
Machine Learning Predictive Modeling of Agricultural Sustainability Indicators

Raden Roro Shafira Meisy Sudarsono* , Harimukti Wandebori
School of Business and Management, Institut Teknologi Bandung

**Corresponding author: raden_oro@sbm-itb.ac.id*

Abstract. Modern-day researchers are provided with data abundance that has its drawback: increased analysis complexity. Approaching this issue through traditional data analysis techniques provides only partial solutions to the complex situation. This research offers analytical and predictive models based on machine-learning algorithms (linear regression, random forest, and generalized additive model) that can be used to assess and improve the Common Agricultural Policy (CAP) impact over agricultural sustainability in European Union (EU) countries, providing the identification of proper instruments that can be adopted by EU policymakers and CAP Council in financial management of the policy. The chosen methodology elaborates custom-developed models based on a dataset containing 22 relevant indicators, considering three main dimensions contributing to the EU sustainable agriculture development goals in the CAP context: social, environment, and economic. The results showed that sustainable agriculture parameters influenced by the relevant indicators could be modeled with both linear and non-linear regression approaches by utilization of real-time data using machine learning. The predictive analytic models provide satisfactory performance and could be adopted by researchers and practitioners as policy impact monitoring and controlling tools, not only the EU but also for other countries that have or plan to adopt similar agricultural policies.

Keywords : Agricultural policies, common agricultural policy, machine learning, rural development, sustainable agriculture

1. INTRODUCTION

The United Nations Sustainable Development Goals (UN SDGs) optimistically commit to ensure year-round safe, nutritious, and sufficient food access (SDG target 2.1) and to eradicate all kinds of malnutrition (SDG target 2.2) for all people by 2030. Yet, according to the Food and Agriculture Organization (UN FAO), the world has not progressed an inch closer to it [1]. The 2020 State of Food Security and Nutrition in the World (SOFI) reported that the COVID-19 pandemic continues exposing the world's food system weaknesses, which threaten food security status of many [2], as it will result in devastating impacts on health, productivity, and economy. Sustainable agriculture practices are needed to end world hunger and achieve food security, but recent studies showed that the current agriculture practices are far from sustainable [3].

Sustainability has three dimensions: social, environment, and economic [4]. Being the sector with strong influence on those three dimensions, agriculture is considered one of the engines of sustainable development. For over six decades, in partnership between agriculture and society, the Common Agricultural Policy (CAP) has shaped the European countries' agricultural sector [5].

The CAP is structured into two pillars aiding the delineation of policy's elements which focused on direct payments and market measure to support agriculture productivity; and promote rural areas' development, competitiveness, and natural resources sustainable management [46]. Since agriculture is an essential sector with significant influence on rural areas, the implementation of CAP and its positive impact on rural employment and economic growth [6] demonstrates that it is indeed shaping rural development despite the critics. Research interest in agricultural policy reforms has increased, especially in investigating the CAP and its effects on different output performances. However, the scientific debates around CAP were fragmented due to the different approaches and perspectives evolved on the topic [5]. As to help policy-makers evaluate indicators being used in the current policy, this research aims to enrich the body of knowledge of the common monitoring and evaluation framework (CMEF). The CMEF is updated annually [54] as it provides policy-makers with set of evidence and guideline for decision-making to improve CAP's utility and efficiency, transparency, effectiveness, and learning.

Machine learning (ML) has been implemented in several agricultural research to model and predict specific scenarios. It is considered a useful approach in providing a better understanding of the complexity that lies behind the input, process, and output of studied phenomenon; especially when working with real-time data which are non-linear in nature. Statistical modeling approaches based on ML algorithms can be used as an alternative to traditional methods and overcome their limitations [6]. While several previous studies have shown that using non-linear approaches has given better predictive performance than linear modeling [6]–[9], real-time changes causing alteration of data behavior need to be handled and interpreted with caution. Hence, to make sure that the available equations are reproducible, it needs further tuning. This research aims to propose predictive analytic models of the CAP's impact on agricultural production support (Pillar I) and rural development (Pillar II) using non-linear (random forest and generalized additive model) and linear (multiple linear regression) algorithms. It also gives insight into which variables are best to represent the predictive models better to assess the effects of CAP at hand, and to offer future researchers and policymakers valuable information based on digital technologies necessary to develop a similar agricultural supporting program for non-European countries with lower AOI level [10], which considered as far behind in achieving the agricultural-related goals of sustainable development.

2. MATERIALS AND METHODS

2.1. Data Exploration

To investigate the CAP's impact on agricultural sector and to identify proper indicators accounting for the instruments that can be used by policymakers, especially in the CAP financial management, this research developed ML models based on variables described in Table 1. The data from 2011-2019 financial period being used were collected from official databases such as the Eurostat, European Commission, and FAO. For the relevance and comparability of present

research’s primary objective of analyzing the CAP: 2014-20 and with respect to data availability of all variables observed, this time range is deemed appropriate.

The agricultural sector's economic situation is related to sustainability: it should be economically viable and not degrade the environment [11]. In [12], the authors focused on the agriculture’s sustainability dimensions and used a series of variables such as employment rate and income, workforce stability, agriculture inputs, insured area, agricultural risks, economic share to GDP, and economic dependence of agricultural activity. Other study [7] with different perspectives in elaborating agriculture economic development used agriculture share of GDP, GVA, farming subsidies, and agriculture production as the variables. Therefore, to assess the sector’s rural environment and economic importance, financial variables such as consumer price index (CPI), GDP, GDP in rural areas (GDP_Rural), GVA, agricultural GVA (GVA_Agri), and GVA registered for rural areas (GVA_Rural) were included; as well as productivity variables consisted of crop yield, fertilizer usage, labor productivity (LPA), total factor productivity (TFP), ammonia (NH₃) and carbon dioxide (CO₂) emissions from farming.

The European countries had high AOI levels, reflecting their government’s high orientation toward the sector through a share of spending relative to the agricultural GVA [10]. This helped increasing productivity and growth by reducing budget constraints and increasing capital [13]. One of the spending schemes is through research and development (R&D). To narrow the gap of possible future food insecurity caused by projected high-calorie demand in 2050 is the adaptation of advanced technology by agricultural producers [14]. Hence, agriculture R&D (ARD) is included as indicator. The CAP’s social goal is to tackle unemployment and poverty rate in rural areas since it was reported higher than in urban areas [15]. Recent evidence showed that CAP successfully increased GDP per capita, rural employment rate, and decreasing rural poverty rate [6]. In this sense, these variables were considered to channel the social impact of CAP: direct payment (DP), agricultural entrepreneurial income (AEI), agricultural factor income (AFI), degree of rural poverty (DRP), rural employment rate (RER), and agriculture employment (AgriEmployment).

Table 1. Dataset Descriptive Statistics

Variable	N	Minimum	Maximum	Mean	SD
GDP	99	431,120.3	3,966,094.0	2,199,334.2	2,498,576.3
GDP_Rural	99	35,980.0	36,300.0	36,140.0	226.3
GVA	99	6,043.9	27,870.4	16,9957.2	15,433.7
GVA_Agri	99	6,043.9	7,759.7	6,901.8	102.6
GVA_Rural	99	97.8	107.3	102.6	6.7
DP	99	714.6	4,708.5	2,711.5	2,824.1
CPI	99	1.4	3.6	2.5	1.6
AFI	99	19,831.0	24,523.0	22,177.0	3,317.7
AgriEmployment	99	424.2	514.4	469.3	63.8
RER	99	1,603.9	9,784.7	5,694.3	5,784.6
RER_M	99	878.6	5,275.1	3,076.8	3,108.8
RER_F	99	725.4	4,509.6	2,617.5	2,675.8
DRP	99	15.6	16.1	15.8	0.4
AEI	99	2,325.7	8,509.2	5,417.4	4,371.4
LPA	99	114.2	115.7	114.9	1.1

Variable	N	Minimum	Maximum	Mean	SD
ARD	99	37.4	917.5	477.4	622.3
TFP	99	102.2	104.8	103.5	1.8
Yield	99	18,919.0	70,567.0	18,919.0	36,520.6
Fertilizer	99	180,464.0	2,039,392.0	1,109,928.0	1,314,461.0
NH3E	99	56,360.0	498,210.0	277,285.0	312,435.1
CO2E	99	1,150,132.0	10,899,927.0	6,025,029.0	6,894,147.0

2.2. Data Preprocessing

This research used an 80:20 ratio for data splitting. The training data were used for fitting the model or algorithm training. The testing data were used to validate the fitted models or the trained algorithm for prediction purposes [16]. The 10-Fold approach was chosen as it delivers good model assessment in ML [17] (Figure 1). Three model fit parameters were assessed: root mean square error (RMSE), mean absolute percentage value (MAPE), and mean absolute error (MAE).

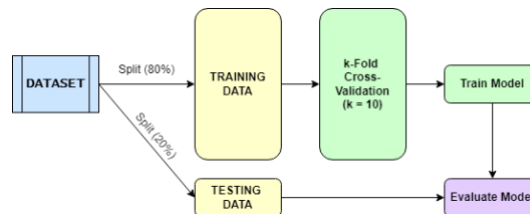


Figure 1. Dataset proportion for data training, testing, and cross-validation

2.3. Data Analysis

This research’s analytical and predictive models, developed for tuning and analyzing observed parameters to assess CAP’s impacts on the EU agricultural sector, were computed using an open-source R programming language software, RStudio, based on the research design shown in Figure 2. Microsoft Excel was utilized for data preprocessing, descriptive statistic calculation, and data storage.

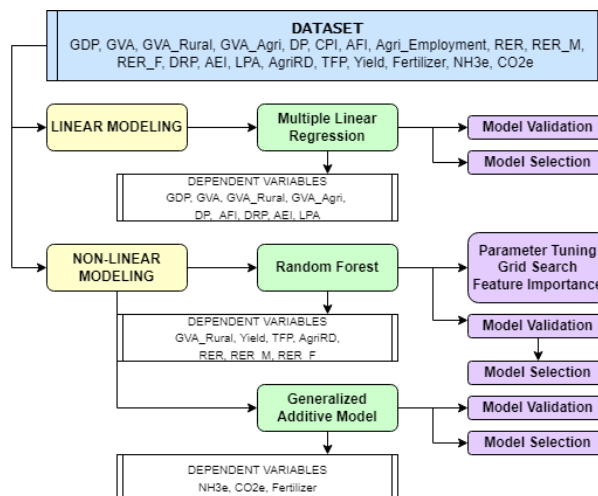


Figure 2. Dataset modeling research design

2.4. Linear Modeling

Regression analysis is used to model single dependent variable and one (simple regression) or more (multiple regression) independent variables [17]. In multiple linear regression (MLR), the model being fitted is linear (Equation 1) and it was computed using RStudio with package: lm and stepAIC. Feature selection step using stepwise regression (stepAIC) technique is used to identify the optimal linear models from variables being studied [16]. To determine the best-fitted model, regression related metrics were analyzed: R-sq and adjusted R-sq.

$$Y = \alpha_1 X_1 + \dots + \alpha_n X_n + \beta + e \quad (1)$$

2.5. Non-Linear Modeling

Many scholars have adopted non-linear ML method, which are more flexible and useful where econometric models fail to deliver relevant interpretability [18]. Developing ML models require the selection of optimal ‘features’ (variables/predictors) used to obtain better learning performance and more cost-effective data generation process [19]. Feature selection was performed for reducing model overfit, improving model accuracy, and providing predictors with interpretability results. This research used random forest (RF) and generalized additive model (GAM), computed using the RStudio with libraries: randomForest, gam, and mgcv. RF is an extension of decision tree where it will collect each tree classification and choose the most voted one as a result [16], [20]. RF has several hyperparameters that must be set in advance with parameter tuning [21] for a better learning process [22]. The result of parameter tuning was used in the algorithm for data training and testing in each RF model. GAM modeling has been applied in agricultural studies such as ecological, land allocation, climate-related issues [23], [24], and for predictive analytics of crop yield [25] or pesticide use [26]. The advantage of GAM stems from its interpretability which similar to MLR, presented as a sum of functions obtained from fitting the additive model, replacing the beta coefficient from linear regression with flexible functions that allow for non-linear modeling relationships [17].

3. RESULTS AND DISCUSSION

3.1 Comparative Analysis of Models within Three Dimensions of Sustainability: Social, Environment, and Economic

Gross Value Added (GVA). The analysis of GVA revealed notable disparities between EU countries (Figure 3b). GVA registered from the agricultural (consists of agriculture, forestry, fishing, and hunting) sector influence on the national economy, which presented by GDP, is considered moderate (Figure 3c). This emphasizes that most EU countries were not entirely dependent on agricultural activities. The result showed similar with previous study [7] evaluating rural economic accounts with financial data from 2007 to 2013. Hence, overall regional growth rate was stagnant. The CAP: 2014-20 reform has had a positive impact on rural GVA (Figure 3a), observed in all studied EU countries.

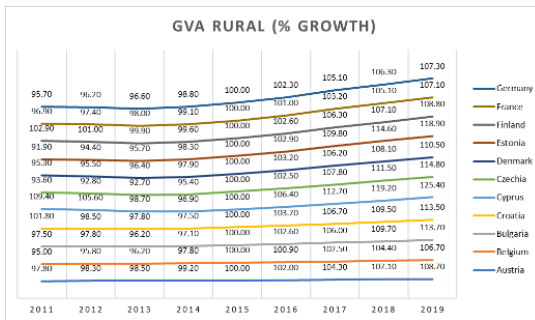
Agricultural Land and Productivity. In terms of average cereal crop yield, the best results were recorded in Belgium, followed by Finland and Austria (Figure 4a). This projects similarity with previous study [7], where Central and Western EU produced high crop yields, even though not all countries in the region live up to the same claim. For example, in the Western EU, France projects a low yield while it had higher land size and production quantity (Figure 3e). For this

case, other crops might contribute more to the total production. Thus, crop yield as a single measure of productivity could lead to a misleading indication of the degree of agricultural productivity [27]. The International Food Policy Research Institute (IFPRI) stated that the TFP index is considered the most suitable indicator to measure the agricultural system. From the data, the Eastern EU region had a higher TFP index (Figure 3e), even with lesser crop yields. This might be explained by lower production factor prices in the Eastern EU, as it had smaller agricultural land to manage (Figure 3f).

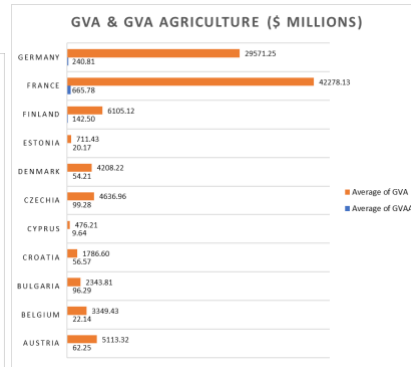
Direct Payment. Direct payment (DP) is considered the main instrument in the CAP aimed at supporting farmers' income and contributing to rural vitality. Many farms depend heavily on this subsidy as farming is a risky and costly business [28]. From the data, countries with high DP resulted in higher agricultural entrepreneurship income (AEI) (Figure 3d). It is also consistent with the size of agricultural land and the production quantity of all crops (Figure 3f). Similar to the previous study [7], Western EU received the largest amount of DP and had higher AEI.

Labor Productivity, Income, Fertilizer Use, and Emissions. Agricultural labor productivity (LPA) plays a fundamental role in economic development processes since higher LPA extends agricultural labor from food production to the production of other forms of goods and services [29]. High labor and land productivity are related to fertilizer usage and high income, respectively. High productivity was observed in Eastern EU countries, consistent with their high TFP (Figure 3e, 3h) and fertilizer usage (Figure 3g). However, the agricultural factor income (AFI) was lower. A possible explanation is related to smaller agricultural land that is linked to lesser DP subsidies granted. The fertilizer usage was synchronized with emissions from production activities (Figure 3g), except Bulgaria. Its emissions were lower compared to their fertilizer usage. This might indicate a better sustainable use of resources. This finding further emphasizes the country's high TFP index despite receiving a lesser amount of DP subsidies.

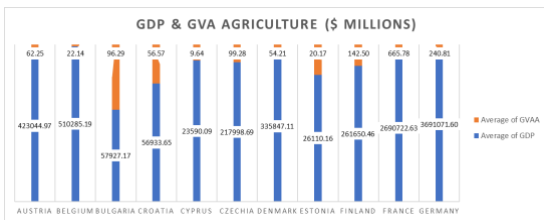
Consumer Price Index, Rural Poverty, Employment Rate, and R&D. Both degree of rural poverty (DRP) and consumer price evolution index (CPI) (Figure 3i), or in this context: food products, were lower in countries with lower AFI (Figure 3h) and higher DP subsidies (Figure 3d), such as Germany, France, and Denmark. While in Eastern EU countries, both DRP and CPI were higher. This reveals that the rural economy is strongly related to agricultural production activities as high food prices may affect the affordability of consumption, labor demand, and income. Noticeably in Figure 3j, the Western EU countries had higher government budget allocated for agriculture expenditure and R&D spending than others. The findings correlated with previous study [7], concluding that the trend has not changed even during CAP: 2014-20. To improve this situation, R&D to execute innovation is vital, and the government was proven to be the enabler for R&D outcomes [30].



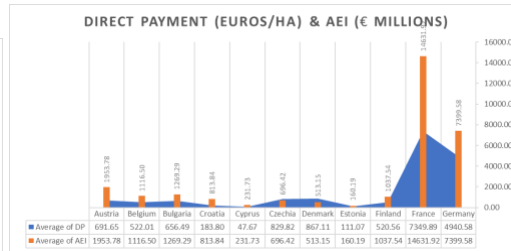
(a)



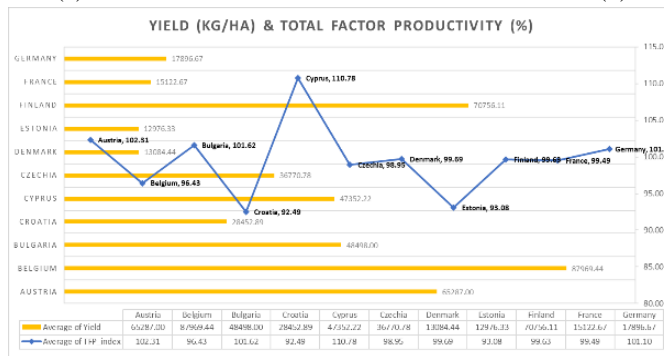
(b)



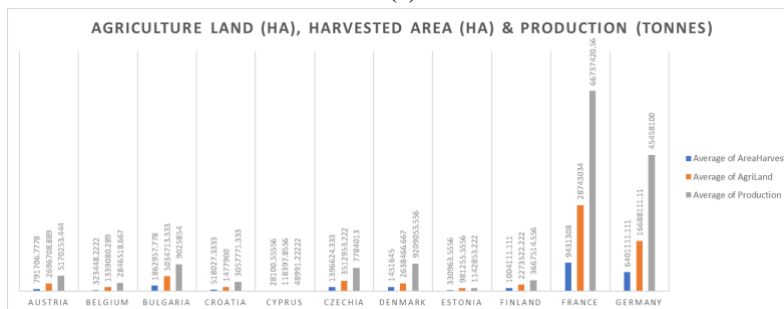
(c)



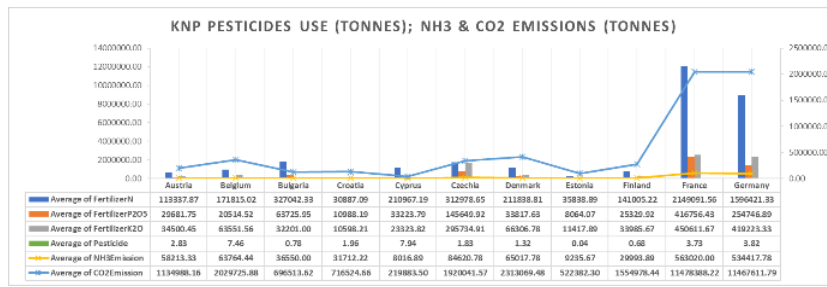
(d)



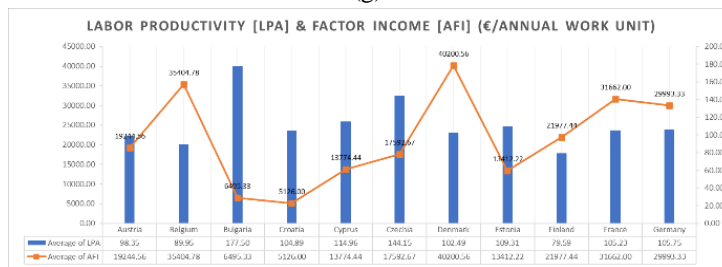
(e)



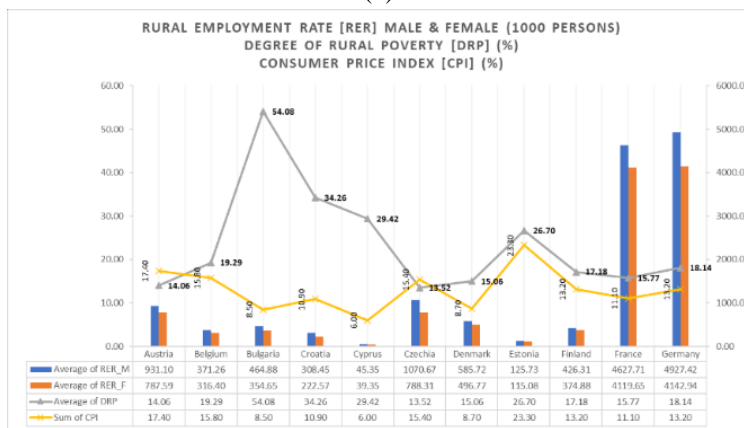
(f)



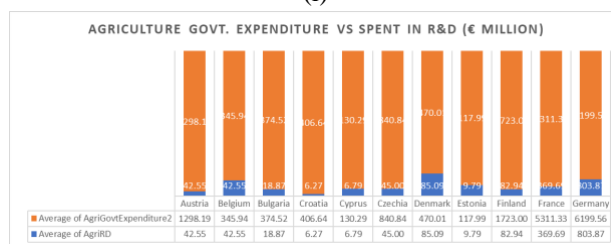
(g)



(h)



(i)



(j)

Figure 3. (a) GVA registered for rural areas (% growth); (b) total GVA and agricultural GVA (in \$ millions); (c) GDP and agricultural GVA (in \$ millions); (d) direct payment (EUR/ha) and agriculture entrepreneurship income (AEI) (in € millions); (e) cereal crop yield and agricultural TFP index; (f) land used for agriculture, the reported area of harvest, and total agriculture production; (g) fertilizers use, ammonia, and carbon dioxide emissions from agriculture; (h) labor productivity in agriculture (LPA) and agricultural factor income (AFI); (i) rural employment rate of male and female per 1000 people in the population, degree of rural poverty (DRP) [▲], and consumer price index (CPI) [X]; (j) government expenditure in agricultural sector and spending in agricultural R&D

3.2 The Social Dimension

The degree of rural poverty (DRP) is an important aspect to be understood as its reduction could bring benefits not only for the rural population but for the country. The model's linear regression (Equation 2) shows very good accuracy metrics, having R-sq at 87.2% and adjusted R-sq at 85.9%. The result indicates that DRP decreases with the increase in female employment rate, employment in agriculture, CPI, and GDP per capita registered for rural areas. Previous study [31] identified that the agricultural GVA and RER are important factors in alleviating rural poverty. However, considering Equation (2) below, the DP and LPA negatively influence the DRP. This might be due to the high employment rate that can decrease the sector's average income as well as its labor productivity [7].

$$\ln(\text{DRP}) = 0,442 - 0,008\ln(\text{AgriEmployment}) - 1,384\ln(\text{CPI}) - 0,0007\ln(\text{GDP_Rural}) + 0,002\ln(\text{DP}) + 0,045\ln(\text{LPA}) + 0,025\ln(\text{RER_M}) - 0,031\ln(\text{RER_F}) \quad (2)$$

GDP per capita (GDP) is a prosperity measure of a nation. Hence, it is an important parameter to be studied. The model has very good accuracy metrics: R-sq 85,5% and adjusted R-sq 84,8%. Considering Equation (3), the reduction in the DRP can improve the GDP per capita. This indicator is also affected by the consumer price evolution of food products or the consumer price index (CPI) [7]. While the rural population's income remains constant and the price increase, people's consumption automatically decreases. This will lead to a direct negative impact on the GDP. As presented on Figure 3c which compares the contribution of GVA_Agri to GDP, most EU countries were not dependent on their agricultural activities. In terms of employment rate and GDP per capita, other authors [32] reported a strong relation between the two.

$$\ln(\text{GDP}) = 55,715 - 8,3983\ln(\text{AgriEmployment}) - 974,8825\ln(\text{DRP}) - 1,234,8016\ln(\text{CPI}) \quad (3)$$

The 'GVA_Rural' parameter was modeled with random forest (RF). Through feature selection analysis, it was concluded that the following five parameters were the most important (based on their weights): DRP, GDP, Fertilizer, GDP_Rural, AFI (Table 2). The 'GDP' and 'Fertilizer' displays higher weights, indicating agriculture played a vital role in the development of agriculture activities in rural areas. 'AFI' is also observed to have a significant influence on rural GVA, enforcing the hypothesis for the European rural development. AFI is best suited for evaluating the impact of changes in the level of public support on the capacity of farmers to reimburse capital, pay for wages, and reward their production (Bank 2008). The CAP, in World Trade Organization definition, is a multifunctional policy—it gives numerous benefits for a country. In agricultural context, it includes food security, environmental protection, and rural employment [33]. As illustrated in Figure 3i, the RER for women is noticeably lower than the male one. Additionally, it is easy to spot countries with the highest RER, such as France and Germany, with also higher DP subsidies. Through RF, this research presents three rural employment rate (RER, RER_M, RER_F) parameters with the following characteristics described in Table 2. According to [34], large-scale commercial farms show little impact on the reduction of rural poverty compared to areas dominated by small farms, which provide job opportunities for the locals. Thus, in the CAP: 2013-27 reform [35], targeting a fairer distribution of DP was the right move. DP being one of the important variables is shown in the RER model (Table 2).

Table 2. Random Forest Feature Importance and Its Weight

Variable	Important Features
RER	7,73 for GDP; 7,55 for RER_F; 7,47 for RER_F; 6,99 for ARD; 6,59 for DP
RER_M	12,13 for RER; 10,09 for DP; 9,98 for RER_F; 9,43 for Fertilizer; 9,33 for NH3E
RER_F	11,69 for AEI; 11,27 for RER_M; 10,19 for GDP; 10,14 for RER; 9,16 for ARD
Yield	7,79 for TFP; 5,41 for GDP_Rural; 3,85 for AEI; 3,41 for LPA; 3,29 for AgriEmployment
TFP	4,84 for ARD; 4,37 for Yield; 4,33 for CPI; 3,29 for AEI; 2,48 for CO2E
GVA_Rural	1,55 for DRP; 1,14 for GDP; 0,89 for Fertilizer; 0,56 for GDP_Rural; 0,53 for AFI

3.3 The Environment Dimension

Fertilizers are being used by the farmers to help crops to produce more yield. But still, the excessive use of fertilizer could result in environmental degradation [36]. ‘Fertilizer’ was modeled using MLR. The model on Equation 4 below displayed a satisfactory R-sq of 98,5% and an adjusted R-sq of 97,9%.

$$\text{Ln}(\text{Fertilizer}) = -0,00006 - 0,425\text{Ln}(\text{GDP}) + 0,143\text{Ln}(\text{GVA}) + 0,001\text{Ln}(\text{LPA}) + 6,330\text{Ln}(\text{NH3E}) + 0,004\text{Ln}(\text{TFP}) \quad (4)$$

Theoretically, higher TFP values are related to lower inputs (e.g., fertilizer) usage due to efficiency improvements [37]. Although the model’s beta coefficient is low, it is worth an investigation. Taken from a study [38] that used the US Department of Agriculture data, the authors stated that measuring TFP growth is challenging, and knowledge gaps still exist in understanding indicators contributed to the TFP index measurement, as there are inputs other than fertilizer usage. This research emphasizes that the agricultural sector’s ammonia (NH₃) emission is directly affected by agricultural investment, such as DP. Additionally, fertilizer usage, labor productivity, GDP, and GVA positively influence the rate of NH₃ emission. This could be explained by when the DP were high, farmers purchased more agriculture inputs which resulted in higher productivity. However, this model has peculiarities. The DRP and employment in agriculture are observed to have a negative influence on NH₃ emission. This implies that an increase in agricultural employment rate will decrease emissions. A possible explanation might be when there are more people employed in the agricultural sector, the use of fertilizers will decrease as people will depend more on physical labor in farming. A similar finding was also found in [39], where farmers used less fertilizers in a situation where the rural employment rate is high.

Equation (5) delivers excellent accuracy metrics with R-sq at 99,7% and adjusted R-sq at 99,6%. The environmental dimension includes carbon dioxide (CO₂) emissions, as several studies showed that agricultural activities are responsible for an overall increase in carbon emissions [40], [41], much more than total NH₃ emitted from the sector (Figure 6a). A study [40] stressed the need for a reliable quantification to assess necessary resource allocation for climate research and land-use management in the agricultural context.

$$\text{Ln}(\text{NH3E}) = -0,0005 + 0,134\text{Ln}(\text{DP}) - 0,022\text{Ln}(\text{DRP}) - 3,240\text{Ln}(\text{AgriEmployment}) + 0,059\text{Ln}(\text{Fertilizer}) + 0,072\text{Ln}(\text{GDP}) + 2,499\text{Ln}(\text{GVA}) + 0,047\text{Ln}(\text{LPA}) \quad (5)$$

In Equation (6), computed using MLR, the CO₂ emission was influenced by several parameters: DP, GDP, agricultural R&D, employment rate, and industrialization aspects. The above model has excellent accuracy metrics with R-sq at 99,7% and adjusted R-sq at 99,5%. Countries with higher TFP index and labor productivity will lower their CO₂ emission through modern technology utilization to increase agricultural productivity. The adaptation of R&D innovation gives the opportunity to provide good quality seeds, fertilizer, and better machinery to farmers, which enhances the level of food security by increasing productivity, availability, and accessibility to food products [30].

$$\text{Ln}(\text{CO}_2\text{E}) = 0,000004 + 0,003\text{Ln}(\text{ARD}) + 0,077\text{Ln}(\text{DP}) + 0,223\text{Ln}(\text{AgriEmployment}) + 0,919\text{Ln}(\text{Fertilizer}) + 0,688\text{Ln}(\text{GDP}) - 0,007\text{Ln}(\text{LPA}) - 0,452\text{Ln}(\text{RER}) - 0,0003\text{Ln}(\text{TFP}) \quad (6)$$

3.4 The Economic Dimension

The present research modeled GVA in agriculture (GVA_Agri) with the GAM method as MLR gave low accuracy level due to non-normal data distribution. The model of ‘GVA_Agri’ gave a satisfactory model fit with an adjusted R-sq of 98,6%. As the model shows, the GVA_Agri parameter is significantly influenced by: DP, LPA, NH₃ emission, crop yield, and employment in agriculture (Equation 7).

$$\text{GVA_Agri} = \text{s}(\text{DP}) + \text{s}(\text{LPA}) + \text{s}(\text{NH}_3\text{E}) + \text{s}(\text{Yield}) + \text{s}(\text{AgriEmployment}) \quad (7)$$

In [42], the authors conducted the study across EU-27 countries using the k-means clustering method resulted in the agricultural production intensities after extra investments given. The overall ‘GVA’ parameter can be expressed through MLR with satisfactory model accuracy of R-sq at 98,9% and adjusted R-sq at 98,8%. According to the model, an increase in GDP, DP, employment in agriculture, and the sector’s GVA would lead to the overall ‘GVA’ increase. An increase in government expenditure on R&D will decrease the GVA. A possible explanation for this finding is the unclear cut of R&D contribution to agricultural productivity. It is reported [43] that while there is a positive relationship between knowledge created from R&D investment and productivity growth, there is still a challenge to exploit the knowledge with capability to capture the utmost benefits of the investment. Therefore, a high level of investment in R&D does not guarantee a positive return in GVA.

$$\text{Ln}(\text{GVA}) = 0,016 - 0,0007\text{Ln}(\text{CO}_2\text{E}) - 0,212\text{Ln}(\text{ARD}) + 0,009\text{Ln}(\text{GDP}) + 4,097\text{Ln}(\text{DP}) + 0,113\text{Ln}(\text{GVA_Agri}) + 0,531\text{Ln}(\text{AgriEmployment}) \quad (8)$$

From Equation (8) above, it is observed that CO₂ emission negatively impacts GVA. Similar finding can be seen in [44], where it reported higher GVA in agriculture negatively impacted CO₂ emission. In contrast, industrial and service sector positively impacting CO₂ emission. The crop yield measures agricultural production harvested per unit of land area (kg/ha). It helps optimize food security programs as part of sustainable development goals. Modeling crop yield is necessary as it offers the policymakers with timely information for rapid decision-making during the growing season in relation to trade and policy changes related to food security. Previous authors [30] modeled crop yield through RF, and the current research adopted the same method. The ‘Yield’ predictive analytic model consists of the following important indicators based on its weight: TFP, GDP_Rural, AEI, LPA, AgriEmployment (Table 2). Intensive

agriculture activity associated with a high value of crop yield will generate higher LPA due to the high automatization degree of production processes included [7]. This implies that higher TFP index results in higher yield as it measures productivity, a critical condition for the sector's development and the overall economy. Thus, 'TFP' was modeled with RF, resulting in ARD, Yield, CPI, AEI, CO2 as important indicators to study (Table 2). Similar findings from [9] support this result where agriculture R&D clearly influences the TFP index. The unavailability of appropriate machinery and farmers' dependency on traditional methods were found as the main barrier to production growth [14].

The CAP supports the agriculture sector development through a combination of DP schemes to farmers, financial aid in rural development, and environmental protection measures. DP is meant to support European farmers and receive even more attention during the CAP: 2014-20 to accelerate the transition towards green agriculture [45], [46]. Other authors [47] examined the role of government intervention for the financial efficiency of the organic farms, resulting in DP indeed playing a vital role in the financial viability of organic farms.

$$\text{Ln}(\text{DP}) = 0,398 + 0,086\text{Ln}(\text{GVA}) + 0,0005\text{Ln}(\text{CO2E}) - 0,152\text{Ln}(\text{AgriEmployment}) - 0,007\text{Ln}(\text{AFI}) + 0,185\text{Ln}(\text{RER}) - 0,001\text{Ln}(\text{GDP}) \quad (9)$$

Based on the previously described dataset, the 'DP' parameter was modeled through MLR approach. The model (Equation 9) shows excellent accuracy metrics with R-sq at 98.9% and an adjusted R-sq of 98,8%. In [48], DP can generate negative factors regarding rural sustainability, this might explain the negative influence of GDP to the model. A significant financial influx into unprepared economies might cause a deficit of skilled labor forces, which eventually cause a decrease in GDP. Similar with [7], the DP value increases with the increase of GVA, CO2, and RER, which represent productivity. However, the negative influence of AFI on DP might explain the need for a fairer DP redistribution [35] in the future CAP: 2023-27 reform.

Agriculture R&D (ARD) investment importance was emphasized in several studies [30]. The 'ARD' parameter was modeled using RF and shows excellent accuracy with an adjusted R-sq of 99.9%. In Equation (10), it can be noticed that overall GVA, DP, GDP per capita registered for rural areas, agriculture GVA, and CO₂ emission from agricultural production are important indicators in predicting ARD. Agricultural R&D is critical for ensuring food security in the coming decades; thus, policymakers should pay more attention to R&D and its impact on agriculture sector performance and overall economic growth.

$$\text{ARD} = s(\text{GVA}) + s(\text{DP}) + s(\text{GDP_Rural}) + s(\text{GVA_Agri}) + s(\text{CO2E}) + s(\text{GDP}) \quad (10)$$

Although the development and practice of CAP have been using an instrument that covers the wide range of policy options, there are still issues found regarding its implementation, which led to the sector's low economic, social, and environmental sustainability [49], [50]. ML technique can offer the ability to measure the policy's impact in real-time and geo-specific manner. The challenge for policymakers at different administrative levels is to capture all heterogeneous territorial contexts in the EU and to include the spatial dimension in future policy development [51]. The synergic approach of applying ML and big data utilization could reduce information asymmetry and the transactional cost of policy implementation [52], as stated in [53], an integrated CAP is a key to achieve food security both in the EU and the UN sustainable development goals.

4. CONCLUSION

The present research proposed predictive analytic models, both linear and non-linear, characterized by very good to excellent performance metrics computed through machine learning algorithms. In this context, future iterations of this research should consider re-training the models based on the extended dataset to reconfirm and enhance its predictive strength. Other than contributing to the agriculture and sustainability literature, this research offers managerial implications of possible policy impact monitoring and controlling tools for strategy evaluation. Additionally, non-European countries with similar or plan to adopt the CAP could refer to this research findings for policy development. This study is not without limitations. Firstly, since CAP has multidimensional implications, future research should consider exploring more variables and try to model the phenomenon with another approach, such as structural equation modeling, with the aim of capturing the complexity. Secondly, the CAP is specific to European countries in nature, so does the results. However, the developed models can be extended, and the research methodology can be adopted to further geo-specific elaboration or more advanced model development. The frameworks developed were intended to have general specificity since they are categorized as monitoring and controlling tools based on predictive analytic algorithms. Therefore, other countries which plan to develop similar policy could refer to CAP studies including this research.

REFERENCES

- [1] UNDP, “Goal 2: Zero hunger,” 2022. https://www1.undp.org/content/seoul_policy_center/en/home/sustainable-development-goals/goal-2-zero-hunger.html (accessed May 24, 2023).
- [2] FAO, IFAD, UNICEF, WFP, and WHO, “The State of Food Security and Nutrition in the World 2020”, vol. 1. FAO, IFAD, UNICEF, WFP and WHO, 2020.
- [3] W. Ma and A. Abdulai, “IPM adoption, cooperative membership and farm economic performance: Insight from apple farmers in China,” *China Agric Econ Rev*, vol. 11, no. 2, pp. 218–236, Jun. 2019, doi: 10.1108/CAER-12-2017-0251/FULL/XML.
- [4] B. Purvis, Y. Mao, and D. Robinson, “Three pillars of sustainability: in search of conceptual origins,” *Sustain Sci*, vol. 14, no. 3, pp. 681–695, 2019, doi: 10.1007/S11625-018-0627-5/FIGURES/1.
- [5] G. Fusco, “Twenty Years of Common Agricultural Policy in Europe: A Bibliometric Analysis,” *Sustain* 2021, vol. 13, no. 19, p. 10650, 2021, doi: 10.3390/SU131910650.
- [6] T. Grodzicki and M. Jankiewicz, “The role of the common agricultural policy in contributing to jobs and growth in EU’s rural areas and the impact of employment on shaping rural development: Evidence from the Baltic States,” *PLoS One*, vol. 17, no. 2, p. e0262673, 2022, doi: 10.1371/JOURNAL.PONE.0262673.
- [7] D. S. Cristea *et al.*, “Modelling the Common Agricultural Policy Impact over the EU Agricultural and Rural Environment through a Machine Learning Predictive Framework,” *Agron* 2021, vol. 11, no. 11, p. 2105, Oct. 2021, doi: 10.3390/AGRONOMY11112105.
- [8] N. Hayat, A. Al Mamun, N. A. M. Nasir, G. Selvachandran, N. B. C. Nawi, and Q. S. Gai, “Predicting Sustainable Farm Performance—Using Hybrid Structural Equation Modelling with an Artificial Neural Network Approach,” *L* 2020, vol. 9, no. 9, p. 289, 2020, doi: 10.3390/LAND9090289.

- [9] J. H. Jeong *et al.*, “Random Forests for Global and Regional Crop Yield Predictions,” *PLoS One*, vol. 11, no. 6, 2016, doi: 10.1371/JOURNAL.PONE.0156571.
- [10] M. Wielechowski, “GOVERNMENT EXPENDITURE ON AGRICULTURE – A EUROPEAN, REGIONAL AND WORLD PERSPECTIVE,” *Ann Polish Assoc Agric Agribus Econ*, vol. XXI, no. 4, pp. 561–570, 2019, doi: 10.5604/01.3001.0013.5732.
- [11] OECD, *Growing Unequal?*, vol. 9789264044197. OECD, 2008.
- [12] J. A. Gómez-Limón and G. Sanchez-Fernandez, “Empirical evaluation of agricultural sustainability using composite indicators,” *Ecol Econ*, vol. 69, no. 5, pp. 1062–1075, 2010, doi: 10.1016/J.ECOLECON.2009.11.027.
- [13] UN SDG, “2.a.1 Agriculture Orientation Index for Government Expenditures - Global SDG Indicator Platform,” 2022. <https://sdg.tracking-progress.org/indicator/2-a-1-agriculture-orientation-index-for-government-expenditures-2/> (accessed May 24, 2023).
- [14] P. Conceição, S. Levine, M. Lipton, and A. Warren-Rodríguez, “Toward a food secure future: Ensuring food security for sustainable human development in Sub-Saharan Africa,” *Food Policy*, vol. 60, pp. 1–9, 2016, doi: 10.1016/J.FOODPOL.2016.02.003.
- [15] D. Weziak-Bialowolska, “Spatial Variation in EU Poverty with Respect to Health, Education and Living Standards,” *Soc Indic Res*, vol. 125, no. 2, p. 451, 2016, doi: 10.1007/S11205-014-0848-7.
- [16] N. Salkind, “Encyclopedia of Research Design,” *Encycl Res Des*, 2012, doi: 10.4135/9781412961288.
- [17] A. J. Izenman, *Modern Multivariate Statistical Techniques*. New York, NY: Springer New York, 2008.
- [18] H. Storm, K. Baylis, and T. Heckelei, “Machine learning in agricultural and applied economics,” *Eur Rev Agric Econ*, vol. 47, no. 3, pp. 849–892, 2020, doi: 10.1093/ERAEB/JBZ033.
- [19] I. Guyon and A. Elisseeff, “An Introduction to Variable and Feature Selection,” *J of Machine Learn Res*, vol. 1, pp. 1157–1182, 2003, doi: 10.1162/153244303322753616.
- [20] L. Breiman, “Random Forests,” *Mach Learn* 2001 451, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [21] P. Probst, M. Wright, and A.-L. Boulesteix, “Hyperparameters and Tuning Strategies for Random Forest,” *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 9, no. 3, 2018, doi: 10.1002/widm.1301.
- [22] E. Scornet, “Tuning parameters in random forests,” *ESAIM Proc Surv*, vol. 60, pp. 144–162, 2017, doi: 10.1051/PROC/201760144.
- [23] J. J. Lawler, D. White, R. P. Neilson, and A. R. Blaustein, “Predicting climate-induced range shifts: model differences and model reliability,” *Glob Chang Biol*, vol. 12, no. 8, pp. 1568–1584, Aug. 2006, doi: 10.1111/J.1365-2486.2006.01191.X.
- [24] R. Stauffer, G. J. Mayr, J. W. Messner, N. Umlauf, and A. Zeileis, “Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model,” *Int J Climatol*, vol. 37, no. 7, pp. 3264–3275, 2017, doi: 10.1002/JOC.4913.
- [25] K. Chen, R. A. O’Leary, and F. H. Evans, “A simple and parsimonious generalised additive model for predicting wheat yield in a decision support tool,” *Agric Syst*, vol. 173, pp. 140–150, 2019, doi: 10.1016/J.AGSY.2019.02.009.
- [26] J. A. Rosenheim, B. N. Cass, H. Kahl, and K. P. Steinmann, “Variation in pesticide use

- across crops in California agriculture: Economic and ecological drivers,” *Sci Total Environ*, vol. 733, p. 138683, 2020, doi: 10.1016/J.SCITOTENV.2020.138683.
- [27] J. D. Mullen and T. L. Cox, “Measuring productivity growth in Australian broadacre agriculture,” *Aust J Agric Resour Econ*, vol. 40, no. 3, pp. 189–210, 1996, doi: 10.1111/J.1467-8489.1996.TB00594.X.
- [28] European Commission, “Income support explained,” 2022. https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/income-support/income-support-explained_en (accessed May 24, 2023).
- [29] A. Dorward, “Agricultural labour productivity, food prices and sustainable development impacts and indicators,” *Food Policy*, vol. 39, pp. 40–50, 2013, doi: 10.1016/J.FOODPOL.2012.12.003.
- [30] S. Kousar, F. Ahmed, A. Pervaiz, and Š. Bojnec, “Food Insecurity, Population Growth, Urbanization and Water Availability: The Role of Government Stability,” *Sustain* 2021, vol. 13, no. 22, p. 12336, 2021, doi: 10.3390/SU132212336.
- [31] M. Q. Dao, “Rural poverty in developing countries: An empirical analysis,” *J Econ Stud*, vol. 31, no. 6, pp. 500–508, 2004, doi: 10.1108/01443580410569244/FULL/XML.
- [32] A. G. Papadopoulos, “The Impact of the CAP on Agriculture and Rural Areas of EU Member States:,” vol. 4, no. 1, pp. 22–53, 2015, doi: 10.1177/2277976015574054.
- [33] A. Spaziante, “Rural Development. Challenges and opportunities for Europe,” *E3S Web Conf*, vol. 2, p. 03005, 2014, doi: 10.1051/E3SCONF/20140203005.
- [34] J. W. Mellor and S. J. Malik, “The Impact of Growth in Small Commercial Farm Productivity on Rural Poverty Reduction,” *World Dev*, vol. 91, pp. 1–10, Mar. 2017, doi: 10.1016/J.WORLDDEV.2016.09.004.
- [35] S. Ciliberti and A. Frascarelli, “The CAP 2013 reform of direct payments: redistributive effects and impacts on farm income concentration in Italy,” *Agric Food Econ*, vol. 6, no. 1, pp. 1–18, 2018, doi: 10.1186/S40100-018-0113-5/TABLES/8.
- [36] X. Ju, B. Gu, Y. Wu, and J. N. Galloway, “Reducing China’s fertilizer use by increasing farm size,” *Glob Environ Chang*, vol. 41, pp. 26–32, Nov. 2016, doi: 10.1016/J.GLOENVCHA.2016.08.005.
- [37] S. P. Dimelis and M. Dimopoulou, “Evaluating Productivity Growth Measures in the EU,” *Econ Plan* 2002 352, vol. 35, no. 2, pp. 161–181, 2002, doi: 10.1023/A:1020121503587.
- [38] Y. Gu *et al.*, “Consequences of agricultural total factor productivity growth for the sustainability of global farming: accounting for direct and indirect land use effects,” *Environ Res Lett*, vol. 14, no. 12, p. 125002, 2019, doi: 10.1088/1748-9326/AB4F57.
- [39] R. L. Lamb, “Fertilizer Use, Risk, and Off-Farm Labor Markets in the Semi-Arid Tropics of India,” *Am J Agric Econ*, vol. 85, no. 2, pp. 359–371 2003, doi: 10.1111/1467-8276.00125.
- [40] C. Oertel, J. Matschullat, K. Zurba, F. Zimmermann, and S. Erasmi, “Greenhouse gas emissions from soils—A review,” *Geochemistry*, vol. 76, no. 3, pp. 327–352, 2016, doi: 10.1016/J.CHEMER.2016.04.002.
- [41] K. Paustian, C. V. Cole, D. Sauerbeck, and N. Sampson, “CO2 Mitigation by Agriculture: An Overview,” *Clim Chang* 1998 401, vol. 40, no. 1, pp. 135–162, 1998, doi: 10.1023/A:1005347017157.
- [42] E. Giannakis and A. Bruggeman, “The highly variable economic performance of European

- agriculture,” *Land use policy*, vol. 45, pp. 26–35, 2015, doi: 10.1016/J.LANDUSEPOL.2014.12.009.
- [43] J. Hall and G. M. Scobie, “The Role of R&D in Productivity Growth: The Case of Agriculture in New Zealand: 1927 to 2001 (WP 06/01),” 2006. <https://www.treasury.govt.nz/publications/wp/role-rd-productivity-growth-case-agriculture-new-zealand-1927-2001-wp-06-01-html> (accessed May 24, 2023).
- [44] J. Alam, “Impact of Agriculture, Industry and Service Sector’s Value Added in the GDP on CO2 Emissions of Selected South Asian Countries,” 2015.
- [45] A. Buckwell, A. Matthews, D. Baldock, and E. Mathijs, “About the RISE Foundation,” Brussels, 2017.
- [46] C. Nazzaro and G. Marotta, “The Common Agricultural Policy 2014–2020: scenarios for the European agricultural and rural systems,” *Agric Food Econ*, vol. 4, no. 1, 2016, doi: 10.1186/S40100-016-0060-Y.
- [47] F. Offermann, H. Nieberg, and K. Zander, “Dependency of organic farms on direct payments in selected EU member states: Today and tomorrow,” *Food Policy*, vol. 34, no. 3, pp. 273–279, 2009, doi: 10.1016/J.FOODPOL.2009.03.002.
- [48] M. Morkunas and P. Labukas, “The Evaluation of Negative Factors of Direct Payments under Common Agricultural Policy from a Viewpoint of Sustainability of Rural Regions of the New EU Member States: Evidence from Lithuania,” *Agric 2020*, vol. 10, no. 6, p. 228, 2020, doi: 10.3390/AGRICULTURE10060228.
- [49] A. Navarro and J. V. López-Bao, “Towards a greener Common Agricultural Policy,” *Nat Ecol Evol* 2018 212, vol. 2, no. 12, pp. 1830–1833, 2018, doi: 10.1038/s41559-018-0724-y.
- [50] G. Pe’er *et al.*, “A greener path for the EU Common Agricultural Policy,” *Science* (80-), vol. 365, no. 6452, pp. 449–451, 2019, doi: 10.1126/SCIENCE.AAX3146/SUPPL_FILE/AAX3146_RAWDATA.XLSX.
- [51] S. Alexiadis, C. Ladias, and N. Hasanagas, “A regional perspective of the Common Agricultural Policy,” *Land use policy*, vol. 30, no. 1, pp. 665–669, 2013, doi: 10.1016/J.LANDUSEPOL.2012.05.013.
- [52] M. H. Ehlers, R. Huber, and R. Finger, “Agricultural policy in the era of digitalisation,” *Food Policy*, vol. 100, p. 102019, 2021, doi: 10.1016/J.FOODPOL.2020.102019.
- [53] F. Recanati, C. Maughan, M. Pedrotti, K. Dembska, and M. Antonelli, “Assessing the role of CAP for more sustainable and healthier food systems in Europe: A literature review,” *Science of the Total Environment*, vol. 653. Elsevier B.V., pp. 908–919, 2019, doi: 10.1016/j.scitotenv.2018.10.377.
- [54] European Commission, “Common monitoring and evaluation framework,” 2023, https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cmef_en, (accessed January 10, 2024).