Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development
Vol. 21, Issue 2, 2021
PRINT ISSN 2284-7995, E-ISSN 2285-3952

# THE IMPACT OF THE COMMON AGRICULTURAL POLICY ON THE EUROPEAN AGRICULTURAL SECTOR SUSTAINABILITY BY USING A MACHINE-LEARNING APPROACH

**Ștefan-Mihai PETREA, Alina MOGODAN, Ira-Adeline SIMIONOV, Aurelia NICA, Dragoș Sebastian CRISTEA, Mihaela NECULIȚĂ**

"Dunărea de Jos" University of Galați, 47 Domnească Street, Galați, Romania, Emails: Stefan.Petrea@ugal.ro, alina.antache@ugal.ro, ira.simionov@gmail.com, anica@ugal.ro, dragoscristea@yahoo.com, neculitam@yahoo.fr

***Corresponding author***: stefan.petrea@ugal.ro

***Abstract***

*The present research evaluates the sustainability of EU agricultural sector in relation to Common Agricultural Policy by using a custom-developed analytical framework, based on relevant indicators: gross domestic product in rural areas (GDP), gross value added (GVA), GVA for agriculture, direct payments (DP), agricultural factor income (AFI), agriculture employment rate (AER), rural employment rate (RER), degree of rural poverty (DRP), agricultural entrepreneurial income (AEI), agriculture research and development investments (ARD), labour productivity in agriculture (LPA), total factor productivity (TFP), cereal crop yield (CCY), fertilizers use (FU), $CO_2$ emissions and ammonia emissions (EA). The EA are mostly related to CCY (+0.47%), DP (+0.45%), ARD (+0.49%), AFI (-0.59%) and GDP (-0.52%). The $CO_2$ emissions are influenced mostly by TFP (+2.82%), RER (+2.49%), DRP (+0.53%) and EA (+0.46%). The FU most significant feature importance weights are CCY (0.79%), $CO_2$ (0.21%) and TFP (0.19%). The GVA model is related mostly to RER (+1.30%), GVA Agriculture (+0.67%), AER (-0.35%) and FU (+0.31%). The ARD most significant feature importance weights are GVA (0.23), DP (0.16) and GVA agriculture (0.15). The results revealed excellent performance metrics.*

**Key words:** *agricultural sector, sustainability, gross value added, analytical framework, Common Agricultural Policy*

## INTRODUCTION

Agricultural expenditures represent an important component of the European Union budget. Still, as it could be observed during the last E.U. financial exercises, this tendency seems to be diminishing. The budgetary cost of the Common Agricultural Policy (CAP) in relation to the gross national income (GNI) of the European Union (EU) decreased from 0.54% in year 1990 to 0.34% in year 2020 (European Parliament, 2019). The CAP is the only policy mainly financed from the budget of the EU, and, given the high degree of integration with the other policies of the Union, it has become a catalyst for the development, providing good practice examples in areas coordinated by other policies as Rural Development Policy (RDP). According to the Regulation of the Council of Europe no. 1290/2005 regarding financing under the Common Agricultural Policy, two European funds were set up to finance agriculture: EAGF (European Agricultural Guarantee Fund) for financing agricultural markets and EAFRD (European Agricultural Fund for Rural Development) implemented to finance rural development programs. The EAGF is used to finance expenditure jointly managed by the Member States and the European Commission and centrally managed expenditure by the European Commission, while EAFRD finances the rural development programs implemented in accordance with Regulation (EC) no. 1698/2005 of the European Council, only if the expenses are jointly managed.

The influence of CAP over national economies was analysed in several research papers [18, 24, 33].

Thus, some authors [24] models the impact of policy measures within the CAP on farm production, income and prices. Also, the study of [24] describes how CAP subsidies and regulations are conducting changes in land use and land-use intensity. Thus, this methodology linked the impacts of economic policy instruments with changes in agricultural production, prices and incomes, having land-use and land-use intensity as the connecting drivers. Other authors [18] identifies the impact of CAP payments on crop diversity of agricultural holdings, in different Slovakian regions. The study [18] used an econometric approach and concluded that the second pillar of the CAP can improve the crop diversity in Slovakia. Also, according to this research [18], the most important factors influencing crop diversity are total crop area, irrigation, geographical location including the soil quality and the legal structure of agricultural holdings, while not important factors are found as managers education, distance of the farm from the city and membership in farmers' associations. In other study [33] is investigated the integration of environmental objectives into the CAP by conducting an economic lab-in-the-field experiment with farmers, in Germany. Thus, the research [33] analysed the impact of different policies on farmers' decisions to adopt sustainable agricultural practices.

The use of machine learning (ML) modelling techniques for improving farm management [16] or performing economic analysis in a more broadly manner [3, 6] has significant potential. Thus, as presented in other papers [13], a type of approach is to assess the influence each explanatory variable on the resulting prediction. Other research [9] targeted to determine the lowest change of an explanatory variable that would cause changes to certain model prediction. [17] had identified the characteristics of an observation that need to be changed for avoiding false predictions. Normally, the model complexity depends on the studied phenomenon. Many phenomena in agricultural economics are inherently non-linear, due to different social, economic or biophysical processes. Thus, some research subjects which confirmed this hypothesis are, as follows: the effect of weather variables on crops yield [27], the groundwater extraction effect on pumping costs [4] or the health effects of pollution [37]. Other study [31] have focused on the estimation of heterogeneity specific aspects, as understanding the distributional effects of an intervention (identifying the variables which reduce consumption in response to food warnings). Other research [32] also emphasizes on the fact that economic theory does not provide clear insights over specific functional form of the estimations, but only information about shape restrictions such as curvature or monotonicity. As such, it is very important to identify models that are able to capture non-linearities, as to avoid misspecification bias. It was emphasized that ML models are extremely flexible and may be helpful in various settings where classical econometric models fail to perform. He identifies three different approaches being particularly relevant to applied economists: ensembles of trees, particularly random forests (RF) and gradient boosting approaches; neural networks; variational inference methods. First two methods are flexible and efficient ML methods that can be applied to a large variety of tasks, while the third one is relevant in a Bayesian context. Several studies [14, 19, 36] revealed that the first two methods are effective prediction tools for scenarios like credit scoring and corporate bankruptcy prediction. Other study [8] develops a Bayesian geo-additive quantile regression model, estimated with gradient boosting.

For the agricultural economics, some studies [21] also used ensemble methods to predict the farmland rental rates. Thus, considering this situation, automatic data-driven parameter selection was applied. The neural networks as being also capable of capturing highly non-linear relationships [14]. An important difference between neural networks and tree-based methods like RF comes from the neural network complexity that is requiring more attributes to be specified, like number of layers and neurons. Also, another important difference comes from the number of samples required to determine a feasible model, which in the case of neural network should be high. A study [14] revealed that with cross-sectional data, in several cases, neural networks were outperformed by tree-based methods. Still, when enough data is available, neural networks could discover more complex non-linear dynamic relationships. Other study [12] used ensemble trees, respectively RF, to provide accurate prediction models for crop yield for enhancing food policies at the regional and global scales. Thus, in this study [12], RF model for predicting crop yield responses to climate and biophysical variables, at global and regional scales, in wheat, maize, and potato, was evaluated.

Thus, in this context, the present research provides an analytical and predictive framework, based on state-of-the-art ML algorithm, that evaluates the sustainability of EU agricultural sector in relation to CAP.

## MATERIALS AND METHODS

*Multiple linear regression and model selection*

The current research proposed framework was implemented using Python programming language with various specific libraries and the Minitab statistical software. The approaches towards data modelling were performed by using multiple linear regression (MLR) for assessing the relation between one dependent variable and several independent variables, fitting the data set into a linear equation (eq. 1), but also by applying tree-based ensemble learning models, respectively RF algorithms, as the predicted results were very conclusive and, also, the parameter importance was a lot easier to apprehend.

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_p X_p + \beta + e \quad ..(1)$$

The resulted algorithms were fed only with important features that can explain the dependent variable. The feature selection was used in order to reduce model complexity, improve model accuracy by selecting the right predictors subset and, also, reduce overfitting. Stepwise regression and best subsets were considered in implementing feature selection for the linear modelling. Stepwise technique assures the inclusion of significance level that will be used to accept or not a parameter. After applying 'best subsets' technique, interpreting the key results is required. By default, the model with the highest $R^2$ for one predictor, two predictors and so on can be chose, where each presented model has several defining predictors 'X' marked.

Selecting the final model requires further examination of residual plots. The histogram of residuals is used to determine if data is skewed or outliers exists. Normal probability plot of residuals verifies the assumption that the residuals are normally distributed. Residuals versus fits is applied to verify the assumption that the residuals have a constant variance and residuals versus order of data is applied to verify the assumption that the residuals are uncorrelated with each other.

Durbin-Watson value tests are applied in order to check if the errors of adjacent observations are correlated. The best model is selected after the interpretation of the statistics related to Best Subsets Regression technique.

*Non-linear prediction models and feature importance determination with RF*

The RF Regression, used for the current study, represents one of the best tools in the ML repertoire due to its high accuracy and ability to handle a large number of features when few samples are available.

The RF aggregates two concepts, respectively Bagging and Random Selection of Features, generating a set of T regression trees where the training set for each tree is selected based on Bootstrap sampling from the original sample set and the features considered for partitioning at each node is a random subset of the original set of features.

Regression trees are nonlinear regression models having samples partitioned at each node of a binary tree based on the value of one selected input feature. The bootstrap sampling for each regression tree generation and the random selection of features considered for partitioning at each node reduces the correlation between the generated regression trees and, thus, the averaging of their prediction responses is expected to reduce the variance of the error.

Due to the feature selection algorithm and noise robustness, RF provides highly accurate predictions.

According to [26], the RF algorithm has several hyperparameters that have to be set by the user, as follows: number of variables for each split, number of observations drawn randomly for each tree, the splitting rule, minimum number of samples for a node and number of trees. Normally, RF work well with default values specified by different software packages, still, by tuning the hyperparameters it is possible to improve the performance of the random forest. Several reviews were published during the last decade on RF hyperparameter tuning [5]. Also, [26] emphasizes that the RF method is often presented as being insensitive to parameter choice when compared to other competing algorithms. The present study uses RF modelling both for building predictive models and to assess variable importance, especially for the cases where linear regression cannot be applied. Thus, by Python, the parameter importance through a RF approach is determined by 'feature importance' technique provided by scikit-learn's RF, based on gini impurity and by 'permutation importance' approach (as in the present research) in which variable importance is directly measured by observing the effect on model accuracy of randomly shuffling each predictor variable. Therefore, for measuring the performance of a RF model, the current study followed several steps, as follows: define and implement model evaluation function, base model accuracy vs real values, hyperparameter tuning – randomized search-grid search and RF model with fewer features.

*Dataset structure*

The dataset structure of present research includes a number of 18 relevant indicators [7] as follows : gross domestic product in rural areas (GDP) - (*thousand $*), gross value added (GVA) – [*billion $*], GVA for agriculture – [*billion $*], rural GVA - [*billion $*], direct payments (DP) – [*€/ha*], agricultural factor income (AFI) – [€/annual work unit], agriculture employment rate (AER) – [%], rural employment rate (RER) – [%], degree of rural poverty (DRP) – [%], agricultural entrepreneurial income (AEI) – [€/annual work unit], agriculture research and development investments (ARD) – [mil. €], labor productivity in agriculture (LPA) – [€/annual work unit], total factor productivity (TFP) – [€/annual work unit], cereal crop yield (CCY) – [kg], fertilizers use (FU) – [kg/ha], $CO_2$ emissions ($CO_2$) – [thousand tons] and ammonia emissions (EA) - [thousand tons], total factor productivity in agriculture – (TFP) [€/annual work unit].

## RESULTS AND DISCUSSIONS

According to other studies [10], in Europe, agriculture accounts for more than 47% of the total territory. In recent decades, European agriculture has experienced a continuous decrease in the number of farms, while the farm size shows a tendency towards larger holdings. According to the authors, a similar tendency was observed also for the share of agriculture in the gross value added (GVA), but with some differences between European regions. In present study, the overall Gross Value-Added parameter can be expressed through a multiple linear regression model with high accuracy (S value 0.20, adjusted R-sq 97.83%, predicted R-sq 97.67%) (see eq 2).

$$Ln(GVA) = -4.37 + 0.09\,Ln(CO_2) + 0.10Ln(rural\ GVA) + 0.31\,Ln(FU) - 0.35Ln(AER) + 0.67Ln\,(GVA\ for\ Agriculture) + 0.17Ln(DP) + 1.30Ln(RER) \tag{2}$$

According to the above model, the overall gross value added is positively influenced by several parameters. According to these, an increase in the rural employment rate, agricultural $CO_2$ emissions, agricultural gross value added, fertilizer use, or direct payments, would lead to an increase of the gross value-added indicator. The only negative influence comes from the 'Employment in Agriculture' parameter, that could be explained by a significant agricultural employee reconversion to services or industrial fields. This explanation is sustained also in other study [29], where the authors identifies that even if the number of agricultural co-operatives has been continuously decreasing, they continued to assure a high rate of job opportunities.

The importance of Agricultural Research and Development funding is emphasized in different research studies. For example, in [28], the authors describe the agricultural innovation platforms as promising vehicles to foster a paradigm shift in agricultural research for development. The digital farming, presented in [30], represents also a mandatory direction for the modern agricultural systems. Modern technologies like sensors, robotics, and data analysis helps moving from tedious operations to continuously automated processes. The authors are highlighting task planning algorithms, digitalization, sensors optimization, multi-robots, human-robot collaboration, environment reconstruction from aerial images and ground-based sensors for the creation of virtual farms as facing challenges in the context of digital farming.

The ARD parameter could not be modelled by using a linear approach, still the RF approach showed a good model accuracy (Fig. 1): BM accuracy at 86.01%, RS accuracy at 81.52%, GS accuracy at 82.18% and model validation at 70.18%. The feature selection analysis identified that, according to their weights, the following parameters were the most important for the ARD parameter prediction: 0.23 GVA, 0.16 DP, 0.15 GVA from agriculture, 0.06 $CO_2$ and 0.03 RER. Based on this numbers, it can be noticed that direct payments, total gross value added and agricultural gross value added are important parameters in ARD prediction.

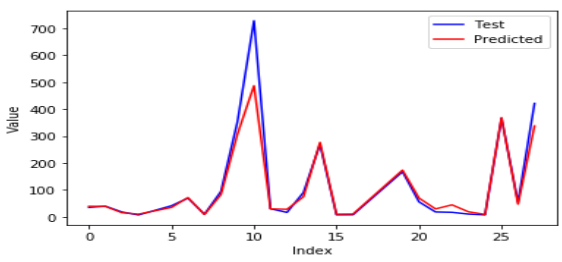


Fig. 1. Real values vs Predicted Values in Random Forest ARD model
Source: Original.

As such it seems, that direct payments overall managed to influence the dynamics of ARD investments. These can be correlated to information revealed by other study [1] which emphasize that ARD is critical in order to ensuring sufficient food for the world in the coming decades.

According to the European statistics, 95% of ammonia emissions come from agriculture. The National Emission Reductions Committee (NEC) criticizes the way the agricultural sector uses ammonia reduction methods and refers specifically to large industrial farms. Investments made in agriculture (through the direct payments / direct investments), along with the results of innovation research activities, can have a direct impact proportional to ammonia emissions (eq. 3), in the sense that the development of agricultural production takes place on a large scale and increases crop yields. On the contrary, the increase of the rural poverty degree, by the lack of jobs, of the presence of investors in the rural / agricultural space, the lack of crops and agricultural production and its effect on environmental, have an effect of reducing the ammonia emissions.

In [35] Velthof et al. (2012), the authors state that the agriculture is the major source of ammonia ($NH_3$), different methodologies being needed to quantify national $NH_3$ emissions and to identify the most effective options to mitigate $NH_3$ emissions. In Europe, according to [2], there has been little progress in controlling agricultural ammonia emissions and their share in European air pollution constantly increased, with values between 85 and 99% in countries with high agricultural activity, such as e.g. Denmark [11].

For the current research, after applying the best subset selection technique, the following linear model was identified for explaining the values of ammonia emissions in agriculture (EA):

$$Ln(EA) = 1.73 + 0.47\, Ln(Crop\ Yield) + 0.45 Ln(DP) - 0.59\, Ln(AFI) + 0.24 Ln(LPA) + 0.49 Ln\,(ARD) - 0.25 Ln(DRP) - 0.52 Ln(GDP) \qquad (3)$$

This model is described by excellent accuracy metrics, having the S value at 0.29, adjusted R-sq at 93.29% and predicted R-sq 92.17%

As it can be observed, the ammonia emissions from agriculture are positively related to cereal crop yield (0.47%), direct payments (0.45%), labor productivity in agriculture (0.24%) and agriculture research and development (0.49%) and negatively influenced by variables like the agricultural factor income per annual work unit in real terms (0.59%), the degree of rural poverty (0.25%) and the gross domestic product per capita in rural areas (0.52%). Direct payments can be used for upgrading fertilizers use technology, improving therefore the research and development activity in this direction and, thus, crop yield and labor productivity in agriculture. However, the increase of labor productivity may decrease the labor demand and accentuate the degree of rural poverty, situating which affects economic growth.

It should be mentioned that in other study [34], the authors are also investigating the relation between crop yield, agricultural factor income and greenhouse ammonia emissions from agriculture, identifying that the agricultural emission mitigation would be possible if a technological path is followed, this allowing also an increase of the farming income.

The $CO_2$ emissions in the agriculture were included in the current analysis as different studies showed that agricultural activities contribute significantly to the increase of $CO_2$ in the atmosphere. For example, [25] present as one of the main causes for $CO_2$ upward trend, the conversion of native ecosystems to agricultural uses. Their study shows that there are also major opportunities for $CO_2$ and other greenhouse gases mitigation through changes in the use and management of agricultural land. [23] emphasized on the fact that soil act as sources and sinks for greenhouse gases such as carbon dioxide ($CO_2$), methane ($CH_4$), and nitrous oxide ($N_2O$) and precise quantifications are needed to obtain reliable global budgets that are necessary for land-use management (agriculture, forestry), global change and for climate research.

Similar with the ammonia emissions from the agriculture activities, the $CO_2$ emissions are influenced by factors like the industrialization of the agricultural sector, agricultural research and development funding schemes, know-how improvements and rural area development. A higher volume of fertilizers could lead to a significant increase of the production volume, respectively production yield, but it will also increase the overall amount of emissions of the sector.

According to the present research dataset, the $CO_2$ emissions could be linearly modelled, eq. 4 expressing the most important predictors and their influence over $CO_2$ agricultural emissions.

$$Ln(CO_2) = -21.35 + 0.16\,Ln(DP) + 0.21Ln(ARD) + 2.82\,Ln(TFP) + 0.53Ln(DRP) + 0.46Ln\,(EA) + 2.49Ln(RER) + 0.30Ln(Fertilizer\ Use) \quad (4)$$

The above model has excellent accuracy metrics (S value: 0.33, adjusted R-sq: 90.45%, predicted R-sq: 89.37%).

Regarding the $CO_2$ emissions from the agriculture parameter, as it can be observed, it is positively influenced by the direct payments (0.16%), the agriculture research and development (0.21%), the total factor productivity in agriculture (2.82%), the degree of rural poverty (0.53%), the ammonia emissions from agriculture (0.46%), the rural employment rate (2.49%), and the fertilizers use in agriculture (0.30%).

As observed from the above coefficients, none of the parameters is negatively influencing the value of $CO_2$ value. That means an increase of any of the independent parameters will lead to a certain increase of $CO_2$ emissions.

From the beginnings of agriculture, fertilizers were used extensively by farmers and families to help different crops grow in different soil and weather conditions. Still, the fertilizer use is very expensive and can harm the environment if not used correctly. It is well known, as described by [15], that the excessive use of fertilizer leads to important environmental degradation and to a high health cost. The technological innovations are expected to improve fertilizers efficiency in crop production, still the socioeconomic constraints should be better understood. [15] emphasizes that the fertilizer use decreased with the increase of farm size and also that the crop yield was higher in large-scale farms. This situation is a consequence of the fact that a very low machinery level in small farms inhibited the application of precise fertilization.

In the current research, the fertilizer use modelling was performed by using a non-linear random forest model, as the linear modelling proved inconclusively, displaying low accuracy metrics.

Thus, the fertilizer use prediction model emphasized high accuracy metrics (Fig. 2), with the base model (BM) accuracy is at 86.71%, random search accuracy (RS) at 88.94%, grid search accuracy (GS) at 89.12% and model validation (MV) accuracy at 82.13%. According to the determined feature importance weights, the most important parameters predicting the fertilizers use are the following: crop yield (0.79), $CO_2$ (0.21), TFP (0.19), GVA (0.07) and DRP (0.05).

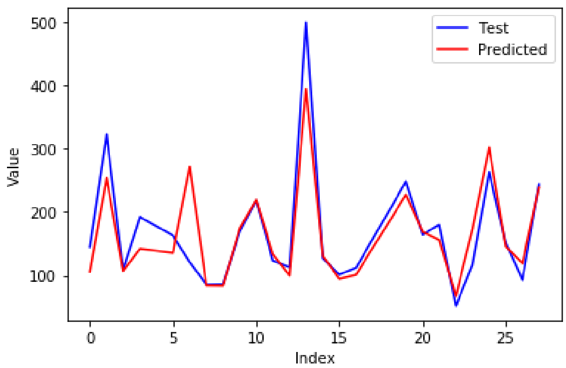


Fig. 2. Real values vs Predicted Values in Random Forest FU model
Source: Original.

As it can be observed, for the prediction of the fertilizers use value, the crop yield parameter displays the highest weight. The pesticides use – crop yield relation was to be expected as several studies are emphasizing on its importance. For example, in other paper [20], the authors emphasize that after the Green Revolution, the global crop production has doubled on the average, but crop yield has stagnated or even declined in some regions. In a context defined by an increased food demand, the use of pesticides and mineral fertilizers has improved crop yields but also contaminated food and the environment. According to their findings: fertilizer and pesticide consumption increased for 35–40 % of the countries; cereal production in 38 % of countries and yields in 47 % of countries either stagnated or decreased from 1961 to 2010; countries showing stagnated or decreased yields are countries with low gross domestic product per capita, mainly situated in Africa, South America, and West Asia – this would explain also the presence of the GVA parameter amongst the most important predictors for fertilizer use parameter. Another example can be found in the work of [22], where the authors are estimating the role of agronomic inputs in cereal yield improvements, their findings clearly suggesting the role for fertilizer, modern seeds and water in boosting yields.

Also, the model presented above confirms the relations previously identified in eq. 3, between fertilizers - crops and labor productivity, as well as $CO_2$ emissions and the degree of poverty.

## CONCLUSIONS

The dataset used in present research is able to generate prediction models, by using both multi-linear and random forest techniques, with excellent performance metrics, for the evaluation of agricultural sector sustainability. The environmental sustainability can be predicted through fertilizers use, $CO_2$ and ammonia emissions models. However, the economic sustainability is represented by the synergy of both GVA and ARD models. The results can be used for assessing insights over the dataset parameters, in terms of parameter relations and importance.

## ACKNOWLEDGEMENTS

The work of Simionov Ira-Adeline was supported by the project "ANTREPRENORDOC", Contract no. 36355/23.05.2019, financed by The Human

Capital Operational Programme 2014-2020 (POCU), Romania. This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS/CCCDI – UEFISCDI, project PN-III-P2-2.1-PTE-2019-0697, within PNCDI III. The authors are grateful for the technical support offered by ReForm - MoRAS through the Grant POSCCE ID 1815, cod SMIS 48745 (www.moras.ugal.ro).

## REFERENCES

[1]Alston, J.M., Beddow, J.M., Pardey, P.G., 2009, Agricultural research, productivity, and food prices in the long run. Science. Vol.325, Issue 5945, 1209-1210.

[2]Backes, A., Bieser, J., Aulinger, A., Matthias, V., Quante, M., 2016, Ammonia emissions in Europe, part I: Development of a dynamical ammonia emission inventory. Atmos Environ. 131, 55-66.

[3]Bajari, P., Nekipelov, D., Ryan, S.P., Yang, M., 2015, Machine learning methods for demand estimation. American Economic Review. Vol.105(5): 481-485.

[4]Burness, H.S., Brill, T.C., 2001, The role for policy in common pool groundwater use. Resour Energy Econ.Vol.23(1):19-40.

[5]Criminisi, A., Shotton, J., Konukoglu, E., 2011, Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Foundations and Trends in Computer Graphics and Vision. Vol. 7(2–3): 81-227.

[6]Einav, L., Levin, J., 2014, Economics in the age of big data. Science. Vol. 346, Issue 6210, 124389.

[7]European Comission Report - Agriculture and Rural Development, 2017.

[8]Fenske, N., Kneib, T., Hothorn, T., 2011, Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. J Am Stat Assoc. Vol.106, Issue 498: 494-510.

[9]Fisher, A., Rudin, C., Dominici, F., 2018, All Models are Wrong but many are Useful: Variable Importance for Black-Box, Proprietary, or Misspecified Prediction Models, using Model Class Reliance. arXiv Prepr arXiv180101489.

[10]Giannakis, E., Bruggeman, A., 2015, The highly variable economic performance of European agriculture. Land use policy. Vol.45: 26-35.

[11]Hutchings, N.J., Sommer, S.G., Andersen, J.M., Asman, W.A.H., 2001, A detailed ammonia emission inventory for Denmark. Atmos Environ. 35:1959-1968.

[12]Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.-M., Gerber, J.S., Reddy, V.R., Kim, S.-H., 2016, Random forests for global and regional crop yield predictions. PLoS One. 11(6).

[13]John, Lu. ZQ, 2010, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. J Royal Stat Soc Series A of Statistics Soc., Vol.173(3):693-694.

[14]Jones, S., Johnstone, D., Wilson, R., 2017, Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks. J Bus Financ Account. Vol.44(1-2):3-34.

[15]Ju, X., Gu, B., Wu, Y., Galloway, J.N., 2016, Reducing China's fertilizer use by increasing farm size. Glob Environ Chang. Vol.41:26-32.

[16]Kamilaris, A., Prenafeta-Boldú, F.X., 2018, Deep learning in agriculture: A survey. Computers and Electronics in Agriculture. Vol.147:70-90.

[17]Kim, B., Khanna, R., Koyejo, O., 2016, Examples are not enough, learn to criticize! Criticism for interpretability.Advances in Neural Information Processing Systems 29.

[18]Lazíková, J., Bandlerová, A., Rumanovská, L., Takáč, I., Lazíková, Z., 2015, Crop diversity and common agricultural policy-The case of Slovakia. Sustain.Vol.11(5):1-14.

[19]Lessmann, S., Baesens, B., Seow, H.V., Thomas, L.C., 2015, Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. Eur J Oper Res. Vol.247(1):124-136.

[20]Liu, Y., Pan, X., Li, J. A., 2014, 1961–2010 record of fertilizer use, pesticide application and cereal yields: a review. Agronomy for Sustainable Development. 35:83-93.

[21]März, A., Klein, N., Kneib, T., Musshoff, O., 2016, Analysing farmland rental rates using Bayesian geoadditive quantile regression. Eur Rev Agric Econ.

[22]McArthur, J.W., McCord, G.C., 2017, Fertilizing growth: Agricultural inputs and their effects in economic development. J Dev Econ. Vol.127: 133-152.

[23]Oertel, C., Matschullat, J., Zurba, K., Zimmermann, F., Erasmi, S., 2016. Greenhouse gas emissions from soils—A review. Chemie der Erde.Vol.76(3):327-352.

[24]Overmars, K.P., Helming, J., van Zeijts, H., Jansson, T., Terluin, I., 2013, A modelling approach for the assessment of the effects of Common Agricultural Policy measures on farmland biodiversity in the EU27. J Environ Manage. Vol.126:132-141.

[25]Paustian, K., Cole, C.V., Sauerbeck, D., Sampson, N., 1998, CO2 mitigation by agriculture: An overview. Clim Change. Vol.40: 135-162.

[26]Probst, P., Wright, M.N., Boulesteix, A.L., 2019, Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. Vol.9(3).

[27]Schlenker, W., Roberts, M.J., 2009, Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. Proc Natl Acad Sci U S A. Vol.106(37): 15594-15598.

[28]Schut, M., Klerkx, L., Sartas, M., Lamers, D., Campbell, M.M.C., Ogbonna, I., et al., 2016, Innovation platforms: Experiences with their institutional embedding in agricultural research for development. Exp Agric. pp.1-22.

[29]Schwarcz, P., Mura, L., Pätoprstý, M., Schwarczová, L., 2012, The impact of employment in agriculture on overall employment and development: a case study of the district of Topol'čany, Slovakia. SEER. 457-482.

[30]Shamshiri, R.R,, Weltzien, C., Hameed, I.A., Yule, I.J., Grift, T.E., Balasundram, S.K., et al., 2018, Research and development in agricultural robotics: A perspective of digital farming. Int J Agric Biol Eng. Vol.11(4)1-14.

[31]Shimshack, J.P., Ward, M.B., Beatty, T.K.M., 2007, Mercury advisories: Information, education, and fish consumption. J Environ Econ Manage. Vol.53(2): 158-179.

[32]Signorino, C.S., Yilmaz, K., 2003, Strategic misspecification in regression models. In: American Journal of Political Science. Vol.47(3):551-566.

[33]Thomas, F., Midler, E., Lefebvre, M., Engel, S., 2019, Greening the common agricultural policy: a behavioural perspective and lab-in-the-field experiment in Germany. Eur Rev Agric Econ. Vol.46(3):367-392.

[34]Valin, H., Havlík, P., Mosnier, A., Herrero, M., Schmid, E., Obersteiner, M., 2013, Agricultural productivity and greenhouse gas emissions: Trade-offs or synergies between mitigation and food security? Environ Res Lett. Vol.8(3).

[35]Velthof, G.L., van Bruggen, C., Groenestein, C.M., de Haan, B.J., Hoogeveen, M.W., Huijsmans, J.F.M., 2012, A model for inventory of ammonia emissions from agriculture in the Netherlands. Atmos Environ. Vol.46:248-255.

[36]Xia, Y., Liu, C., Li,Y.Y., Liu, N., A boosted decision tree approach using Bayesian hyper-parameter optimization for cred scoring, Expert Systems with Applications, DOI: 10.1016/j.eswa.2017.02.017

[37]Zivin, J.G., Neidell, M., 2013, Environment, health, and human capital. Journal of Economic Literature. Vol.51(3):689-730.