Extended Abstract
Please do not add your name or affiliation

Paper/Poster Title	**The determining factors of farmers' participation in insurance schemes: a comparative analysis of machine learning tools.**

Abstract prepared for presentation at the 97th Annual Conference of the Agricultural Economics Society, The University of Warwick, United Kingdom

27th – 29th March 2023

Abstract	***200 words max***

Identifying the factors influencing insurance participation plays a vital role in ensuring the success of such schemes. Many scholars have tried to unravel the question with the most varied approaches. We address a gap in the literature by Using Machine Learning. The analysis relied on 10,926 individual fieldcrop farm observations from the Italian Farm Accountancy Data Network from 2016 to 2019.
We use three Machine Learning (ML) tools: Least Absolute Shrinkage and Selection Operator (LASSO), Boosting and Random Forest, and compare them to the Generalised Linear Model (GLM) usually used in insurance modelling and adopted in this study as the baseline. In addition, we have adopted a cross-validation process to determine the reliability of the findings. The results revealed the outstanding performance of ML: these tools efficiently perform variable selection with an accurate prediction of insurance compliance. In particular, Boosting performs better than other tools, especially in comparison with GLM, using a smaller set of regressors. At the same time, the reliability of variable frequency selection is found to be high. Results can be used to focus on the subset of information that best explains insurance participation and, in turns, to reduce the cost of designing insurance schemes.

Keywords	Machine Learning, Insurance, Risks
JEL Code	G22, C55, Q18, Q12, G22 see: www.aeaweb.org/jel/guide/jel.php?class=Q)

Introduction	***100 – 250 words***

External shocks, such as extreme events in weather conditions, markets or policy, significantly impact agriculture. Farmers use a variety of risk management tools to face these risks, where insurance takes the lion's share. The agricultural insurance literature has analysed several aspects concerning the relationship between farmers and insurance. One goal of the analysis is to assess which variables affect adhesion in insurance schemes. This allows the design of the insurance contracts structure to meet special insurance requirements based on the unique features of individual farms. Hence, this can support insurance companies and policymakers in creating contracts that satisfy farmers' needs and make the scheme financially sustainable. This study analyses the (many) characteristics that potentially influence farmers' behaviour if we consider the involvement in an insurance scheme using different Machine Learning tools. Participation choice is usually affected by a large number of variables, making the task challenging. Furthermore, these factors are interrelated and can easily mask the influence on any prediction of adhesions.

Performing an accurate prediction and identifying the factors that affect farmers' participation are the main objectives of this analysis. Unfortunately, traditional methodologies (GLM) cannot satisfactorily use this large set of variables because of problems such as multicollinearity and overfitting. ML tools could overcome these problems.

Methodology	***100 – 250 words***

Click here to enter text.Relyed on the literature, we investigate 66 farm characteristics, including economic, technical, financial, topographic and climatic features affecting farmers' participation in insurance schemes. The analysis utilises individual data from the Italian FADN from 2016 to 2019. We restrict the analysis to fieldcrop farms (type of farming 1) in order to analyse subjects with comparable risk exposure. This selection yields 10,926 observations. We employ three Machine Learning (ML) tools: LASSO, Boosting and Random Forest. These ML tools were chosen because they are able to select variables. The performance of ML is compared to that of GLM, a model usually used in the insurance sector. Furthermore, to evaluate the robustness of the results, we implemented a cross-validation (CV) procedure. This technique divides the entire sample into two sets: the training-set to find the model setting and the test-set used to predict the outcome by adopting the setting found previously. This procedure is performed several times. We consider two aspects to assess the variable selection capabilities of models: the frequency of selection and the relative significance of each variable. In addition, the model performance also is also assessed by analysing the confusion matrices, mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE).

Results	***100 – 250 words***

Boosting reaches high performance using only 41 variables of the 66 available. In contrast, the other tools are less selective. The confusion matrix reports that ML overcome GLM considering sensitivity, negative prediction value, detection rate and balanced accuracy. In particular, Boosting obtained the best performance, followed by Random Forest and Lasso. MAE, MSE and RMSE confirm the outstanding predictive capacity of the ML Tools. However, these differ in this regard Boosting performs the best, GLM and LASSO are the worst, while Random Forest stays in between. Results allow us to identify which variables are selected the most and are the most important in explaining farmers' participation. The are in order of importance: farm economic size, presence of other gainful activities, amount of land, kW of available machinery, production diversification (Herfindahl index), degree of intensification (as total revenue per unit of land), fixed capital on total capital, and mechanical expenses.

Discussion and Conclusion	***100 – 250 words***

Although participation in an insurance scheme is a complex decision, ML ensures a relatively good prediction way better than GLM models. Furthermore, Boosting offers better performances within the considered ML approaches by utilising a small set of regressors. Conversely, the settings of Boosting can be challenging, and the evaluation of trade-offs between selection capacity and performance must be considered. Nevertheless, the proposed ML tools allow us to identify the variables

that affect participation in insurance, their robustness of selection and their importance.The general conclusion show as ML is a helpful tool for exploring the factors that explain farmers' participation in insurance schemes. Furthermore, the results obtained can help design better insurance schemes and, hopefully, boost farmers' involvement. Therefore, ML approaches can be useful in supporting the design of insurance schemes and should be done carefully considering the characteristics of the empirical case study. Insurance companies can use these tools to create a more suitable insurance product that only considers the characteristics that affect insurance adherence or to create a more specific marketing program.