Unemployment in Rural Europe: A Machine Learning Perspective

Mehmet Güney Celbiş[1,2]

Abstract

This paper aims to provide policy-relevant findings that can contribute to the resilience of rural regions by discovering the main individual-level factors related to unemployment in those areas through the use of a set of machine learning techniques. Unemployment status is predicted using tree-based classification models: namely, classification tree, bootstrap aggregation, random forest, gradient boosting, and stochastic gradient boosting. The results are further analyzed using inferential techniques such as SHAP value analysis. Results suggest that access to training programmes can mitigate the labor market inequalities caused by differences in education levels, gender, age, alongside with parental education levels. The results also show how such inequalities are even larger for various subgroups detected by the employed algorithms.

Introduction

Rural labor markets are perpetually in transformation, resulting from the incessant and rapid evolution of their industrial compositions (Lindsay et al., 2003). Consequently, the ways of approaching the issue of rural unemployment, and the remedies proposed, have been context-sensitive. There were times when unemployment in rural areas was attributed to drastic changes in seasonality patterns of agricultural activities arising from advances in technology and the availability of capital (Gash, 1935). More recent views on the other hand, have drawn attention to network effects, individual employability assets (e.g. basic and key skills, personal characteristics), use of ICT, business size, and accessibility constraints posed by the remoteness from

✉ Mehmet Güney Celbiş
  celbis@merit.unu.edu

1 Department of Economics, Yeditepe University, Istanbul, Turkey

2 Maastricht Economic and Social Research Institute on Innovation and Technology
  (UNU-MERIT), Maastricht, The Netherlands

urban markets (Lindsay et al., 2003; McQuaid & Lindsay, 2003; McQuaid et al., 2004; Halden et al., 2005; Jones, 2004). Further interest on disadvantages of rural populations was rekindled in the recent years due to political outcomes related to rural populism which is rooted in the urban-rural disparities and the marginalization of rural populations (Wuthnow, 2018; Bernard et al., 2019). However, not all rural areas are equally subject to the negative effects of the urban-rural divide. For instance, while advanced urbanization benefited rural labor markets that are adjacent to cities, remote rural populations were left behind (Bock, 2018; Bernard et al., 2019). The urban-rural divide has become even more relevant, particularly due to the pandemic conditions. For example, a recent study by Türk (2021) shows that the digital divide between the rural and urban areas has created unequal access to distance education opportunities during the COVID-19 pandemic.

Apart from leading to direct negative social effects, rural unemployment also triggers problems indirectly by pushing individuals from rural to urban areas (Zenou, 2011; Lyu et al., 2019). This phenomenon is particularly pronounced in the case of young individuals that face unemployment in rural locations (Cartmel and Furlong, 2000; Jones, 2004; van Twuijver et al., 2020). The attractiveness of urban areas are also a factor for the outmigration of women from rural areas, as the agglomeration of industry in large urban areas often creates favorable conditions for female employment (Özgüzel, 2020; Alonso-Villar & Del Río, 2008). Especially in the global North, such population movements are frequently related to the insufficiency of public and private services and may lead to depopulation and permanent unemployment (Kühn, 2014; Bock et al., 2016; Bock, 2018; van Twuijver et al., 2020). The downward spiral created by the loss of the young labor force often transforms rural locations into concentrations of older people, leading to further challenges in attracting specialist workers (Steiner et al., 2021) Emigration from rural areas, while draining out rural economies, may in turn lead to problems related to congestion in urban areas. On the other hand, while bringing about a multitude of social consequences, unemployment itself is subject to a wide array of socioeconomic factors that range from the macroeconomic scale to the personal level. Against the background of such a broad spectrum of influences, machine learning (ML) algorithms have emerged as techniques that can provide a deeper understanding of sophisticated relationships (Athey, 2018; Harding & Hersh, 2018; Mullainathan & Spiess, 2017; Varian, 2014).

Applications of ML models even on the general topic of unemployment are rare. Some examples are the studies by Xu et al. (2013), Cook and Hall (2017), and Kreiner and Duca (2019) who use techniques such as neural networks and support vector machines, among other methods, for predicting unemployment in the USA. A similar approach is taken by Katris (2019) who used ML methods to predict unemployment rates in a set of European countries, Gogas et al. (2021) who forecast the Euro area unemployment, and the study by Chakraborty et al. (2020) where a set of ML approaches is employed to predict the unemployment rates in seven countries. In an individual-level study, Montanez and Hurst (2020) used smart meter data to predict personal employment status for a set of individuals in Ireland. The present study on the other hand, uses five different ML applications for investigating the mechanisms behind rural unemployment. To be more precise, rural unemployment is predicted through the use of stochastic and non-stochastic gradient boosting

machines, random forest and bootstrap aggregated algorithms in addition to a simple classification tree that serves as a root model for them. The findings of these models are further analyzed using an adaptation of Shapley values from cooperative game theory, alongside with partial dependence and clustering techniques. The single tree, bagging, and random forest models can be considered as providing relatively more descriptive results in relation to the research question, while the interpretable techniques that we subsequently employ constitute the empirical side of this paper. The aforementioned algorithms allow us to profile the individuals in rural areas who are unemployed, through the algorithmic assessment and selection from a large set of predictors. Furthermore, this study introduces alternatives to traditional approaches (e.g. econometric methods) frequently used in regional economic research, and contributes to the presently scarce ML applications in the field. Through the use of these techniques, the study aims to explore the individual-level reasons behind rural unemployment, assess the size and direction of the effects related to unemployment, and identify specific subgroups of individuals that are particularly vulnerable to job loss. The study also aims to algorithmically search and discover additional patterns in the data that may be of interest in addressing challenges relating to unemployment in rural locations.

The rest of this study is structured as follows. "The Data" provides a description of the data. "Binary Recursive Partitioning" constructs a classification (decision) tree and discusses its findings. "Bagging Prediction and Random Forest" expands the analytic work to ensemble models and explains the results of a bagging prediction together with the findings of a random forest implementation. "Non-stochastic and Stochastic Gradient Boosting Machine Applications" elaborates on the results of additive sequential tree models; that is to say, the gradient boosting algorithms (both stochastic and non-stochastic). "Interpretable Machine Learning Applications on the Gradient Boosting Results" implements a series of interpretable ML techniques on the results for the purpose of providing more detailed observation-specific results. "Concluding Remarks" discusses the conclusions of the study.

The Data

Measuring overall unemployment in rural locations – in comparison to urban areas – is relatively problematic, as major rural occupations are often not entitled to unemployment compensation claims which are often used to estimate official rates (Lasley & Korsching, 1984). The representativeness of rural data may be further hampered by other uncertainties that pertain to the measurement official unemployment rates, such as the discouraged worker effect. Furthermore, the commonly used designation "actively searching for a job" used when classifying the unemployed may result in loss information, as employment opportunities are often shared through informal networks (Lasley & Korsching, 1984). As a result, understanding the circumstances that affect rural unemployment in particular necessitates data at the individual level.

The European Social Survey (ESS) series directed by the European Research Infrastructure Consortium (ERIC) collects micro-level data on individual employment status among other themes. This study uses the eighth round of the ESS survey

program (ESS8, 2016). In ESS8, interviewees were asked about their employment status. Two of the options presented were related to being unemployed where the respondent could specify whether they are "unemployed and actively looking for job" or "unemployed, but not actively looking for job." In order to account for the discouraged workers, we merged these two options such that all individuals who have given any one of these two answers are categorized as unemployed. As a result, the dependent variable is constructed through a binary classification if individuals into the "Unemployed" and "Employed" categories. The data set has been subsequently limited to include only the individuals aged 15 to 64 (i.e. the working age population based on the OECD (2020) definition). Retired individuals are excluded from the dataset.

As a means to reduce heterogeneity in the sample that may exist due to significant dissimilarities of institutional structures unobserved in the data, the sample has been restricted to the EU countries that are included in ESS8, Norway, and the UK. Subsequently, the sample is further subsetted so that only the respondents located in rural areas are included.

The full ESS8 data set includes a multitude of variables (i.e. features) with comparatively very sparse observations, leading to costly limitations on the number of individuals that can be included in the analysis. A great number of survey questions, which are unfeasible to list within the text of this paper, can be grouped under general categories such as: political views, opinions regarding social freedoms, views on religion, experiences regarding discrimination, concerns regarding the environment and climate, alongside of the features selected by the models in the present study. Features with missing values greater than 10% have been excluded from the analysis, as collectively they rendered the dataset unusable, causing the prediction models to drop almost all rows. The ESS8 data also comprises many administrative features (e.g. interview duration, edition, production date, etc.), in addition to rows corresponding to refusals to respond. Such observations were also removed. Lastly, for each categorical predictor, all classifications were encoded as binary features.

The above outlined subsetting stages have reduced the full ESS8 data to a sample comprised of 4,622 observations and 958 features. However, as one would expect, only a small portion of these observations (about 8%) were in the "Unemployed" category. This large imbalance can cause the ML algorithms to be biased towards predicting many individuals as "Employed." This issue has been counteracted by including – instead of the total available number – a random sample consisting of 10% of the employed individuals in the data set, resulting in both a reduction in imbalance and an increase in computational efficiency.[1] The resulting sample consists of 762 observations and 958 predictors.

Despite the reduction in the full data, the ML models applied in this study analyse a large number of variables. We therefore define only the features that are chosen by the ML algorithms. The full ESS8 data set, accompanied by the code book for all variables is downloadable in the website of the ESS. For comparability, the sample

---

[1] Searching for the best tuning parameters to be used in the ML algorithms, through running grid searches using average home computers, can take several days.

used in our analysis is available for download.[2] In the following sections, a set of ML models are employed in order to extract information from the aforementioned dataset. In addition to models oriented towards prediction, we also apply model-agnostic interpretable ML techniques to further elaborate on the result, explore hidded effects, and identify disadvantages faced by subgroups of individuals that may be in need of specific policy approaches.

Binary Recursive Partitioning

With the aim of setting a point of departure for the ML models applied in the present study, we begin by fitting a classification tree through the application of a binary recursive partitioning algorithm on the training data, based upon the Classification and Regression Trees (CART) method by Breiman et al. (1984).[3,4] In the classification tree, the criterion of impurity for a node j is given by a Gini score $G_j = \sum_{k=1}^{K} p_{jk}(1 - p_{jk})$ and $p_{jk} = \frac{1}{N_j}\sum_{i \in D_j} \mathbf{1}(y_i = k)$ where y_i is the observed outcome for the i'th individual in the training data ($i = 1,...,N$), k is the class index, and D_j is the set of all individuals that are in the j'th node, and N_j is the number of observations in j (James et al., 2013; Friedman, 2001).[5] At each split, a splitting variable z_m from the feature space $Z = z_1,...,z_M$ and $m = (1,...,M)$ alongside with its split value c are chosen such that the sum of weighted Gini scores of the nodes produced by the binary partitioning are minimized:

$$\min_{m,c} \left[\frac{N_{j_1}(m,c)}{N} G_{j_1}(m,c) + \frac{N_{j_2}(m,c)}{N} G_{j_2}(m,c) \right] \tag{1}$$

where j_1 and j_2 are the two daughter nodes of j, and N is the number of observations (Friedman, 2001; James et al., 2013; Breiman et al., 1984). The partitioning described in Eq. 1 is not performed if the split does not yield data regions with a

---

[2] For replication purposes, the dataset and the code can be downloaded from the folowing links: https://drive.google.com/file/d/1G14AHu12c8ngLERsbyL7P-pBT7y18pxP/view?usp=sharing and https://drive.google.com/file/d/1TDTzHJCNeNtj4hyrcYf7u0LYqwhkT1Df/view?usp=sharing. The complete ESS8 data set and the code book including all feature definitions is available at: https://www.europeansocialsurvey.org/data/multilevel/guide/bulk.html.

[3] The training data corresponds to the randomly sampled 70% of the complete data, and the test data consists of the remaining thirty percent.

[4] The R functions used in this study are: randomForest by Liaw and Wiener (2002) for random forest and the associated proximity matrix, gbm by Ridgeway (2004) for the gradient boosting machine, Xgboost by Chen et al. (2015) for the stochastic gradient boosting machine, train by Kuhn (2008) for the bootstrap aggregated model, rpart by Atkinson and Therneau (2000) for binary recursive partitioning, rpart.plot by Milborrow (2019) for visualizing the classification tree, pdp by Greenwell (2017) for the individual conditional expectation graphs, shapforxgboost by Liu and Just (2020), and ggplot2 by Wickham (2011) for graphs.

[5] Since the dependent variable has a two-class structure, G_j is equal to $2p_{j1}(1 - p_{j1})$.

total Gini score that is lower that the Gini value for the observations in D_j (James et al., 2013; Friedman, 2001; Breiman et al., 1984).[6]

A full tree produced in this manner may overfit the data and lead to poor out-of sample predictions (James et al., 2013). To cope with this problem, the complexity of a classification tree is penalized through the application of a K-fold cross-validation process.[7] Based on the descriptions by Friedman (2001) and James et al. (2013) and Sutton (2005) this process can be summarized along the following lines: also referred to as "cost complexity pruning" or "weakest link pruning" the complexity reduction process is initiated by the generation of an unrestricted classification tree R and a series of nested subtrees $r \subseteq R$ where the complexity level of each r is represented by the parameter $\gamma \geq 0$. Greater values of γ indicate lower tree complexity such that in fact each subtree of R is in turn also nested in the immediate more complex subtree. Following the random partitioning of the training data into Q equally sized folds ($q = 1,...,Q$), the prediction is performed Q times by using each q internally as the validation set in each q'th round. The parameter of complexity γ corresponding to the lowest weighted aggregate misclassification error contingent on $\gamma|r|$ (the penalizing term) is determined for each r where $|r|$ is the number of the leaves of the classification tree (i.e. the number of terminal tree nodes $\bar{j}$):

$$E_q(\gamma) = \sum_{\bar{j}}^{|r|} e_{\bar{j}} \frac{N_{\bar{j}}}{N} + \gamma|r|$$
$$\text{where} \quad e_{\bar{j}} = \frac{1}{N_{\bar{j}}} \sum_{i \in D_j} \mathbf{1}(y_i \neq k_{\bar{j}}^*) \tag{2}$$

where $E_q(\gamma)$ expresses the error of subtree r in predicting the observations in fold q using all other $Q - 1$ folds adjusted by the complexity of r, $k_{\bar{j}}^*$ is the class that is in majority in the terminal node $\bar{j}$, $\mathbf{1}$ denotes the indicator function, and $\bar{j} = \{1, ..., |r|\}$. Next, the complexity-adjusted subtree errors are averaged for each level of γ over the Q predictions performed on the left-out folds. The γ value indexing the r'th subtree with the lowest complexity-adjusted error is:

$$\gamma^* = \arg\min_{\gamma} \left[\frac{1}{Q} \sum_{q=1}^{L} E_q(\gamma) \right] \tag{3}$$

Finally, from among the initially generated subtrees using the full training data, the subtree $r^* \subseteq R$ with a corresponding complexity value $\gamma = \gamma^*$ is used as the final tree model. (Friedman, 2001; Sutton, 2005; James et al., 2013).[8] In addition, the maximum depth of a node, and the minimum size of a node are determined through a grid search process.[9] As a consequence, a tree with a maximum node depth and a

---

[6] If z is a categorical feature, c separates the variable into classes or sets of classes.

[7] In order to avoid confusion with the class index which is also denoted by k, we index the number of data folds in the cross-validation step by q. The term "internal" is used to distinguish a given fold q from the test data which is left out earlier.

[8] It is possible that the cross-validation may indicate that the optimal tree is the unrestricted R (i.e. $\gamma|r|$ = 0 in Eq. 2). Therefore, the set of subtrees is expressed as "$r^* \subseteq R$" as opposed to "$r^* \subset R$" (James et al., 2013).

[9] The depth and size parameters were iterated by increments of 1 in a range of two to twenty.

minimum node size of 5, and a γ parameter of 0.02 is grown. The test data accuracy of the resulting classification tree is 61%.

The predictors selected by the binary recursive partitioning algorithm are defined in Table 1, and the classification tree is shown in Fig. 1. For each terminal node, the share of the predicted category is displayed. The single tree is dominated by predictors related to education. Individuals in rural areas who have taken steps to improve their skills and knowledge – such as by attending courses and training programs (represented by the variable `atncrse`) – are predicted to be employed. Alongside attendance to such activities, results imply that the skills and knowledge needed for employment may be transferred through personal relationships; an individual is predicted to be employed if the level of education of her/his father (`eiscedf`) is at least "upper secondary." This is not an implausible finding; Lindsay et al. (2003) note in their interview-based study that the work record of an individual's father – which may also be linked to his level of education – can be a very important determinant of employment status in remote rural labor markets. For the group of individuals whose fathers are less educated (i.e. `eiscedf<3`) only those with minimum nine years of education who hold at least an advanced vocational degree are predicted as employed.

While the above initial findings related to training and education may seem trivial, they are particularly relevant in the rural context. Chandler (1989) and Cartmel and Furlong (2000) had highlighted the insufficiency of training facilities, specifically for young individuals, in selected rural locations in England and Scotland respectively. Bock (2004) on the other hand, shows how women in particular are affected by the inadequacy of training opportunities in rural Europe. The ML techniques presented in the following sections, particularly the interpretable ML methods, discover and assess the effect directions and sizes of a larger set of features and examine the interactions between them.

Bagging Prediction and Random Forest

The sole tree presented in Fig. 1 is merely one possible representation of the multitude of classification trees used either in an ensemble or a sequential manner by the more sophisticated ML algorithms employed in the present study. This process, however, can yield very dissimilar trees when the underlying data set is modified (e.g. by resampling the training and test samples), and leave out important predictors if there is a high degree of correlation among the features (Athey and Imbens, 2019; James et al., 2013). In a bagging prediction, developed by Breiman (1996) and also called the bootstrap aggregation algorithm, multiple classification trees are combined. This is done by drawing, with replacement, a large number of random samples from the training data and constructing a corresponding classification tree t for each of those T samples $(1,...,T)$. The bagging prediction for individual i is the majority vote of the ensemble made of T classification trees. The optimum number of trees (T) is determined by using the estimate for the classification test error, calculated through the prediction of out-of- bag (OOB) observations corresponding to each t (James et al., 2013). We have identified the number of trees that yield the

Table 1 Algorithmically Selected Predictors - Classification Tree (out of 958 Features)

Name	Description	Values
atncrse	Binary variable indicating whether the respondent has taken any course or attended any conference or lecture to improve her/his knowledge or skills for work.	**1:** Yes. **2:** No.
eduyrs	Years of full-time education completed by the respondent.	Numeric predictor.
eisced	A 1 to 7 scale measuring the highest education level attained by the survey participant.	**1** indicates "less than lower secondary" and **7** indicates "Master's level or more".
eiscedf	A 1 to 7 scale measuring the highest education level attained by the survey participant's father.	**1** indicates "less than lower secondary" and **7** indicates "Master's level or more".

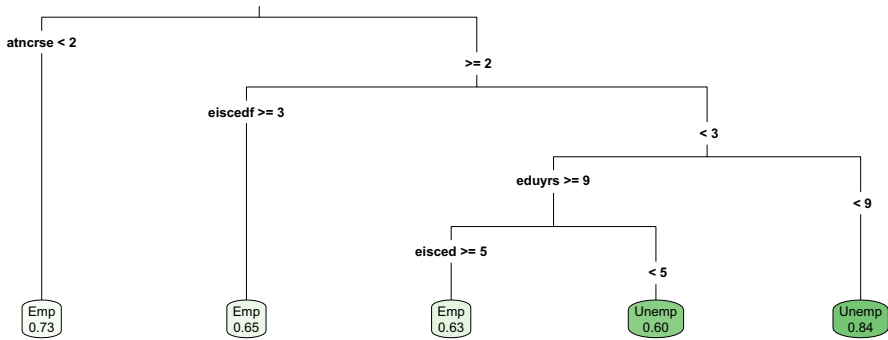Some of the above definitions may be identical to those in the ESS8 codebook (ESS8, 2016)

Fig. 1 Individual Classification Tree

lowest classification test error as 16 after making bagging predictions that consist of minimum 10, and maximum 500 classification trees. The accuracy of this bagged model is 63.5%.

The bootstrap aggregation method can be also used to calculate the variable importance values. The overall decrease in node impurity resulting from the usage of a given predictor as the splitting feature, presented in Eq. 4, is averaged over the T trees and scaled into an interval ranging from 1 to 100 (James et al., 2013):

$$\Delta G = G_j - \left[\frac{N_{j1}}{N} G_{j1} + \frac{N_{j2}}{N} G_{j2} \right] \tag{4}$$

where G_j is the Gini value of the parent node. The most important 20 predictors identified by the bootstrap aggregation algorithm and their definitions are presented in Fig. 2a and Table 2 respectively (the variables already defined in "Binary Recursive Partitioning" are not included). The findings of the bagged prediction is discussed in combination with the results of the below outlined random forest application.

The random forest algorithm developed by Breiman (2001) instills extra randomness into each tree – in addition to the random sampling employed by bagging – for the purpose of reducing the correlation between the trees in the ensemble (James et al., 2013; Friedman, 2001). When the splitting of a node is attempted in a random forest tree, the set of variables considered is restricted to a random subset $L \subset Z$, and $|L|$ is generally taken as $\sqrt{|Z|}$ in the classification context (Friedman, 2001; Breiman, 2001). Our random forest algorithm (with 500 trees) performs with an accuracy level of 65%. The variable importance degrees are presented next to those of the bagging prediction in Fig. 2b. While random forest and bagged predictions confirm the importance of training and education related factors, both models highlight the role of age in predicting rural unemployment. For the bootstrap aggregation algorithm, the age of an individual has been the predictor that provides the highest reduction in node impurity. The random forest application on the other hand, also reports age as one of the top predictors. In general, the two ensemble learning methods provide further information on several other features that were not selected by the classification tree. For instance, a person's opinions on the importance of

Table 2 Algorithmically Selected Predictors - Random Forest and Bagging (out of 958 Features, Cont'd)

Name	Description	Values
agea	Age of the participant.	Numeric predictor.
chldhm	Categorical feature indicating whether there are children living in the respondent's household.	**1**: Yes. **0**: No.
cntry.	The participant's country of residence (followed by the country abbreviation). Binary variable.	Equals 1 if the participant resides in the country and 0 otherwise.
consbuild	Binary variable indicating whether the respondent's current or past main job or the kind of work she/he did or does most of the time is defined as "construction of buildings."	**1**: Yes. **0**: No.
crophunt	Binary variable indicating whether the sector of the respondent's current or past main job is defined as "crop and animal production hunting and related service activities."	**1**: Yes. **0**: No.
domicil	Binary variable categorizing the type of the rural area where the respondent lives in.	Types are: "Country village," "Home or farm in countryside."
eduunmp	Score indicating the how strongly the respondent is in favor of the government spending more on training and education programs for unemployed individuals at the cost of a reduction in unemployment benefits.	A scale of 1 to 4 where **1** is "strongly against" and **4** is "strongly in favor."
eiscedm	A 1 to 7 scale measuring the highest education level attained by the survey participant's mother.	**1** indicates "less than lower secondary" and **7** indicates "Master's level or more".
gndr	Gender of the participant.	Binary predictor.
hhmmb	Number of regular members of the household.	Numeric predictor.
hlthhmp	Score measuring how badly by the respondent's daily activities are hampered by health and/or disability related problems.	A scale of 1 to 3 where **1**: "Very much" and **3**: "Not at all."
impdiff	Willingness of the respondent to try different and new things in life.	A scale of 1 to 6 where **1** means strong willingness and **6** means strong unwillingness.
ipcrtiv	Score measuring the level of importance that the respondent places on being creative and seeking new ideas.	A scale of 1 to 6 where **1** signifies very high importance and **6** signifies very low importance.
lptransport	Binary variable indicating whether the sector of the respondent's current or past main job is defined as "land transport and transport via pipelines."	**1**: Yes. **0**: No.

Table 2 (continued)

Name	Description	Values
pfirm	Binary variable indicating whether the respondent has worked or is working in the private sector.	**1**: Yes. **0**: No.
sbeqsoc	Score measuring the extent of agreement by the respondent that the social services and benefits in the country lead to a higher degree of equality in the society.	A scale of 1 to 5 where **1**: "Agree strongly" and **5** "Disagree strongly."
servbuild	Binary variable indicating whether the respondent's current or past main job or the kind of work she/he did or does most of the time is defined as services to buildings and landscape activities.	**1**: Yes. **0**: No.
uemplip	Binary variable indicating whether the respondent's partner has been unemployed and not looking for a job during the last seven days.	**1**: Yes. **2**: No.
wrkprbf	Score indicating the how strongly the respondent is in favor of the government introducing extra services and social benefits directed towards working parents to combine family and work life at the cost of much higher taxes for all.	A scale of 1 to 4 where **1** is "strongly against" and **4** is "strongly in favor."

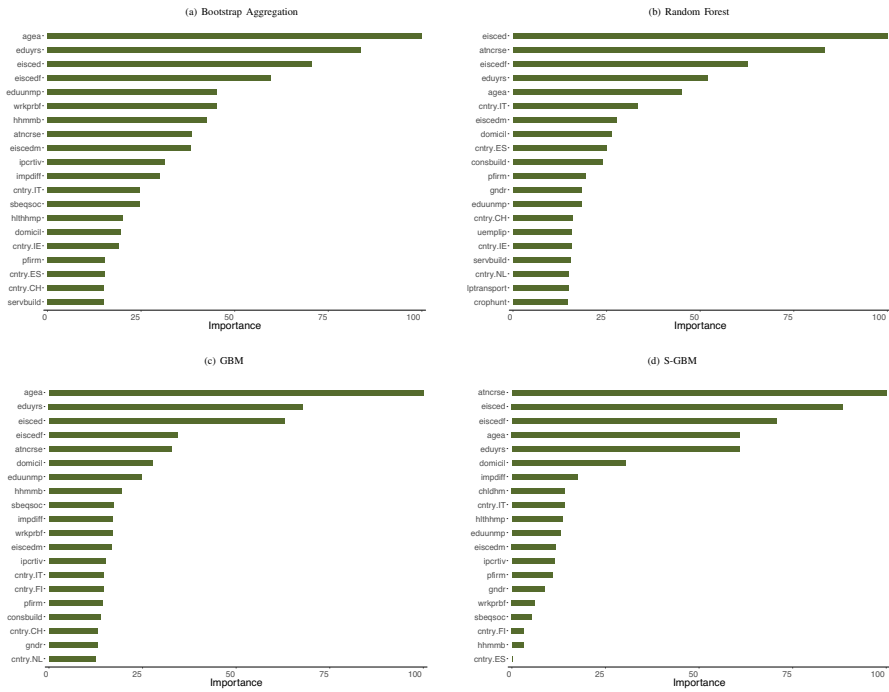Some of the above definitions may be identical to those in the ESS8 codebook (ESS8, 2016)

Fig. 2 Variable Importance Plots

being creative, the number of people in the household, and being residents of Italy or Spain are also important predictors. Regarding the findings on Italy and Spain, partial dependence plots suggest that individuals in the rural areas of these countries are more likely to be unemployed.[10]

Further information can be harnessed from the random forest results by generating a matrix of proximities between observations and plotting these proximities in reduced dimensional forms (Friedman, 2001; Breiman and Cutler, 2020). The employment status of the individuals that are excluded from a given classification tree t within the random forest (i.e. those that are the OOB observations for tree t as outlined earlier) are predicted by t, and each time in an iteration t the predicted outcomes of two OOB observations i and l ($l = 1,...N$, $i \neq l$)[11] fall into the same terminal region $D_{\bar{l},t}$, the element a_{il} of the initially zero $N \times N$ matrix **A** is increased by 1 (Friedman, 2001; Breiman & Cutler, 2020; Aldrich & Auret, 2013). **A** is subsequently divided by the number of trees in the random forest |T| and expressed in terms of dissimilarities in the form of a matrix **P**= 1 −**A** (Aldrich & Auret, 2013).

---

[10] The partial dependence plots for these variables are not reported in this study and can be provided if requested.

[11] The second observation is indexed by l instead of the convenient j for avoiding confusion with the tree nodes j.

(a) Two Dimensions

(b) Three Dimensions

Fig. 3 Proximity Plots: Random Forest

The visualization of the information contained in **P** in lower dimensions (typically three or two) is done through metric multidimensional scaling (Friedman, 2001). The resulting "random forest proximity plots" are presented in Fig. 3a and b in two and three dimensions respectively. In the proximity plots, the darker circles indicate the unemployed individuals vis-à-vis the lighter colored which indicate the employed persons. The larger a circle, the older is the corresponding individual.

Visualizing proximities between observations as outlined above also allows for the detection of clusters (Friedman, 2001; Aldrich & Auret, 2013). Both proximity plots suggest that the random forest algorithm has been able to separate the two categories to a certain extent; clusters of unemployed individuals are hinted on the left-hand-sides of both plots.

Non-stochastic and Stochastic Gradient Boosting Machine Applications

The ensemble methods presented in the previous section uncovered useful information on rural unemployment using information from large collections of separately grown trees. Sequential models on the other hand, allow the classification trees to learn from past prediction errors. Two sequential models that we apply in this study are the gradient boosting machine (GBM) developed by Friedman et al. (2001), and its extension, the stochastic gradient boosting machine (GBMS), established by Friedman (2002). The gradient boosting algorithms differ from the random forest and bootstrap aggregation models by growing the trees sequentially and attributing greater weights on individuals whose employment status were previously misclassified. Following Friedman (2001) and Friedman (2002) and Friedman et al. (2001), we construct a sequence of gradient boosted trees and compute the final predictions by implementing the following procedure: Firstly, the negative likelihood of the loss function $L(y_i, f(Z))$ is minimized with respect to the logarithm of odds of being unemployed $f(Z)$, and the resulting value is used as the approximation of the initial $\hat{y}_{i,t}$ for all i individuals collectively:

$$\hat{y}_{i,1} = \underset{f(Z)}{\arg\min} \sum_{i=1}^{N} L(y_i, f(Z)) \tag{5}$$

To allow a gradient boosted tree t belonging to sequence $1, ..., T$ to partially correct the errors made by its preceding counterpart $t - 1$, the errors of the latter is expressed in the form of pseudo-residuals ϵ_{it} for each person i in the training data:

$$\epsilon_{it} = -\left[\frac{\partial L(y_i, f(Z_i))}{\partial f(Z_i)} \right]_{f=f_{t-1}} for \ \ i = 1, ..., N \tag{6}$$

Subsequently, a regression tree is fitted to predict the pseudo-residuals.[12] The tree generates terminal nodes $\bar{j}$ comprised of the residuals of all i individuals $Z_i \in D_{\bar{j}}$. The output of a terminal node $\epsilon_{\bar{j}r}$ is given by Friedman (2001) and Friedman et al. (2001):

---

[12] As opposed to a classification tree, the binary criteria used in the recursive partitioning steps performed by a regression tree is often a loss function with a squared error form instead of node impurity. Regression trees, not covered in the present study, are formulated by Breiman et al. (1984) within the CART structure alongside with classification trees.

$$\epsilon_{\bar{j}r} = \arg\min_{\epsilon} \sum_{z_i \in D_{jr}} L(y_i, f_{t-1}(Z_i) + \epsilon) \tag{7}$$

Using the predicted errors of the previous tree, the subsequent tree then learns, to some degree, from the predictions f_{t-1} made in the previous iteration, and revises the predictions for individual i accordingly:

$$\hat{y}_{it} = f_{t-1}(Z_i) + \alpha\epsilon_{\bar{j}r}\mathbf{1}(Z_i \in A_{\bar{j}r}) \tag{8}$$

where the degree of learning, also referred to as the learning rate, is denoted by α, such that $\alpha\epsilon_{\bar{j}r}$ is the improvement made by tree t on the predictions of tree $t - 1$, and $\mathbf{1}$ is the characteristic function (Friedman et al., 2001; Friedman, 2001). The adjustment done by tree t is itself limited ($0 < \alpha < 1$) as improvements of a given tree may also be subject to errors. In fact, a low α combined with a large number of sequential trees is shown to improve predictions (Friedman, 2001; Friedman et al., 2001). We implement the GBMS model using the programming enhancements provided in the Extreme Gradient Boosting Machine algorithm (Xgboost) developed by Chen et al. (2015) which enables the user to introduce further regularization (such as pruning) of the classification trees as outlined in "Binary Recursive Partitioning". Xgboost also makes it possible for the GBM to incorporate additional stochasticity through the selection of a random sample when constructing each tree in addition to selecting a random subset of variables $X \subset Z$ at each binary split in each tree in the sequence, similar to the random forest approach.

In order to cope with potential overfitting caused by too complex trees, Friedman (2002) suggests the usage of a learning rate (i.e. shrinkage parameter) of 0.005, a random sample proportion of 50%, and maximum number of leaves (terminal nodes) of 6. In addition, we limit the predictors considered at each split to a subsample corresponding to 50% of the full set of features, set the minimum $N_{\bar{j}}$ to 5, and the maximum iteration number to 5000. The resulting GBM and GBMS algorithms are 66.2% and 61.4% accurate in predicting the employment status of the individuals in the test data. Figure 2c and d present the variable importance levels assessed by the two models. The sequential learning results are fairly consistent with those of the earlier applied ensemble tree algorithms. The variables pertaining to education related variables and age are selected as top predictors by both the GBMS and the GBM. However, the variable importance values presented in the previous sections provide information only on the importance of a feature for achieving the best possible predictions, and they do not provide information on how age, education, or any of the remaining features are related to employment status. In this regard, we elaborate on the relationship between a feature and the predicted employment status using partial dependences (Friedman, 2001) and Shapley value methods in the next section.

Interpretable Machine Learning Applications on the Gradient Boosting Results

Interpretable ML methods allow us to retrospectively understand how a variable is related to the outcome after the predictions are generated. We examine our results using partial dependences, conditional expectations, and Shapley values. The first of these approaches, the partial dependence plots, are constructed for typically two variables at most by grouping the features of interest Z_C and the remaining variables Z_S separately in the training data and computing:

$$f_C(Z_C) = \frac{1}{N} \sum_{i=1}^{N} f(Z_C, Z_{i,S}) \tag{9}$$

where the $Z_{i,S}$ are values of variables in Z_S that remain constant, and predictions $f_C(Z_C)$ are redone for each feature value in Z_C and averaged over individuals i (Friedman, 2001; Aldrich and Auret, 2013). While useful as a basic look at relationships between input variables and the outcome, this approach can result in implausible implications for features strongly correlated with variables in Z_S (Friedman, 2001; Molnar, 2019). An additional shortcoming of the partial dependence approach is that the partial effects are averaged, and therefore, individual patterns cannot be observed (Molnar, 2019). As formulated by Goldstein et al. (2015), the disaggregation of the global partial dependence for a feature and plotting the conditional expectations for each observation i yields Individual Conditional Expectation (ICE) plots where the heterogeneity across the individuals in the training data can be visualized (Goldstein et al., 2015; Molnar, 2019). However, stacked ICE curves for all individuals i in the training data will have separate intercepts. The expression can become visually more informative with regards to the heterogeneity of the effect of the variable of interest for different individuals upon centering all the ICE curves $f_{i,C}$ at a particular anchor point z^*, and therefore, removing the variation in levels caused by the different $Z_{i,S}$ values by plotting the centered individual conditional expectations $f_{i,C}^{centered}$ (Goldstein et al., 2015; Molnar, 2019):

$$f_{i,C}^{centered} = f_{i,C} - f(z^*, Z_{i,S}) \tag{10}$$

Figure 2 displays the centered individual conditional expectation (ICE) graph based on the GBMS model for the feature agea, alongside with the two-way partial dependence plot (PDP) for agea and eduyrs.[13] Figure 4a indicates that the probability of being unemployed is on average lower for middle-aged persons (shown by the red partial dependence curve), compared to the individuals aged between 16-22 which is within the range defined by the OECD (2020) for youth unemployment (15-24). The probability of unemployment further drops for the individuals between the ages of about 50 and 60. The black ICE curves on the other hand, imply considerable heterogeneity. This observation is an expected

---

[13] Partial dependence plots are developed by Friedman (2001), and the individual conditional expectation plot is developed by Goldstein et al. (2015).
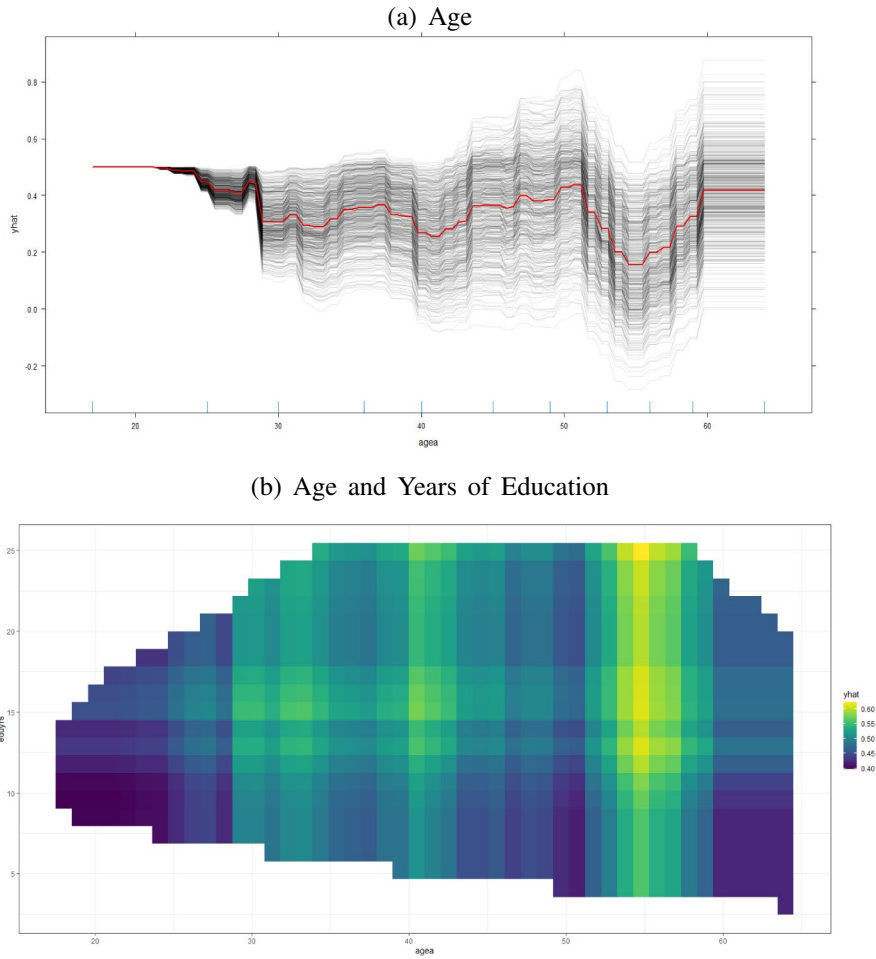
(a) Age

(b) Age and Years of Education

Fig. 4 Individual Conditional Expectation Plots: Stochastic Gradient Boosting

outcome; our ML models precisely address this situation among other issues, by allowing for all possible interactions in the data. The bright yellow pixels in the two-way PDP in Fig. 4b represent high probabilities of employment, and are concentrated mostly around the areas that correspond to middle-aged individuals (up to around age 60) and to durations of education that are, approximately, higher then 10 years.

The Shapley Additive Explanation (SHAP) method is a novel approach adapted into the machine learning framework by Lundberg and Lee (2017) based on the use of Shapley values from cooperative game theory developed by Shapley (1953). The SHAP approach is able to summarize both the sizes and the directions of the effects of each feature for each data instance. In this regard, the framework presented by Lundberg and Lee (2017) can be summarized as follows:

the contribution of the value v of a given variable z_m to the deviation of the predicted outcome from the average, for the i'th individual, is denoted by ϕ_v and computed as

$$\phi_v = \sum_{S \subseteq Z \setminus \{v\}} \frac{|S|!(|Z| - |S| - 1)!}{|Z|!} \big[g(S \cup \{v\}) - g(S) \big] \tag{11}$$

where S is a subset of variable values from the features for observation i which are to be held constant, while multiple GBM predictions for i are calculated by replacing the features except those in S by randomly drawn observations from the training data, and subsequently calculating the deviation of the expected value for individual i from the average prediction (Lundberg & Lee, 2017; Molnar, 2019).[14] More specifically, this step is performed by the function g as follows:

$$g(S) = \int f(Z) dP_{z \notin S} - E(f(Z)) \tag{12}$$

whereas in the term $g(S \cup \{v\})$ in Eq. 11 the actual feature value for observation i is preserved alongside with those in S, and the step is repeated for all possible combinations that can produce a subset $S \subseteq Z \setminus \{v\}$ and averaged over all the combinations (Molnar, 2019). As Štrumbelj and Kononenko (2013) and Molnar (2019) point out, calculating multiple predictions with v versus with a randomly drawn v for all combinations $\frac{|S|!(|Z| - |S| - 1)!}{|Z|!}$ is computationally very inefficient. An approximation of the SHAP value for individual i is formulated by Štrumbelj and Kononenko (2013) and is presented in Eq. 13:

$$\bar{\phi}_v = \frac{1}{B} \sum_{b=1}^{B} \big[f^*(\hat{y}(i)_{+v}^b) - f^*(\hat{y}(i)_{-v}^b) \big] \tag{13}$$

where the outcome for the i'th individual ($f^*(\hat{y}(i)_{+v}^b)$) is predicted by replacing a random subset of the variable values for this person while preserving its specific true feature value v (hence the subscript $+v$). The prediction is repeated, but this time the value of the variable of interest, v, is also replaced by the corresponding value of the randomly drawn observation l, and this predicted outcome $f^*(\hat{y}(i)_{-v}^b)$ is subtracted from the prediction that preserves the true v. The difference between the predicted outcomes for person i with and without v is computed in this randomized fashion B times ($b = 1,...,B$) and the average $\bar{\phi}_v$ is calculated (Molnar, 2019; Štrumbelj & Kononenko, 2013). All SHAP outcomes in the present study computed using the R package `SHAPforxgboost` (Liu & Just, 2020) which also calculates the global SHAP importance levels for every feature and observation by aggregating the absolute SHAP values for a given feature v for all i and the variables in Fig. 5 are ordered based on their SHAP importance levels computed as (Molnar, 2019):

---

[14] The cardinality of Z as denoted in Eq. 11 is an alternative way to express the total number of features M as defined in "Binary Recursive Partitioning" ($|Z| = M$).
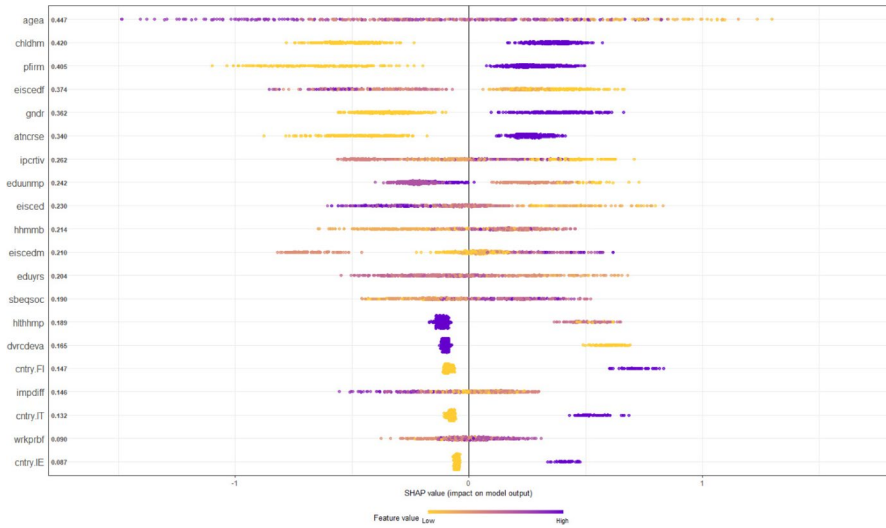
Fig. 5 SHAP Values - Extreme Gradient Boosting Machine

$$\Phi_v = \sum_{i=1}^{n} |\phi_{i,v}| \tag{14}$$

Figure 5 visualizes the resulting SHAP values of each feature and person. Each individual in the training dataset is represented by a color-filled circle where darker circles indicate a higher v (for binary values only two colors are used). The x-axis values indicate the SHAP values. As in all previous importance plots, we report the features with the highest importance scores. In addition to the features selected by the earlier presented models, the SHAP procedure identifies `dvrcdeva`, a binary feature indicating whether the respondent has ever divorced or had a civil union dissolved, as a relevant variable. Consistent with the earlier findings, the SHAP results suggest that age is a major determinant of rural unemployment. A cluster of light colored circles to the far right in the row for `agea` (the first row in Fig. 5) indicate that young age has contributed strongly to the prediction of uneemployment for those individuals. The SHAP plot also shows that individuals with children at home (`chldhm`) – represented by light colored circles – are less likely to be unemployed. However, it is possible that this feature captures the effect of age, as middle-age individuals whom we predict to be mostly employed are more likely to have children at home compared to younger or older persons. The findings also suggest that individuals who work in the private sector (those represented as dark circles in the `pfirm` row) are more vulnerable to job loss compared to people who are working in the private sector or are self-employed. We also observe that the highest level of education of a person's father (`eiscedf`) contributes to the prediction of lower-than-average unemployment probabilities in accordance with our previous results. The SHAP values also underline the disadvantages faced by women in rural labor markets, as gender has
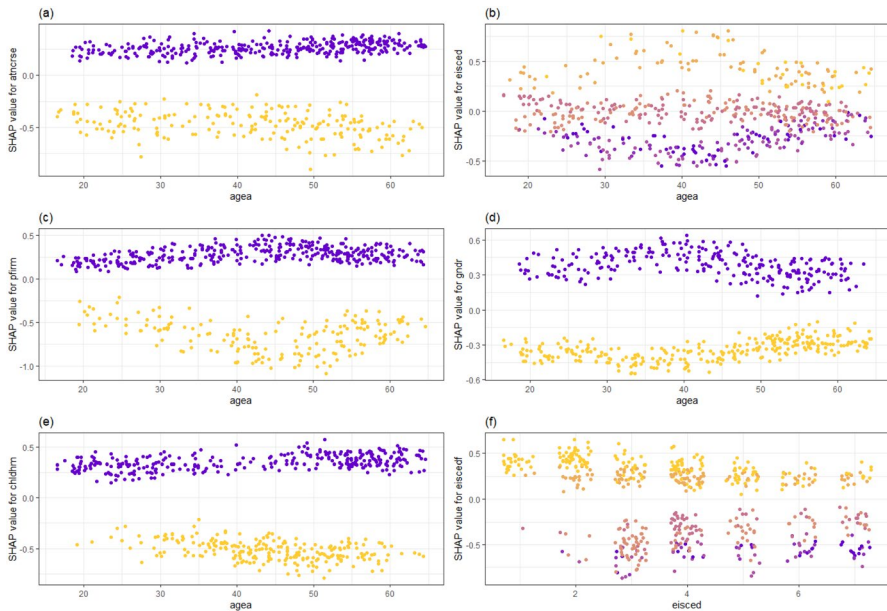
Fig. 6 SHAP Interaction Plots - Extreme Gradient Boosting Machine

contributed positively to the prediction of unemployment for female respondents. One of our main variables of interest that was highlighted by the previous models is `atncrse` which indicates whether the respondent took steps to improve her/his knowledge and skills by attending a course, lecture, or a conference. The SHAP values show that attending such activities lowered the probability of being unemployed (circles marked with light colors).

Instead of assessing the above mentioned major effects in isolation, we examine SHAP interaction plots in Fig. 6 (Molnar, 2019; Liu & Just, 2020).[15] In panel (a) of Fig. 6 the persons who have not attended a course, lecture, or a conference are represented as dark circles while those who did are marked in light colors. We observe that the SHAP value of this variable is positive if the person did not attend such an activity (adds to the prediction of a person's status as unemployed) and negative otherwise. A slight divergence between the two groups is suggested to exist as age increases, indicating that such additional/complementary training activities are more important for older individuals with regard to their employment probabilities. In panel (b), we see that high levels of education (dark colors) yield negative SHAP values, where low levels yield positive SHAP values (i.e. increase the prediction of unemployment). Intermediate levels of education, on the other hand, do not seem to have strong effects. The impact of the level of education depends on age, as the discrepancies are larger for middle aged individuals, whereas they are less pronounced for younger and older groups.

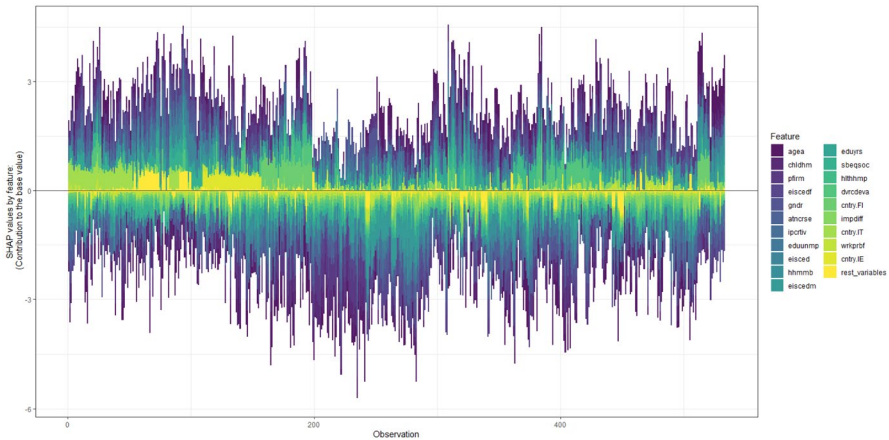[15] A low degree of jitter is used in the plots for better readability.

Fig. 7 SHAP Force Plot - Extreme Gradient Boosting Machine

In panel (c), persons who work in private firms are represented with dark colors, and those who work in the public sector, or are self-employed are shown in light colors. We observe that people who are employed in private firms are particularly vulnerable to unemployment, and that this vulnerability is a higher concern for middle-aged persons. We also observe a similar result for gender, as shown in panel (d) where dark colored circles represent females and light colors represent males. The plot suggests that being female makes a big difference as the SHAP values of women are all above zero. This difference begins to narrow to some extent for persons older than 50 years old. People without children are attributed higher SHAP values (panel (e)), but as earlier noted, this finding may also be capturing the effect of the variable age. Finally, panel (f) shows how a persons own education interacts with the father's education level which was observed to be an important feature yielding high SHAP values. Persons with low levels of father's education (light colors) have positive SHAP values, but this disadvantage in the labor market is mitigated to some extent as the person's own education level increases.

Finally, the SHAP force plot (Molnar, 2019; Liu & Just, 2020) presented in Fig. 7 shows that the top 20 variables with the highest SHAP importance levels account for most of the variation in the SHAP value calculations for each individual (observation) represented by a bar. The additive contribution to the individual specific SHAP value of all remaining features for a given person is grouped under the category "rest_variables" and shown in bright yellow. The force plot suggests that reporting the top 20 variables has been a reasonable approach in our models.

Concluding Remarks

As an issue that is hard to detect, poverty and unemployment in rural areas attracts relatively less political interest in Europe (Chandler, 1989; Bernard et al., 2019). This being said, unemployment in rural Europe has often been identified as a

persistent problem particularly for young individuals (Philip & Shucksmith, 2003; Unay-Gailhard, 2016).

This study provided new evidence with regard to the recent state of rural unemployment in Europe by using up-to-date statistical learning methods. In particular, the research question has focused on understanding *who* are the individuals facing difficulties in finding jobs in rural locations, and why they are unemployed while others are not. The results suggest that issues related to lack of adequate training and education opportunities, and youth unemployment in rural areas – highlighted over the decades in many studies – still persist. In addition to features related to age and education, the algorithms have selected other predictors related to a person's character and household attributes. Further, in depth empirical assessments of the findings showed that persons who come from lower educated families (as shown by the father's education variable), face difficulties in becoming competitive in the labor market, and even for those who have high levels of education themselves this disadvantage is not reduced sufficiently. This result may be related to personal networks and connections which may be weak for even educated persons in rural areas if their fathers, due to low education, were excluded from such circles. Therefore, policies that aim to increase the inclusiveness of children of disadvantaged families, by increasing their chances to match with available jobs in the rural labor markets may present benefits.

We have also observed that while attending complementary training activities are particularly helpful for reducing a person's probability to be unemployed, some evidence suggests that elderly people may benefit the most from such activities. Whereas the general level of education affects employment status, the effect is much more noticeable for persons roughly above the age 30 and below 50 years old. This implies that education does not necessarily provide a way out of unemployment for persons younger that 30 years old, and that the benefit of having a higher level of education is limited for younger persons. The employment discrepancies become particularly pronounced in the case of women, who face clear disadvantages. Establishing or enhancing job placement assistance programs for young persons and women in rural areas may potentially generate welfare-increasing outcomes. We also observe that job security is a concern particularly for persons working in private firms as they are more prone to become unemployed compared to other groups (this finding is especially relevant for middle-aged persons). Therefore, legal improvements aimed to increase job security levels in rural markets may provide further ameliorations to the living standards of the rural workforce. Finally, some country-level effects are found (such as for the cases of Italy and Spain) which call for thorough case studies on the rural labor markets in these countries.

References

Aldrich, C., & Auret, L. (2013). *Unsupervised process monitoring and fault diagnosis with machine learning methods*. London: Springer.

Alonso-Villar, O., & Del Río, C. (2008). Geographical concentration of unemployment: A male–female comparison in Spain. *Regional Studies*, *42*(3), 401–412.

Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda* (pp. 507–547). University of Chicago Press.

Athey, S., & Imbens, G.W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics, 11*.

Atkinson, E.J., & Therneau, T.M. (2000). *An introduction to recursive partitioning using the RPART routines*. Rochester: Mayo Foundation.

Bernard, J., Contzen, S., Decker, A., & Shucksmith, M. (2019). Poverty and social exclusion in diversified rural contexts. Sociologia Ruralis.

Bock, B. (2004). *It still matters where you live: rural women's employment throughout Europe*, (pp. 14–41). Farnham: Ashgate Publishing Limited.

Bock, B., Osti, G., & Ventura, F. (2016). Rural migration and new patterns of exclusion and integration in Europe. *Routledge International Handbook of Rural Studies 71–84*.

Bock, B.B. (2018). Rural futures: Inclusive rural development in times of urbanisation. Technical report.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24* (2), 123–140.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Breiman, L., & Cutler, A. (2020). Random Forests. https://www.stat.berkeley.edu/breiman/RandomForests/cc_home.htm. Accessed 23 Feb 2020.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). *Classification and regression trees*. Monterey: Wadsworth and Brooks.

Cartmel, F., & Furlong, A. (2000). Youth unemployment in rural areas. Number 18. York Publishing Services for the Joseph Rowntree Foundation York.

Chakraborty, T., Chakraborty, A.K., Biswas, M., Banerjee, S., & Bhattacharya, S. (2020). Unemployment rate forecasting: A hybrid approach. *Computational Economics*, *57*(1), 183–201.

Chandler, J. (1989). Youth unemployment in rural areas: Local government and training agencies. *Local Government Studies*, *15*(3), 59–73.

Chen, T., He, T., Benesty, M., Khotilovich, V., & Tang, Y. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2 1–4.

Cook, T., & Hall, A.S. (2017). Macroeconomic indicator forecasting with deep neural networks. Federal Reserve Bank of Kansas City, Research Working Paper, (17-11).

European Social Survey (ESS ERIC). (2016). European social survey (ESS) round 8.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* Vol. 1. New York: Springer Series in Statistics.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, *5*, 1189–1232.

Friedman, J.H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378.

Gash, N. (1935). Rural unemployment, 1815-34. *The Economic History Review*, *6*(1), 90.

Gogas, P., Papadimitriou, T., & Sofianos, E. (2021). Forecasting unemployment in the Euro area with machine learning. *Journal of Forecasting*, *41*(3), 551–566.

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, *24*(1), 44–65.

Greenwell, B.M. (2017). Pdp: An R package for constructing partial dependence plots. *The R Journal*, *9*(1), 421–436.

Halden, D., Mcquaid, R., & Greig, M. (2005). Relationships between transport and the rural economies.

Harding, M., & Hersh, J. (2018). Big data in economics. IZA World of Labor, (451).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* Vol. 112. New York: Springer.

Jones, M.K. (2004). Rural labour markets: The Welsh example. *Local Economy: The Journal of the Local Economy Policy Unit*, *19*(3), 226–248.

Katris, C. (2019). Prediction of unemployment rates with time series and machine learning techniques. *Computational Economics*, *55*(2), 673–706.

Kreiner, A., & Duca, J.V. (2019). Can machine learning on economic data better forecast the unemployment rate? *Applied Economics Letters*, 1–4.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26.

Kühn, M. (2014). Peripheralization: Theoretical concepts explaining socio-spatial inequalities. *European Planning Studies*, *23*(2), 367–378.

Lasley, P., & Korsching, P.F. (1984). Examining rural unemployment. *Journal of Extension*, *22*(5), 32–36.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomforest. *R News*, *2*(3), 18–22.

Lindsay, C., McCracken, M., & McQuaid, R.W. (2003). Unemployment duration and employability in remote rural labour markets. *Journal of Rural Studies*, *19*(2), 187–200.

Liu, Y., & Just, A. (2020). SHAPForxgboost: SHAP Plots for 'XGBoost'. R package version 0.1.0.

Lundberg, S.M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 4768–4777).

Lyu, H., Dong, Z., Roobavannan, M., Kandasamy, J., & Pande, S. (2019). Rural unemployment pushes migrants to urban areas in Jiangsu province, China. *Palgrave Communications, 5*(1).

McQuaid, R., Lindsay, C., & Greig, M. (2004). Reconnecting the unemployed information and communication technology and services for jobseekers in rural areas. *Information, Communication & Society*, *7*(3), 364–388.

McQuaid, R.W., & Lindsay, C. (2003). Delivering job search services for unemployed people in rural areas. The role of ICT.

Milborrow, S. (2019). R package 'rpart. plot'.

Molnar, C. (2019). Interpretable Machine Learning. https://christophm.github.io/interpretable-ml-book/. Accessed 12 March 2020.

Montanez, C.A.C., & Hurst, W. (2020). A machine learning approach for detecting unemployment using the smart metering infrastructure. *IEEE Access*, *8*, 22525–22536.

Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106.

OECD. (2020). Working age population (indicator). https://data.oecd.org/pop/working-age-population.htm. Accessed 4 April 2020.

Özgüzel, C. (2020). Agglomeration Effects in a Developing Country: Evidence from Turkey. Mimeo, Paris School of Economics halshs-02878368.

Philip, L.J., & Shucksmith, M. (2003). Conceptualizing social exclusion in rural Britain. *European Planning Studies*, *11*(4), 461–480.

Ridgeway, G. (2004). The gbm package. R Foundation for Statistical Computing, Vienna. Austria, 5(3).

Shapley, L.S. (1953). A value for n-person games. In K. H.W. T. A.W. (Eds.) *Contributions to the Theory of Games II, Annals of Mathematics Studies No. 28* (pp. 307–317). Princeton: Princeton University Press.

Steiner, A., Caló, F., & Shucksmith, M. (2021). Rurality and social innovation processes and outcomes. A realist evaluation of rural social enterprise activities. Journal of Rural Studies.

Štrumbelj, E., & Kononenko, I. (2013). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, *41*(3), 647–665.

Sutton, C.D. (2005). Classification and regression trees, bagging, and boosting. In *Handbook of Statistics* (pp. 303–329). Elsevier.

Türk, U. (2021). A multilevel analysis of the contextual effects in distance education outcomes during COVID-19. *Eastern Journal of European Studies*, *12*, 149–169.

Unay-Gailhard, İ. (2016). Job access after leaving education: a comparative analysis of young women and men in rural Germany. *Journal of Youth Studies*, *19*(10), 1355–1381.

van Twuijver, M.W., Olmedo, L., O'Shaughnessy, M., & Hennessy, T. (2020). Rural social enterprises in Europe: A systematic literature review. *Local Economy: The Journal of the Local Economy Policy Unit*, *35*(2), 121–142.

Varian, H.R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28.

Wickham, H. (2011). Ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(2), 180–185.

Wuthnow, R. (2018). *The Left Behind*. Princeton: Princeton University Press.

Xu, W., Li, Z., Cheng, C., & Zheng, T. (2013). *Data mining for unemployment rate prediction using search engine query data* (Vol. 7, pp. 33–42). Berlin: Springer.

Zenou, Y. (2011). Rural–urban migration and unemployment: theory and policy implications. *Journal of Regional Science*, *51*(1), 65–82.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.