climate change. The new conditions created by these changes in the climate create additional risks in agriculture [13] and claim for new solutions of credit [14].

Other issues addressed by the researchers, based on artificial intelligence approaches, are, for example, the following: agricultural risks and the respective insurance contexts [15]; agroecosystem research [16] and models [17] and disease identification [18]. The credit programmes created, in certain conditions, to support the farmers [19] need, in some circumstances, to be complemented with other policies [20]. The complementarity between programmes and policies is fundamental for effective rural development [21].

Following the scenario presented previously, this research proposes to identify the main predictors and the respective models that support the explanation of the interest paid in the European Union agriculture, taking into account data at the farm level [3] and following the procedures proposed by the software IBM SPSS Modeller [22].

9.2 Data Investigation

The interest paid (on average over the period 2020–2021) was higher in the representative farms of the following European Union member states (Fig. 9.1): Denmark; Netherlands; Slovakia; Sweden; Czechia; Luxembourg; Germany; Belgium; Finland; Estonia; France; Latvia and Austria. These costs are relatively and significantly higher in Denmark and the Netherlands.

Table 9.1 with the microeconomic data disaggregated at the regional level confirms that Denmark, some German agricultural regions, Netherlands and Slovakia, for example, are the European Union frameworks where the representative farms have the highest averages (2020–2021) for the interest paid indicator. On the other hand, Greece, for instance, is where the representative farms present the lowest averages for the interest and financial costs paid on loans obtained for investments, and financial costs on debts.

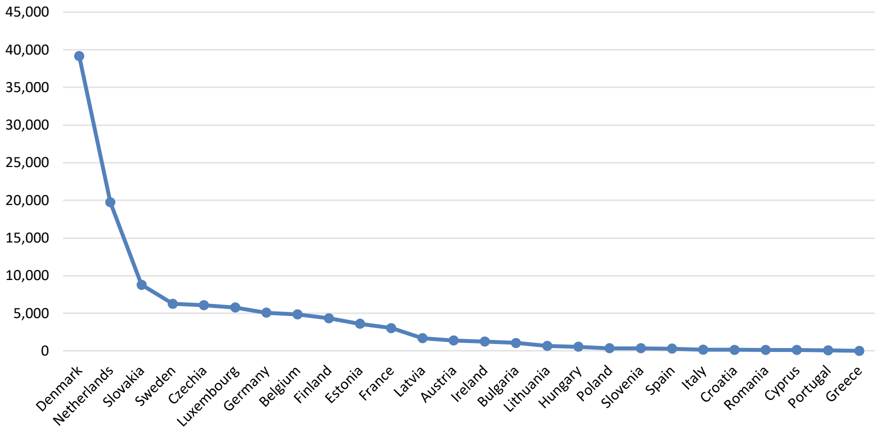

Fig. 9.1 Average values for the interest paid in the European Union countries, with data at the farm level, over the period 2020–2021

Table 9.1 Average values for the interest paid in the European Union farming regions, with data at the farm level, over the period 2020–2021

Member state	Region	Average
Austria	Austria	1382
Belgium	Vlaanderen	5327
Belgium	Wallonie	4072
Bulgaria	Severen tsentralen	1155
Bulgaria	Severoiztochen	1381
Bulgaria	Severozapaden	1983
Bulgaria	Yugoiztochen	1458
Bulgaria	Yugozapaden	641
Bulgaria	Yuzhen tsentralen	387
Croatia	Jadranska Hrvatska	60
Croatia	Kontinentalna Hrvatska	172
Cyprus	Cyprus	120
Czechia	Czechia	6072
Denmark	Denmark	**39,148**
Estonia	Estonia	3601
Finland	Etelä-Suomi	3306
Finland	Pohjanmaa	5178
Finland	Pohjois-Suomi	6611
Finland	Sisä-Suomi	4911
France	Alsace	1966

(continued)

Table 9.1 (continued)

Member state	Region	Average
France	Aquitaine	2628
France	Auvergne	2253
France	Basse-Normandie	5285
France	Bourgogne	3453
France	Bretagne	5423
France	Centre	3264
France	Champagne-Ardenne	2685
France	Corse	928
France	Franche-Comté	3331
France	Guadeloupe	421
France	Haute-Normandie	4485
France	Île-de-France	2899
France	La Réunion	1254
France	Languedoc-Roussillon	1378
France	Limousin	1955
France	Lorraine	3356
France	Midi-Pyrénées	1886
France	Nord-Pas-de-Calais	3875
France	Pays de la Loire	4868
France	Picardie	3162
France	Poitou–Charentes	3587
France	Provence-Alpes-Côte d'Azur	1107
France	Rhône-Alpes	2040
Germany	Baden-Württemberg	2194
Germany	Bayern	2422
Germany	Brandenburg	**24,893**
Germany	Hessen	3358
Germany	Mecklenburg-Vorpommern	**28,440**
Germany	Niedersachsen	6223
Germany	Nordrhein-Westfalen	3737
Germany	Rheinland-Pfalz	2377
Germany	Saarland	2398
Germany	Sachsen	**12,544**
Germany	Sachsen-Anhalt	**16,615**
Germany	Schleswig–Holstein/Hamburg	**7826**
Germany	Thüringen	**18,109**

(continued)

Table 9.1 (continued)

Member state	Region	Average
Greece	Ipiros-Peloponissos-Nissi Ioniou	*0*
Greece	Makedonia-Thraki	*3*
Greece	Sterea Ellas-Nissi Egaeou-Kriti	*2*
Greece	Thessalia	*0*
Hungary	Alföld	420
Hungary	Dunántúl	901
Hungary	Észak-Magyarország	251
Ireland	Ireland	1235
Italy	Abruzzo	*8*
Italy	Alto Adige	1127
Italy	Basilicata	18
Italy	Calabria	*0*
Italy	Campania	*0*
Italy	Emilia-Romagna	83
Italy	Friuli-Venezia Giulia	626
Italy	Lazio	124
Italy	Liguria	*7*
Italy	Lombardia	*3*
Italy	Marche	44
Italy	Molise	31
Italy	Piemonte	198
Italy	Puglia	29
Italy	Sardegna	91
Italy	Sicilia	64
Italy	Toscana	351
Italy	Trentino	135
Italy	Umbria	66
Italy	Valle d'Aosta	389
Italy	Veneto	601
Latvia	Latvia	1688
Lithuania	Lithuania	660
Luxembourg	Luxembourg	5767
Netherlands	The Netherlands	**19,733**
Poland	Malopolska i Pogórze	129
Poland	Mazowsze i Podlasie	250
Poland	Pomorze i Mazury	724
Poland	Wielkopolska and Slask	554

(continued)

Table 9.1 (continued)

Member state	Region	Average
Portugal	Açores e Madeira	128
Portugal	Alentejo e Algarve	87
Portugal	Norte e Centro	42
Portugal	Ribatejo e Oeste	255
Romania	Bucuresti-Ilfov	*0*
Romania	Centru	35
Romania	Nord-Est	65
Romania	Nord-Vest	42
Romania	Sud-Est	290
Romania	Sud-Muntenia	201
Romania	Sud-Vest-Oltenia	224
Romania	Vest	114
Slovakia	Slovakia	**8776**
Slovenia	Slovenia	350
Spain	Andalucía	78
Spain	Aragón	1203
Spain	Asturias	209
Spain	Canarias	235
Spain	Cantabria	79
Spain	Castilla y León	442
Spain	Castilla-La Mancha	274
Spain	Cataluña	840
Spain	Comunidad Valenciana	127
Spain	Extremadura	47
Spain	Galicia	109
Spain	Islas Baleares	172
Spain	La Rioja	477
Spain	Madrid	27
Spain	Murcia	274
Spain	Navarra	978
Spain	País Vasco	427
Sweden	Län i norra Sverige	3361
Sweden	Skogsoch mellanbygdslän	5291
Sweden	Slättbyggdslän	**7001**

Note Bold corresponds to the highest values and italic to the lowest

9.3 Core Findings

Tables 9.2 and 9.3 present the most accurate models, to predict the interest paid averages in the European Union farming regions, considering, respectively, 50/50 and 70/30 for the training set/testing set. The models displayed in the two figures are similar, however, the relative errors are different. The differences in the predictive capacity of the frameworks highlighted by these two figures are also exhibited in Figs. 9.2 and 9.3. In any case, the results presented in Fig. 9.3 are a consequence of the linear support vector machine (LSVM) model, for example, such as highlighting the relative error.

Table 9.2 Models with the highest accuracy (the lowest relative error, considering, respectively, 50/50 for the training set/testing set) for the interest paid in the European Union farming regions, with data at the farm level on average over the period 2020–2021

Model	Build time	Correlation	No. fields	Relative error
CHAID	1	1.000	14	0.000
Linear	1	1.000	25	0.001
XGBoost linear	1	0.997	177	0.006
XGBoost tree	1	0.995	177	0.028
C&R tree	1	0.929	30	0.138
Random forest	1	0.955	177	0.157
Random trees	1	0.919	177	0.161
Regression	1	0.697	6	0.549
SVM	1	0.943	168	1.147
Linear-AS	1	0.806	177	812.752

Table 9.3 Models with the highest accuracy (the lowest relative error, considering, respectively, 70/30 for the training set/testing set) for the interest paid in the European Union farming regions, with data at the farm level on average over the period 2020–2021

Model	Build time	Correlation	No. fields	Relative error
CHAID	3	1.000	22	0.000
Linear	3	0.998	30	0.003
XGBoost linear	3	0.997	177	0.007
XGBoost tree	3	0.995	177	0.020
Neural net	3	0.978	168	0.048
C&R tree	3	0.979	22	0.049
Random trees	3	0.963	177	0.080
SVM	3	0.901	168	1.123
Linear-AS	3	0.472	177	2.138
LSVM	3	− 0.131	177	743,000,000,000.000

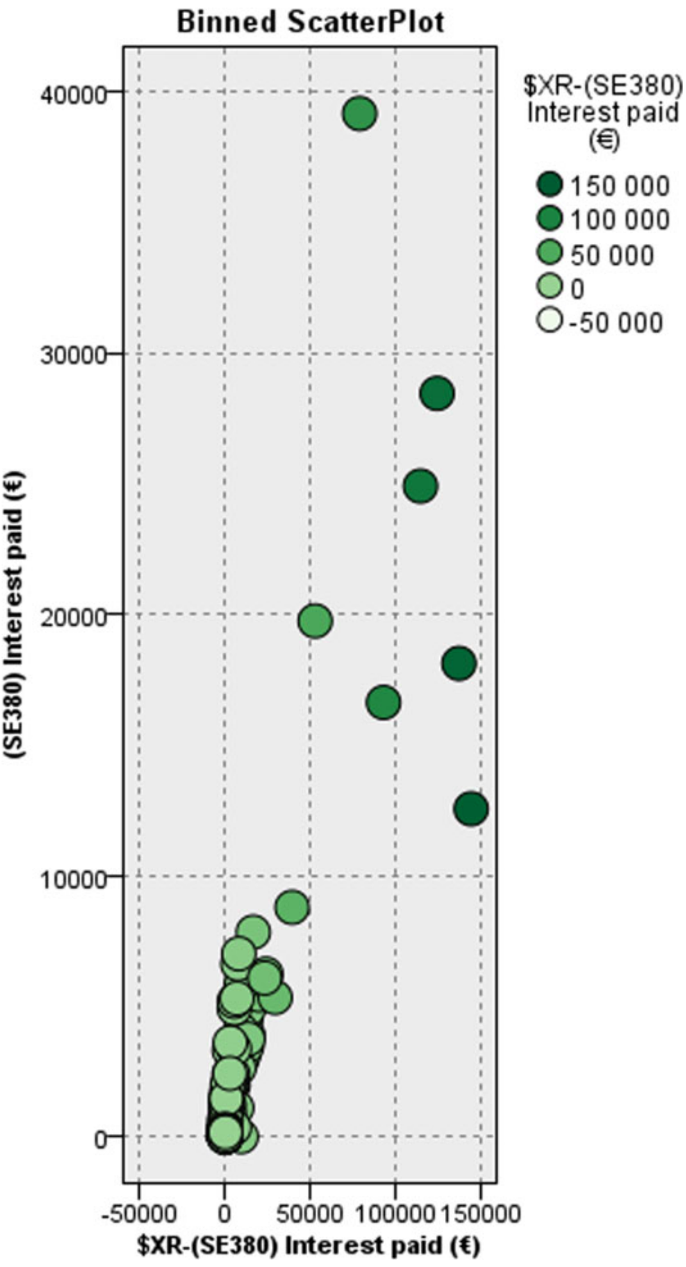

Fig. 9.2 Relationships between the observed values and the predicted ones for the interest paid in the European Union farming regions (considering, respectively, 50/50 for the training set/testing set), with data at the farm level on average over the period 2020–2021

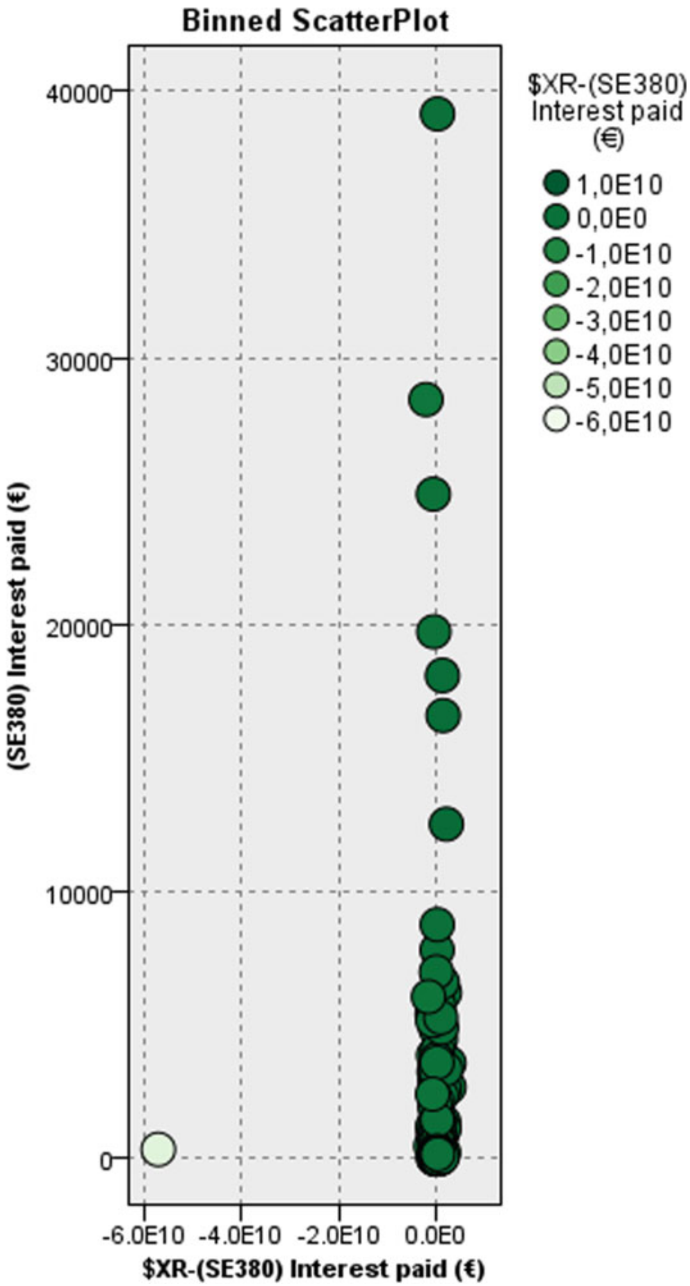

Fig. 9.3 Relationships between the observed values and the predicted ones for the interest paid in the European Union farming regions (considering, respectively, 70/30 for the training set/testing set), with data at the farm level on average over the period 2020–2021

Table 9.4 Importance of the predictors for the interest paid in the European Union farming regions (considering, respectively, 70/30 for the training set/testing set), with data at the farm level on average over the period 2020–2021

Nodes	Importance
Total support for rural development (€)	0.0286
Total assets (€)	0.0298
Net investment on fixed assets (€)	0.0307
Total current assets (€)	0.0364
Cash flow 2 (€)	0.0478
Intangible assets (€/farm)	0.0605
Vegetables and flowers (€/farm)	0.0671
Land, permanent crops and quotas (€)	0.0883
Total inputs (€)	0.1050
Gross investment on fixed assets (€)	0.1074

Nonetheless, the results presented in Table 9.4 were obtained considering that the partition node is set to have 70% train and 30% test, because with a partition node of 50/50 for these sets, the models did not identify the most important predictors.

Gross Investment on fixed assets (€), land, permanent crops & quotas (€), intangible assets (€/farm), net Investment on fixed assets (€) and total assets (€) are among the most important predictors of the interest paid in the representative farms of the European Union agricultural regions (Table 9.4).

To better understand the specific reality of the interest paid in the European Union countries, Table 9.5 presents the relative importance (for the year 2021) of these costs on the total external factors and the relative relevance of the total external factors costs on the total input. The interest paid represents, on average, only about 10% of the total external factors and these costs represent about 17% of the total inputs. In addition, the variability of the interest paid, across the European Union countries, is greater than, for example, that of the total external factors, such as revealing the results for the coefficient of variation.

9.4 Discussion and Conclusions

Interest paid is not among the costs that, in general, motivate more concerns for the farmers in the European Union contexts, considering their reduced relative importance in the total farming costs, including in the costs with external factors, and the diversity of importance in the diverse European frameworks. This statement does not ignore, however, the relevance of these costs in some agricultural contexts and the importance of the bank credits to promote farming investments and the performance of agriculture worldwide, particularly in the European Union sector. Considering these motivations, this research intended to bring more insights into the interest paid

Table 9.5 Relative importance (%) of the interest paid on the total costs

Member state	Total external factors (€)/total inputs (€)	Interest paid (€)/total external factors (€)
Belgium	11.176	13.902
Bulgaria	33.945	3.325
Czechia	25.322	4.683
Denmark	23.269	24.995
Germany	18.443	8.072
Estonia	18.724	10.427
Ireland	9.175	17.263
Greece	15.082	0.000
Spain	23.102	1.531
France	15.549	8.605
Croatia	10.822	4.712
Italy	18.832	1.375
Cyprus	12.098	2.040
Latvia	16.154	11.033
Lithuania	14.512	8.118
Luxembourg	11.504	16.586
Hungary	19.091	3.311
Netherlands	17.579	18.617
Austria	9.000	15.645
Poland	8.298	11.236
Portugal	19.447	1.364
Romania	18.744	2.574
Slovenia	3.625	28.654
Slovakia	24.954	4.291
Finland	12.443	16.959
Sweden	17.715	12.384
Average	16.485	9.681
Standard deviation	6.486	7.671
Coefficient of variation	0.393	0.792

by the European Union farmers' prediction and in this way create conditions that may support the different stakeholders to find better solutions for the interest paid management, working capital management and banking credit access. For that artificial intelligence approaches were considered following the procedures suggested by the software IBM SPSS Modeller and taking into account microeconomic statistical information from the Farm Accountancy Data Network, on average over the period 2020–2021.

The bank credit conditions affect the dynamics of the economic activities in the rural areas, because of the impacts on the investment initiatives (some of them associated with the financial supports created in the framework of the Common Agricultural Policy) and working capital management. Smart solutions may bring relevant contributions to these contexts, particularly in predicting frameworks of financial needs and credit risks. Nonetheless, it is expected that these new approaches bring other credit risks and challenges.

Denmark, Netherlands, Slovakia, Sweden, Czechia, Luxembourg, Germany, and Belgium are examples where the farms present the greatest values for the interest paid in the European Union context. These costs are particularly higher in the farms of Denmark and the Netherlands. On the other hand, Greece is an example where the levels of interest paid are lower.

With the machine learning approaches applied, some difficulties were found in identifying accurate models and important predictors of the interest paid in the European Union farming sector. In any case, with these limitations, the most relevant variables to predict the interest paid in the European Union farms were the following: Gross Investment on fixed assets; land, permanent crops and quotas; intangible assets; net Investment on fixed assets and total assets.

In terms of practical implications, highlighting the importance of variables related to the investment and the assets to predict the level of interest paid. In terms of policy recommendation, it is suggested to improve the interlink between the policy instruments from the Common Agricultural Policy framework with the interest paid, because none of the policy measures with statistical information in the database used has the importance to predict the costs associated with the banking credit. For future studies, it is suggested to identify the real relationships between the important predictors and the target (interest paid).

Acknowledgements This work is funded by National Funds through the FCT–Foundation for Science and Technology, I.P., within the scope of the project Refª UIDB/00681/2020 (https://doi.org/https://doi.org/10.54499/UIDP/00681/2020). This research is also funded by the Enovo company. This study was carried out under the international project "Agriculture 4.0: Current reality, potentialities and policy proposals" (CERNAS-IPV/2022/008). Furthermore we would like to thank the CERNAS Research Centre and the Polytechnic Institute of Viseu for their support. This work is too co-financed by the PRR–Plano de Recuperação e Resiliência (República Portuguesa) and the European NextGeneration EU Funds (https://recuperarportugal.gov.pt) through application PRR-C05-i03-I-000030—"Carb2Soil—Reforçar a Complementaridade entre agricultura e pecuária para aumentar a fertilidade dos solos e a sua capacidade de sequestro de carbono."

References

1. C. Deng, Research on the development and integration of inclusive finance in agricultural economy based on the artificial intelligence machine learning algorithm. Wirel. Commun. Mob. Comput. **2022**, 1758983 (2022)
2. A.V. Araujo, C. Mota, S. Siraj, Using genetic programming to identify characteristics of Brazilian regions in relation to rural credit allocation. Agricult. Basel **13**, 935 (2023)

3. FADN, Several Statistics. https://agriculture.ec.europa.eu/data-and-analysis/farm-structures-and-economics/fadn_en
4. D. Loukatos, M. Kondoyanni, I.-V. Kyrtopoulos, K.G. Arvanitis, Enhanced robots as tools for assisting agricultural engineering students' development. Electronics **11**, 755 (2022)
5. J. Chen, A.L. Katchova, C. Zhou, Agricultural loan delinquency prediction using machine learning methods. Int. Food Agribus. Manag. Rev. **24**, 797 (2021)
6. M.S. Uddin, G. Chi, M.A.M. Al-Janabi, T. Habib, K. Yuan, Modeling credit risk with a multi-stage hybrid model: an alternative statistical approach. J. Forecast. **41**, 1386 (2022)
7. J. Iffy, R. Kuhns, K. Patrick, Can machine learning improve prediction: an application with farm survey data. Int. Food Agribus. Manag. Rev. **21**, 1083 (2018)
8. H.D. Dang, A.H.T. Dam, T.T. Pham, T.M.T. Nguyen, Determinants of credit demand of farmers in lam dong, Vietnam a comparison of machine learning and multinomial logit. Agric. Finan. Rev. **80**, 255 (2020)
9. A. Belhadi, S.S. Kamble, V. Mani, I. Benkhati, F.E. Touriki, An ensemble machine learning approach for forecasting credit risk of agricultural SMEs' investments in agriculture 4.0 through supply chain finance. Ann. Oper. Res. **12**, 1–29 (2021)
10. T. Cukur, F. Cukur, Determining factors affecting cooperative membership of the beekeepers using decision tree algorithms. J. Agric. Sci. Tarim Bilim. Derg. **28**, 25 (2022)
11. M.J. Segovia-Vargas, I.M. Miranda-Garcia, F.A. Oquendo-Torres, Sustainable finance: the role of savings and credit cooperatives in Ecuador. Ann. Public Coop. Econ. **94**, 951 (2023)
12. U. Acharya, R. Lal, R. Chandra, Data driven approach on in-situ soil carbon measurement. Carbon Manag. **13**, 401 (2022)
13. P. Jha, S. Chinngaihlian, P. Upreti, A. Handa, A machine learning approach to assess implications of climate risk factors on agriculture: the Indian case. Clim. Risk Manag. **41**, 100523 (2023)
14. X. He, J. Yan, L.E. Yang, Y. Wu, H. Zhou, Climate change adaptation of smallholders on the Tibetan Plateau under government interventions. J. Clean. Prod. **381**, 135171 (2022)
15. A. Ghahari, N.K. Newlands, V. Lyubchich, Y.R. Gel, Deep learning at the interface of agricultural insurance risk and spatio-temporal uncertainty in weather extremes. N. Am. Actuar. J. **23**, 535 (2019)
16. M.S. Moran, P. Heilman, D.P.C. Peters, C.H. Collins, Agroecosystem research with big data and a modified scientific method using machine learning concepts. Ecosphere **7**, e01493 (2016)
17. M. Nowatzke, L. Damiano, F.E. Miguez, G.S. McNunn, J. Niemi, L.A. Schulte, E.A. Heaton, A. VanLoocke, Augmenting agroecosystem models with remote sensing data and machine learning increases overall estimates of nitrate-nitrogen leaching. Environ. Res. Lett. **17**, 114010 (2022)
18. G. Sambasivam, G.D. Opiyo, A predictive machine learning application in agriculture: cassava disease detection and classification with imbalanced dataset using convolutional neural networks. Egypt. Inform. J. **22**, 27 (2021)
19. K. Jindo, J.A. Andersson, F. Quist-Wessel, J. Onyango, J.W.A. Langeveld, Gendered investment differences among smallholder farmers: evidence from a microcredit programme in Western Kenya. Food Sec. **12**, 1–16 (2023)
20. M. Hossain, M.A. Malek, M.A. Hossain, M.H. Reza, M.S. Ahmed, Agricultural microcredit for tenant farmers: evidence from a field experiment in Bangladesh. Am. J. Agr. Econ. **101**, 692 (2019)
21. A. Wagener, J. Zenker, Decoupled but not neutral: the effects of counter-cyclical cash transfers on investment and incomes in rural Thailand† JEL codes. Am. J. Agr. Econ. **103**, 1637 (2021)
22. IBM SPSS Modeler, SPSS Modeler: Overview. https://www.ibm.com/products/spss-modeler

Chapter 10
Predictive Artificial Intelligence Approaches of Labour Use in the Farming Sector

Abstract It is not expected that the agricultural sector absorbs a great part of the employment in developed economies with a dynamic industry and services sector. When the percentage of employment in agriculture is high, this may be a sign of the weak performance of the farms. Every country wants to have a developed farming sector to not compromise the dynamics and performance of the economy. In any case, agricultural employment plays a fundamental role, particularly in rural spaces and in contexts of temporary crises in the remaining economy. Taking into account these motivations, this chapter aims to highlight the main approaches and variables that may be considered to predict labour use in the European Union farms. To achieve these aims, European Union agricultural statistics were considered, as well as models based on the new technologies associated with the digital transition worldwide. The results found may provide pertinent suggestions for a more sustainable farming sector, where the social contributions may be improved.

Keywords Accuracy · European Union farm accountancy data network · Agricultural employment

10.1 Introduction

Artificial intelligence, economic sectors, including agriculture, and employment are combinations with several dimensions [1], associated with the fourth industrial revolution [2], some positive and others undesirable [3]. In fact, the innovative approaches open new potentialities for the different domains of society [4, 5] and particularly for the agro-food sectors [6], through smart solutions [7].

Some of the opportunities shaped are related to the possibility of increasing the efficiency of the farms [8], improving productivity [9] and mitigating environmental impacts [10]. The smart approaches may bring relevant added value for the sustainability of the economic development [11] and they are the hope to deal with the increased demand for food without additional environmental consequences.

V. J. P. D. Martinho, *Machine Learning Approaches for Evaluating Statistical Information in the Agricultural Sector*, SpringerBriefs in Applied Sciences and Technology, https://doi.org/10.1007/978-3-031-54608-2_10

The farming sector has an important social contribution creating employment for many people [12] worldwide [13], however, the technological changes may reduce the human labour in the farming sector and create new difficulties for the health of the agricultural workforce, at physical and psychological level.

The new technologies require additional competencies and this may create new kinds of conflicts and stress. Another consequence may the social conflicts, due to the changes in the employment structure [14], and the emergence of questions associated with social ethics [15]. In any case, it is expected also that artificial intelligence create new jobs [16].

The question here is to understand if the net employment will increase, or decrease, with the digital transformation. In other words, it will be important in future research to assess whether the employment created by the smart solutions is greater, or less, than that destroyed in the economic activities.

There is a long discussion to be carried out about the relationships between human labour and digital solutions in agriculture [17], but the complementarity among the workforce and robots, for example, is a possibility for the future of crop and livestock activities in the farms [18].

From this literature survey, it seems pertinent to bring more insights about the most important variables and adjusted models to predict the farming labour in the European Union farms, using data from the European databases with data at the farm level [19], consider the procedures proposed by approaches that take into account artificial intelligence methodologies [20] and following the findings of Martinho [21].

10.2 Data Assessment

The representative farms with the highest labour use (on average over the period 2019–2020) belong to European Union countries, such as Slovakia, Czechia, Netherlands, Bulgaria, Germany, Belgium, Denmark, Estonia, Latvia, Romania, Hungary, Luxembourg, Spain, Poland and France (Fig. 10.1). These results are, in part, confirmed, with the results presented in Table 10.1 for the statistical information obtained from the Farm Accountancy Data Network for the representative farms of the European Union agricultural regions. These microeconomic data were obtained considering harmonised bookkeeping principles.

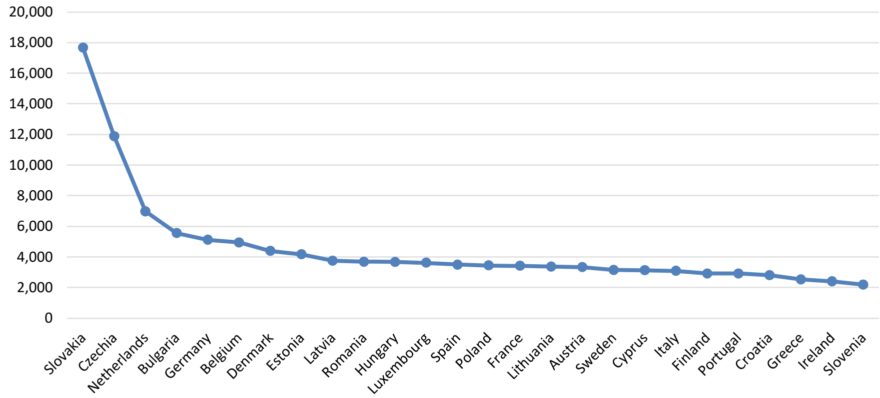

Fig. 10.1 Average values for labour use in the European Union countries, with data at the farm level, over the period 2019–2020

10.3 Results Revealed

Table 10.2 displays the models with the highest accuracy (for the training set) to predict labour use (on average over the period 2019–2020). These accurate models are the following: generalised linear (enlarges the general linear model); generalised linear engine (GLE); linear-AS (linear regression); linear (linear regression); Chi-squared Automatic Interaction Detection (CHAID); XGBoost linear (advanced application of a gradient boosting algorithm with a linear model as the reference model); XGBoost tree (advanced application of a gradient boosting algorithm with a tree model as the reference model); random trees (multiple decision trees); random forest (algorithm with a tree model as the reference) and support vector machine (SVM).

Figure 10.2, for the relationships among the observed values of the labour use (in the farms of the European Union agricultural regions) averages (over the period 2019–2020) and those predicted, confirms the predictive relevance of these models.

The most relevant predictors of labour use in the European Union agricultural regions are, respectively, the following (Table 10.3): total labour input (AWU); paid labour Input (h); fertiliser P_2O_5 (q); oil-seed crops (€/farm); economic size (€'000); unpaid labour input (h); other output (€/farm); single area payment (€); machinery and building current costs (€) and fertiliser N (q). Economic size, for example, may be considered by the stakeholders to predict the labour use in the European Union farming sector.

Table 10.1 Average values for the labour use in the European Union farming regions, with data at the farm level, over the period 2019–2020

Member state	Region	Average
Austria	Austria	3326
Belgium	Vlaanderen	5495
Belgium	Wallonie	4017
Bulgaria	Severen tsentralen	**6643**
Bulgaria	Severoiztochen	6638
Bulgaria	Severozapaden	6089
Bulgaria	Yugoiztochen	5892
Bulgaria	Yugozapaden	4367
Bulgaria	Yuzhen tsentralen	4695
Croatia	Jadranska Hrvatska	3390
Croatia	Kontinentalna Hrvatska	2612
Cyprus	Cyprus	3125
Czechia	Czechia	**11,885**
Denmark	Denmark	4386
Estonia	Estonia	4166
Finland	Etelä-Suomi	2570
Finland	Pohjanmaa	2864
Finland	Pohjois-Suomi	3932
Finland	Sisä-Suomi	3380
France	Alsace	3286
France	Aquitaine	3689
France	Auvergne	2615
France	Basse-Normandie	3181
France	Bourgogne	3835
France	Bretagne	3816
France	Centre	2987
France	Champagne-Ardenne	3026
France	Corse	3835
France	Franche-Comté	3204
France	Guadeloupe	*2514*
France	Haute-Normandie	3293
France	Île-de-France	3494
France	La Réunion	3056
France	Languedoc-Roussillon	3672
France	Limousin	2713
France	Lorraine	2912

(continued)

Table 10.1 (continued)

Member state	Region	Average
France	Midi-Pyrénées	2728
France	Nord-Pas-de-Calais	3537
France	Pays de la Loire	4103
France	Picardie	2927
France	Poitou-Charentes	3083
France	Provence-Alpes-Côte d'Azur	5133
France	Rhône-Alpes	3669
Germany	Baden-Württemberg	4446
Germany	Bayern	3827
Germany	Brandenburg	**15,065**
Germany	Hessen	3814
Germany	Mecklenburg-Vorpommern	**9931**
Germany	Niedersachsen	4761
Germany	Nordrhein-Westfalen	4361
Germany	Rheinland-Pfalz	5599
Germany	Saarland	3860
Germany	Sachsen	**13,650**
Germany	Sachsen-Anhalt	**11,488**
Germany	Schleswig-Holstein/Hamburg	4831
Germany	Thüringen	**16,008**
Greece	Ipiros-Peloponissos-Nissi Ioniou	*2285*
Greece	Makedonia-Thraki	*2488*
Greece	Sterea Ellas-Nissi Egaeou-Kriti	2755
Greece	Thessalia	2595
Hungary	Alföld	3424
Hungary	Dunántúl	4233
Hungary	Észak-Magyarország	3472
Ireland	Ireland	*2397*
Italy	Abruzzo	2869
Italy	Alto Adige	3465
Italy	Basilicata	3611
Italy	Calabria	2904
Italy	Campania	3034
Italy	Emilia-Romagna	3464
Italy	Friuli-Venezia Giulia	3463

(continued)

Table 10.1 (continued)

Member state	Region	Average
Italy	Lazio	3400
Italy	Liguria	2958
Italy	Lombardia	3642
Italy	Marche	2783
Italy	Molise	2998
Italy	Piemonte	3704
Italy	Puglia	2623
Italy	Sardegna	2787
Italy	Sicilia	*2433*
Italy	Toscana	3753
Italy	Trentino	2567
Italy	Umbria	2811
Italy	Valle d'Aosta	4660
Italy	Veneto	3354
Latvia	Latvia	3749
Lithuania	Lithuania	3359
Luxembourg	Luxembourg	3615
Netherlands	The Netherlands	**6973**
Poland	Malopolska i Pogórze	3224
Poland	Mazowsze i Podlasie	3389
Poland	Pomorze i Mazury	3713
Poland	Wielkopolska and Slask	3571
Portugal	Açores e Madeira	*2230*
Portugal	Alentejo e Algarve	3409
Portugal	Norte e Centro	2812
Portugal	Ribatejo e Oeste	2985
Romania	Bucuresti-Ilfov	4115
Romania	Centru	3999
Romania	Nord-Est	3339
Romania	Nord-Vest	3651
Romania	Sud-Est	3942
Romania	Sud-Muntenia	3230
Romania	Sud-Vest-Oltenia	4013
Romania	Vest	3630
Slovakia	Slovakia	**17,666**
Slovenia	Slovenia	*2184*
Spain	Andalucía	3352

(continued)

Table 10.1 (continued)

Member state	Region	Average
Spain	Aragón	3548
Spain	Asturias	3924
Spain	Canarias	**8220**
Spain	Cantabria	3068
Spain	Castilla y León	2902
Spain	Castilla-La Mancha	3930
Spain	Cataluña	3822
Spain	Comunidad Valenciana	*2362*
Spain	Extremadura	4627
Spain	Galicia	*2556*
Spain	Islas Baleares	3348
Spain	La Rioja	3568
Spain	Madrid	4822
Spain	Murcia	4793
Spain	Navarra	*2369*
Spain	País Vasco	3206
Sweden	Län i norra Sverige	3280
Sweden	Skogsoch mellanbygdslän	3098
Sweden	Slättbyggdslän	3141

Note Bold corresponds to the highest values and italic to the lowest

Table 10.2 Models with the highest accuracy (the lowest relative error) for the labour use in the European Union farming regions, with data at the farm level on average over the period 2019–2020

Model	Build time	Correlation	No. fields	Relative error
Generalised linear	1	1.000	177	0.000
Generalised linear engine	1	1.000	177	0.000
Linear-AS	1	1.000	177	0.000
Linear	1	1.000	26	0.000
CHAID	1	1.000	12	0.001
XGBoost linear	1	0.999	177	0.003
XGBoost tree	1	0.997	177	0.018
Random trees	1	0.970	177	0.066
Random forest	1	0.965	177	0.194
SVM	1	0.954	170	1.046

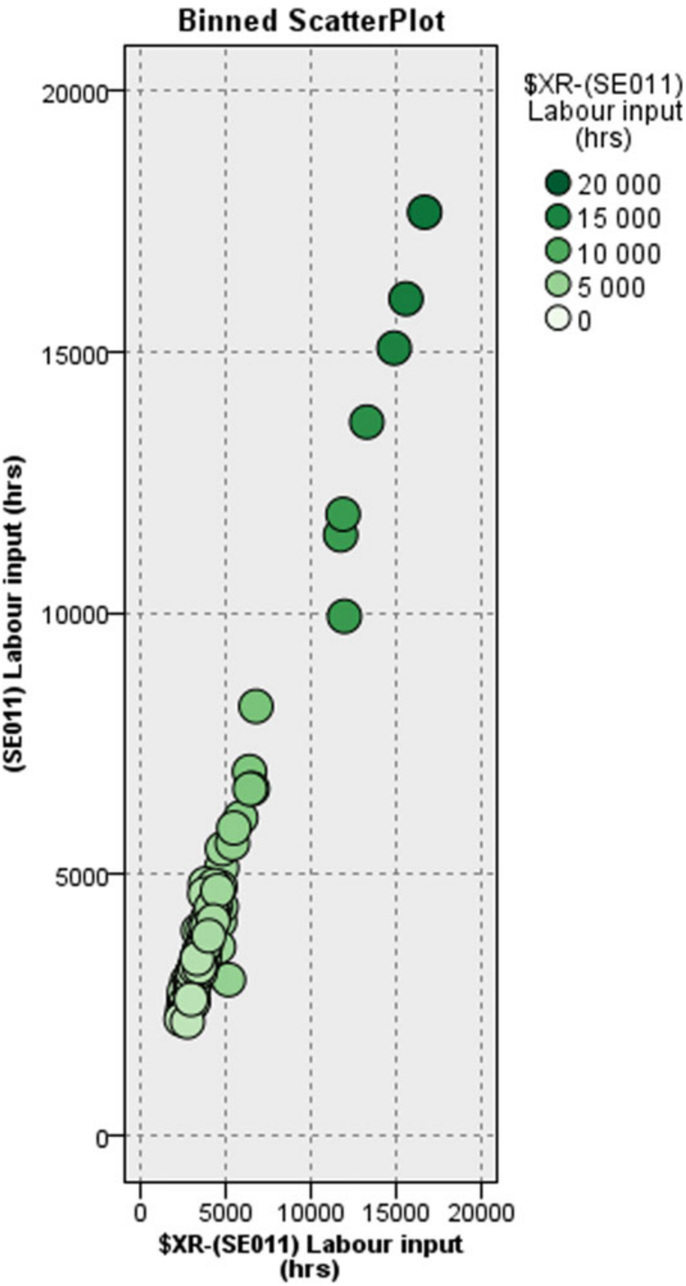

Fig. 10.2 Relationships between the observed values and the predicted ones for the labour use in the European Union farming regions, with data at the farm level on average over the period 2019–2020

Table 10.3 Importance of the predictors for the labour use in the European Union farming regions, with data at the farm level on average over the period 2019–2020

Nodes	Importance
Fertiliser N (q)	0.0064
Machinery and building current costs (€)	0.0065
Single area payment (€)	0.0067
Other output (€/farm)	0.0072
Unpaid labour input (h)	0.0073
Economic size (€'000)	0.0074
Oil-seed crops (€/farm)	0.0089
Fertiliser P_2O_5 (q)	0.0299
Paid labour input (h)	0.0361
Total labour input (AWU)	0.0590

10.4 Discussion and Conclusions

The labour use in the agricultural sector, in general, has undergone several changes over the last decades, due to the mechanisation of the sector, the use of chemical products for fertilisation and crop protection practises, and the transformation in societies and economies. In fact, the economic changes verified worldwide have promoted rural exodus, with the desertification of the rural areas, and the consequent agricultural abandonment, and urban congestion (with implications on the social urban organisations and environmental pressures, because of the pollution). It is expected that the new technologies open new opportunities for labour use management in the different economic sectors, including in agriculture, and create new challenges. From this perspective, this chapter intended to identify accurate models to predict the labour use in the farms of the European Union countries and agricultural regions. To achieve these objectives, artificial intelligence approaches were considered based on new solutions and statistical information from the Farm Accountancy Data Network was taken into account.

It is expected that the new smart solutions may be complementary to human labour in agriculture, supporting several stakeholders to find solutions to deal with the scarcity of the workforce. In any case, the discussion about these topics is not unanimous, because in some contexts digital technologies may solve some needs of human labour in the agricultural sector, but in other frameworks, these new approaches may substitute the human workforce and create new problems related to the health of the workers.

On average, over the period 2019–2020, Slovakia, Czechia, Netherlands, Bulgaria, Germany, Belgium, Denmark and Estonia are examples where the European Union farms present the highest values for human labour use. These contexts are, in general, verified for the microeconomic data at the country level and disaggregated by agricultural region.

Generalised linear, GLE, linear-AS, linear, CHAID, XGBoost linear, XGBoost tree, random trees, random forest, and SVM are the most accurate models to predict labour use in the European Union representative farms. Total labour input, paid labour input, fertiliser P_2O_5, oil-seed crops, economic size, unpaid labour input, other output, single area payment, machinery and building current costs, and fertiliser N.

In terms of practical implications, the findings here obtained highlight the importance of the economic size of the European Union farms to predict labour use, as well as indicators related to some specific productions and the single area payment. For policy recommendation, it is suggested to analyse the relationships between labour use and the single payment area and why not with other policy instruments. In future research, it could be important to assess the effect of the time in these findings, namely considering panel data and/or lagged variables one or more years.

Acknowledgments This work is funded by National Funds through the FCT–Foundation for Science and Technology, I.P., within the scope of the project Refª UIDB/00681/2020 (https://doi.org/https://doi.org/10.54499/UIDP/00681/2020). This research is also funded by the Enovo company. This study was carried out under the international project "Agriculture 4.0: Current reality, potentialities and policy proposals" (CERNAS-IPV/2022/008). Furthermore we would like to thank the CERNAS Research Centre and the Polytechnic Institute of Viseu for their support. This work is too co-financed by the PRR–Plano de Recuperação e Resiliência (República Portuguesa) and the European NextGeneration EU Funds (https://recuperarportugal.gov.pt) through application PRR-C05-i03-I-000030—"Carb2Soil—Reforçar a Complementaridade entre agricultura e pecuária para aumentar a fertilidade dos solos e a sua capacidade de sequestro de carbono."

References

1. T. Ciarli, M. Kenney, S. Massini, L. Piscitello, Digital technologies, innovation, and skills: emerging trajectories and challenges. Res. Policy **50**, 104289 (2021)
2. J.M. Tien, Toward the fourth industrial revolution on real-time customization. J. Syst. Sci. Syst. Eng. **29**, 127 (2020)
3. R. Sparrow, M. Howard, C. Degeling, Managing the risks of artificial intelligence in agriculture. NJAS Impact Agric. Life Sci. **93**, 172 (2021)
4. D.E. Lucero-Prisno et al., Prospects for financial technology for health in Africa. Digit. Health **8**, 20552076221119548 (2022)
5. X. Ma, H. Pan, Improving entrepreneurial skills and professional association of college students using big data analysis and IoT. Soft Comput. **27**, 14253 (2023)
6. V. Dadi, S.R. Nikla, R.S. Moe, T. Agarwal, S. Arora, Agri-food 4.0 and innovations: revamping the supply chain operations. Prod. Eng. Arch. **27**, 75 (2021)
7. S.L. Ullo, G.R. Sinha, Advances in IoT and smart sensors for remote sensing and agriculture applications. Remote Sens. **13**, 2585 (2021)
8. S. AlZu'bi, B. Hawashin, M. Mujahed, Y. Jararweh, B.B. Gupta, An efficient employment of internet of multimedia things in smart and future agriculture. Multimed. Tools. Appl. **78**, 29581 (2019)
9. M. Pathan, N. Patel, H. Yagnik, M. Shah, Artificial cognition for applications in smart agriculture: a comprehensive review. Artif. Intell. Agric. **4**, 81 (2020)
10. S. Ferraris, R. Meo, S. Pinardi, M. Salis, G. Sartor, Machine learning as a strategic tool for helping cocoa farmers in Côte D'Ivoire. Sensors **23**, 7632 (2023)
11. R.I. Mukhamediev et al., Review of some applications of unmanned aerial vehicles technology in the resource-rich country. Appl. Sci. Basel **11**, 10171 (2021)

12. S.K. Apat, J. Mishra, K.S. Raju, N. Padhy, An artificial intelligence-based crop recommendation system using machine learning. J. Sci. Ind. Res. **82**, 558 (2023)
13. H. Khan, I. Ul-Haq, M. Munsif, S.U. Khan, M.Y. Lee, Automated wheat diseases classification framework using advanced machine learning technique. Agricult. Basel **12**, 1226 (2022)
14. A. Akimov, Robotics and development. Mirovaya Ekon. Mezhdunarodyne Otnosheniya **61**, 74 (2017)
15. C. Yang, C. Lin, X. Fan, Cultivation model of entrepreneurship from the perspective of artificial intelligence ethics. Front. Psychol. **13**, 885376 (2022)
16. C. Guang, Development of migrant workers in construction based on machine learning and artificial intelligence technology. J. Intell. Fuzzy Syst. **40**, 6629 (2021)
17. S. Rotz et al., Automated pastures and the digital divide: how agricultural technologies are shaping labour and rural communities. J. Rural Stud. **68**, 112 (2019)
18. V. Marinoudi, C.G. Sorensen, S. Pearson, D. Bochtis, Robotics and labour in agriculture: a context consideration. Biosyst. Eng. **184**, 111 (2019)
19. FADN, Several Statistics. https://agriculture.ec.europa.eu/data-and-analysis/farm-structures-and-economics/fadn_en
20. IBM SPSS Modeler, SPSS Modeler: Overview. https://www.ibm.com/products/spss-modeler
21. V.J.P.D. Martinho, Labour drivers in the agricultural sector of the European Union: the social role of farms, in *Impacts of Climate Change and Economic and Health Crises on the Agriculture and Food Sectors* (IGI Global, 2022), pp. 134–154