mathematics

Article

Machine Learning Models for Predicting Romanian Farmers' Purchase of Crop Insurance

Codruţa Mare [1,*], Daniela Manaţe [1], Gabriela-Mihaela Mureşan [2], Simona Laura Dragoş [2], Cristian Mihai Dragoş [1] and Alexandra-Anca Purcel [1]

[1] Department of Statistics-Forecasts-Mathematics, Faculty of Economics and Business Administration, and The Interdisciplinary Centre for Data Science, Babeș-Bolyai University, 400591 Cluj-Napoca, Romania
[2] Department of Finance, Faculty of Economics and Business Administration, Babeș-Bolyai University, 400591 Cluj-Napoca, Romania
* Correspondence: codruta.mare@econ.ubbcluj.ro or codruta.mare@ubbcluj.ro; Tel.: +40-0264-418-652

Abstract: Considering the large size of the agricultural sector in Romania, increasing the crop insurance adoption rate and identifying the factors that drive adoption can present a real interest in the Romanian market. The main objective of this research was to identify the performance of machine learning (ML) models in predicting Romanian farmers' purchase of crop insurance based on crop-level and farmer-level characteristics. The data set used contains 721 responses to a survey administered to Romanian farmers in September 2021, and includes both characteristics related to the crop as well as farmer-level socio-demographic attributes, perception about risk, perception about insurers and knowledge about agricultural insurance. Various ML algorithms have been implemented, and among the approaches developed, the Multi-Layer Perceptron Classifier (MLP) and the Linear Support Vector Classifier (SVC) outperform the other algorithms in terms of overall accuracy. Tree-based ensembles were used to identify the most prominent features, which included the farmer's general perception of risk, their likelihood of engaging in risky behaviour, as well as their level of knowledge about crop insurance. The models implemented in this study could be a useful tool for insurers and policymakers for predicting potential crop insurance ownership.

Keywords: machine learning; crop insurance; classification

MSC: 62-08; 62P05; 62P20; 68T07

JEL Classification: C38; C51; C52; G22; Q14

Citation: Mare, C.; Manaţe, D.; Mureşan, G.-M.; Dragoş, S.L.; Dragoş, C.M.; Purcel, A.-A. Machine Learning Models for Predicting Romanian Farmers' Purchase of Crop Insurance. *Mathematics* **2022**, *10*, 3625. https://doi.org/10.3390/math10193625

Academic Editor: Denis N. Sidorov

Received: 20 August 2022
Accepted: 29 September 2022
Published: 3 October 2022

1. Introduction

Crop insurance offers farmers and agricultural producers financial protection against crop damage caused by natural events or disasters. Within the EU, in 2019 Romania had the largest number of workers employed in the agricultural sector [1]. Considering the sector size, increasing the crop insurance adoption rate and identifying the factors that drive adoption could present a real interest in the Romanian market. The continuous collection of insureds' data can enable insurance companies to implement machine learning (ML) models for targeting current or potential policyholders or for predicting insured lifetime value or attrition.

Technological developments have had a crucial impact on the insurance market, as on any other financial industry. In this line, machine learning (ML) algorithms receive special attention from researchers for addressing the following issues: insurance fraud detection [2–4], insurance premium prediction [5], underwriting process [6], claim analysis [7], risk prediction [8], sales forecasting [9], customer churn [10], and insurance tariff plans [11] among others.

A large number of studies are devoted to demonstrating the effectiveness of ML algorithms on forecasting. Accurate forecasting has attracted attention in various fields, such as price forecasting (e.g., authors in [12] proposed a novel machine-learning-based electricity price, while the authors in [13] integrated variational mode decomposition and random sparse Bayesian learning to forecast the oil prices), import and export forecasting (e.g., the authors in [14] used an econometrics-based co-integration model to estimate the natural gas demand, and those in [15] proposed NARX and Transformer models with regularization based on neural network models for mid-term forecasting of crop production and export, and showed that the values forecast by the proposed method were more accurate), financial products (e.g., the authors in [16] used machine learning algorithms to study the volatility of Bitcoin, and Hanafy and Ming [17] used ML for auto insurance) and others.

In insurance, the most used databases that use ML are those that come from car insurance [17–19]. We also identified databases from medical/healthcare insurance [20–22], life insurance [23] and agricultural insurance [24].

Specifically, to capture information about socio-environmental patterns in agriculture, ML is increasingly and widely applied, but generally, articles have focused on crop yield density and some other part of farmer behaviour. In this regard, the authors in [25] provide a review of ML in crop yield prediction, and emphasize that the most used ML algorithm seems to be the neural network, and the most widely used deep learning algorithm is the convolutional neural network. Recently, Wu and colleagues [26] explored a nonparametric ML tool based on Gaussian process regressions to predict crop yields over time and its applications to decision-making in crop insurance. They proposed models of non-stationary crop yields in a single stage and showed the utility of their method for insurance companies. Nguyen et al. [5] investigated the efficacy of ML in predicting farmer behaviour. They used data from 534 Vietnamese farmers, and showed that the insurance premium depended on factors such as quantity harvested, cost, province, and the farmer's desire to be insured.

This article focuses on the desire of Romanian farmers to take out insurance. Our research is even more important, as our literature search did not identify any papers on the behaviour of Romanian farmers that applied ML methodology. The main goal of the present paper is to fill this gap by showing the increased efficiency of using ML approaches in order to predict behavioural issues on the Romanian crop insurance market. It is well-known that Romania is a former communist developing country, and these patterns are worth investigating because they can decisively influence farmer behaviour. The articles that have focused on this topic have shown that there are a wide range of factors that affect behaviour. Among the most used factors remain the socio-demographic variables in works led by [27–29] and others. The typology of risks and the complex way in which they act outline the existence of agricultural insurance [30–33]. Other factors that have a significant influence on the insurance decision, but which directly influence the insurance premium that the farmer pays, are related to the characteristics of the land [34,35].

Narrowing the specificity of EU countries, the authors in [36], using a sample of 224 Bulgarian farmers interviewed in 2011, found that regional effects were one of the most influential factors in increasing the demand for crop insurance. Additionally, small and medium-sized farms were less likely to get insured compared to large farms. A similar result was obtained in a study in France [37]. The authors used a Logistic Regression and showed that large farms and risk exposure were predictors of the decision to take out crop insurance. In Spain, Garrido and Zilberman [38] found that premium subsidies explained an important proportion of the differences in farmers' insurance decisions. Hungarian crop farmers were positively influenced by education, size, and indebtedness of the crop, as seen in [39]. The same research also showed that crop-producing farms with an agricultural insurance contract were more efficient than the farmers without insurance. Using a structural model, Trestini et al. [40] obtained an interesting result in terms of the insurance intention of Italians and Poles. Risk aversion seemed to negatively influence the intention to purchase insurance, and previous insurance adoption at farm level as well

level of trust in the insurer were the main factors of future intention. In this regard, Iyer et al. [41] emphasized that in order to make predictions about farmers' behaviour, their risk preference and the heterogeneity in the level of their risk aversion need to be taken into account. Menapace et al. [42] showed, based on the regression analysis of risk and crop insurance purchases, that farmers in the Province of Trento, Northern Italy, were more likely to buy crop insurance if they were more risk averse.

In Romania, Dragos and Mare [43] studied the factors affecting crop insurance using a sample of 308 farmers from 18 villages from the six North-West Region counties (Cluj, Bihor, Bistrita -Nasaud, Maramures, Salaj and Satu Mare). The findings, based on a logit model, showed that education, age, distance from the farm to the nearest important city, size of the village and type of culture significantly influenced the decision to purchase crop insurance. Additionally, Romanian farmers that grew vegetables were more likely to purchase crop insurance. Unfortunately, we did not identify studies that integrated information on Romanians' perception of risk, or their level of knowledge in the field of agricultural insurance. There is already evidence in the insurance literature about the difference between education level and education in the field. Authors in [44] constructed the Index of Annuity Literacy and tested it for the German annuity market, and those in [45] constructed the Index of Insurance Knowledge for the Romanian private pension and life insurance market, and pointed out the important differences given by having or not having knowledge in the insurance field.

The main objective of this research was to identify the performance of ML techniques in predicting Romanian farmers' purchase of crop insurance. The models use crop-level and farmer-level characteristics from crop insurance purchase data collected in the span of one month—September 2021. Additionally, we aimed to identify the main contributing features in some of the models implemented in order to increase the awareness of the actors on this market in respect to what should be treated and how to enhance its development in a country like Romania, with massive agricultural potential. We achieved this by using the feature importance scores for the top 10 variables depicted by each analysis method.

The following Section 2 is devoted to describing the materials and methods used in this study. We explain the data and the methodology applied. The results of our research and their interpretations are given in Section 3. Section 4 concludes the paper.

2. Materials and Methods

2.1. Data Set

The data used in this research were collected from farmer responses to a survey administered by the Romanian Agency for Financing Rural Investments (AFIR) on Romanian farmers' crop insurance purchases. Data were collected in September 2021, both through Computer-Assisted Telephonic Interview (CATI) and an online platform. The final data set contains 721 entries from Romanian farmers. The respondents were aged between 21 and 68 years and lived in both rural and urban areas.

2.1.1. Predictors

The survey collected information regarding crop-level characteristics and farmer-level characteristics, all of which were used in the model. Table 1 shows the primary characteristics of the data set.

Table 1. Predictors.

Predictors		
Crop-level	principal crop type (CROP_TYPE_FIELD, CROP_TYPE_VEGETABLE, CROP_TYPE_TREEVINE, CROP_TYPE_OTHERS), region (CROP_REGION), area under cultivation (CROP_AREA), past experience with damage caused by natural calamities (CROP_CALAM)	
Farmer-level	socio-demographic attributes	AGE, residence (RURAL), education (EDUC), marital status (MARRIED), level of income (INCOME), percent of income attributed to agricultural activities (PERC_INCOME)
	farmer-insurance attributes	past experience with and trust in insurance companies (INS_EXP, INS_TRUST), knowledge about agricultural insurance (INS_EDUC)
	farmer-risk-aversion attributes	perceived risk of losing the crop (RISK_PERCEPTION_CROP), perception about ratio between the value of the total premium and the amount of risk to which the farm is exposed (RISK_OVEREVAL), perceived risk of various activities (RISK_PERCEPTION_GENERAL), likelihood of engaging in risky behavior (RISK_BEHAVIOUR)

Crop-level characteristics considered in this study were principal crop type (mostly field (57.4%), tree and vine (24.5%), vegetable (14.7%), and others), region (the 8 different official regions of Romania—NUTS2 level), and two ordinal variables indicating area under cultivation and past experience with damage caused by natural calamities. A total of 72.2% of respondents had an area under cultivation less than five hectares (compared to the 91.8% national statistic in 2016 [46], and 34.5% of the respondents had indicated that calamities in the past had caused large or very large damages to their crops.

Farmer-level characteristics included socio-demographic attributes, farmer-insurance attributes and farmer-risk-aversion attributes.

In terms of the socio-demographic variables, the mean and median age of the farmers was 46 years, with 77.7% residing in rural areas and 35.6% having completed higher education studies. Other attributes considered were marital status, level of income, as well as percent of income attributed to agricultural activities (with two-thirds of respondents having less than 50% of total income coming from agricultural activities).

Farmer-specific insurance attributes included past experience and trust in insurance companies, which were quantified on a 5-point Likert scale, from very unpleasant to very pleasant experience, and from minimum to maximum trust, respectively. A total of 54.4% of respondents expressed positive trust in insurance companies, while a smaller percentage (41.3%) had a positive experience with insurance companies. Additionally, respondents' agricultural insurance knowledge was evaluated using 7 questions.

The survey also collected data regarding farmers' risk-aversion characteristics. The perceived risk of losing a crop due to economic, legal, calamitous or other reasons was evaluated using a 5-point Likert scale, with 47.5% expressing very little or little fear. Respondents' risk perception about multiple activities (e.g., gambling or investing in stock or crypto markets) and their likelihood to engage in those activities were also evaluated using a 7-point Likert scale. One example of such an activity is gambling the monthly income on a sports event, with 50.3% considering it extremely risky, and 32.2% viewing as extremely unlikely to perform such an action. Additionally, farmers were also asked about the ratio between the value of the total premium and the amount of risk to which the farm is exposed, and whether that was under or over-valued, with 47.2% considering it an equitable ratio.

The RURAL variable (dummy variable indicating rural or urban residence) was highly correlated with MALE (dummy variable indicating male or female), and therefore the latter

was excluded from the analysis. The descriptive features of the data can be visualized in Appendix B.

2.1.2. Target Variable

A total of 423 (58.67%) respondents did not have crop insurance, 193 (26.77%) had standard insurance, while 105 (14.56%) had extended insurance. The target variable considered in all the models was the binary feature indicating whether the respondent owned or did not own crop insurance. Because the data set is not heavily imbalanced between the two classes (as can be seen in Table 2), no sampling techniques for dealing with imbalanced class distributions were considered in this approach.

Table 2. Target Variable.

Target	Value	Count
owns crop insurance	yes	298 (41.3%)
	no	423 (58.7%)

2.2. Data Processing

Responses about perceived risk of multiple activities were aggregated using the mean into a composite variable (RISK_PERCEPTION_GENERAL). The same method was applied for responses about likelihood of engaging in the same activities (RISK_BEHAVIOUR). The sum of the received answers to the 7 questions regarding agricultural insurance knowledge were aggregated into the composite variable INS_EDUC. Only the composite variables were further used in the model-building step. Input features were normalized using MinMaxScaler, which transforms all inputs in the [0, 1] range. In total, 27 input variables were further used in all the models specified further. The definition for MinMaxScaler can be seen below:

$$x\_scaled = x\_std * (max - min) + min \tag{1}$$

where

$$x\_std = \frac{x_i - min(x)}{max(x) - min(x)} \tag{2}$$

and min = lower value of desired transformation range;
max = upper value of desired transformation range;
$min(x)$ = the minimum value of feature x;
$max(x)$ = the maximum value of feature x.

2.3. Model Development

We used various modeling algorithms on the pre-processed data set: Baseline, Logistic Regression (LR), Decision Tree Classifier (DT), Random Forest Classifier (RFC), eXtreme Gradient Boosting Classifier (XGB), Linear Support Vector Classifier (SVC) and Multi-Layer Perceptron Classifier (MLP).

2.3.1. Baseline Model

The Baseline was built as a frame of reference to compare the existing models against, and it predicts the majority class (class 0—does not have insurance) in all situations. The Baseline disregards any patterns in the training data, and we expected all other implemented models to surpass the performance of this random classifier.

2.3.2. Logistic Regression (LR)

The Logistic Regression model is a widely used classification method, and can be written as:

$$h(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \tag{3}$$

where β are the parameters, and x the explanatory variable.

It returns the probability of class membership, and is based on the standard logistic function, which is defined as

$$f(x) = \frac{1}{1 + e^{-x}}, \tag{4}$$

which returns values between 0 and 1.

2.3.3. Decision Tree Classifier (DT)

A DT is a commonly used supervised classification method that has a flowchart structure. The tree achieves classification by splitting the data multiple times based on certain cutoff values in the features. At each split, different subsets of the initial data are created, with each instance belonging to one set. The leaf nodes represent the final subsets, while the intermediary ones are called split nodes. Decision trees carry a large explainability, and they can capture nonlinear relationships.

2.3.4. Tree-Based Models

Two tree-based ensemble methods were also considered—namely, Random Forest Classifier (RFC) and eXtreme Gradient Boosting Classifier (XGB). RFC is a "bagging" (bootstrap aggregating) ensemble method which implements in parallel multiple decision trees on bootstrapped samples, and the results are combined into a final model through a "majority vote" mechanism. XGB is a boosting ensemble model that trains decision trees sequentially, turning weak learners into strong learners.

2.3.5. Multi-Layer Perceptron Classifier (MLP)

An MLP is a feed-forward artificial neural network that maps a non-linear function from an input vector to an output vector. It is composed of a minimum of three layers: an input layer, one or more hidden layers and an output layer. Each layer is fully connected to the next layer, and non-linear activation functions are applied to neurons in each layer (except for the input layer). An advantage of the MLP is that it can distinguish data that is not linearly separable.

2.3.6. Model Implementation

Hyper-parameter optimization was performed using grid search and 5-fold cross-validation on the training set. The final models were evaluated on the test set, which was not used for tuning, in order to achieve an unbiased evaluation.

For the tree-based ensembles (RFC and XGB), the top 10 features contributing to the model were selected (representing 37% of total input features). Feature importance was employed to rank the variables based on their impact upon the decision to buy crop insurance. For robustness reasons, we present the feature importance results of both the RFC and the XGB.

3. Results and Discussions

3.1. Model Evaluation

The data set was split into an 80% training set and a 20% test set. Models were trained on the training set, using the optimal hyper-parameters identified on the same set with grid search. Performance was evaluated on the test set using overall accuracy and F1 score for each class as performance metrics. We also report precision and recall for Class 1 (owns insurance), as well as the area under the ROC curve (AUROC) score.

Accuracy reflects the proportion of correct predictions made out of the total predictions. Precision shows the ratio of true positives out of all cases predicted as positives. Recall indicates the proportion of true positives out of all actual positive cases. F1 score is the harmonic mean of precision and recall. A receiver operating characteristic (ROC) curve is a plot of the true positive rate (sensitivity, recall) against the false positive rate (1-specificity) at various thresholds. The AUROC score highlights the probability that a randomly chosen positive example is ranked higher than a randomly chosen negative example.

In terms of overall accuracy, the MLP had the best performance (0.76), followed by SVC and LR. The tree-based methods had an overall good performance, with the exception of the Decision Tree. MLP surpassed all implemented models in terms of overall accuracy (0.76), as well as F1 scores for both classes. SVC and LR followed in performance, with an overall accuracy of 0.74 and 0.72, respectively. All implemented models (except for the Baseline) had AUROCs above 0.5, indicating that they performed better than a random classifier (in our case, the Baseline model).

The performance comparison of all the models can be seen in Table 3.

Table 3. Performance metrics on the test set.

Algorithm	Accuracy	F1 Score Class 0	F1 Score Class 1	Precision Class 1	Recall Class 1	AUROC
Baseline	0.57	0.72	0.00	0.00	0.00	0.50
LR	0.72	0.76	0.67	0.69	0.65	**0.83**
DT	0.66	0.71	0.60	0.62	0.57	0.73
RFC	0.70	0.77	0.60	**0.73**	0.51	0.82
XGB	0.71	0.76	0.63	0.71	0.56	0.80
SVC	0.74	0.78	0.69	**0.73**	0.65	**0.83**
MLP	**0.76**	**0.79**	**0.72**	0.72	**0.73**	0.82

This is not a surprising result, considering that other articles related to the use of ML techniques in the financial sector also point out the efficiency of MLP against SVC and Logistic Regression (e.g., [47,48]). Bold indicates the maximum value.

3.2. Feature Importance from Ensemble Methods

Tree-based models have the advantage of ease of interpretation, as opposed to deep learning algorithms, which behave as a black box and limit the understanding of how features are combined to make predictions. Decision Trees are highly interpretable, as long as the depth of the tree is not very large. Bagging (e.g., RFC) and boosting (e.g., XGB) methods improve the performance by combining multiple Decision Trees, but are more difficult to interpret. For these cases, feature importance (i.e., the importance score of each variable in the construction process) is determined using computed information gain. For example, in RFC, feature importance is evaluated using the mean decrease impurity—namely, the average across trees of the total decrease in node impurity. In the boosting methods (e.g., XGB), the relative importance of each variable is higher because it is more used to take decisions. We plotted the feature importance to highlight the main contributing factors to the prediction of whether a farmer had or had not purchased crop insurance. Figures 1 and 2 contain the top 10 contributing features in the ensemble models implemented, RFC and XGB, respectively. The definition of each variable can be found in Table 1. Highlighted in green are the features not consistent across models within the top 10 contributions.

The top 10 factors point out several significant factors impacting Romanian farmers' decision to buy crop insurance. Among the crop-level characteristics [34,35], past experience with damage made by calamities was an important factor in the final purchasing decision of the farmer. However, the feature importance scores indicate that farmer-level variables were the most impactful. Among these, both the RFC and the XGB indicated the relative importance of risk aversion (consistent with [30–33,41,42] and education in the field [45]). Therefore, our results are in line with the literature.

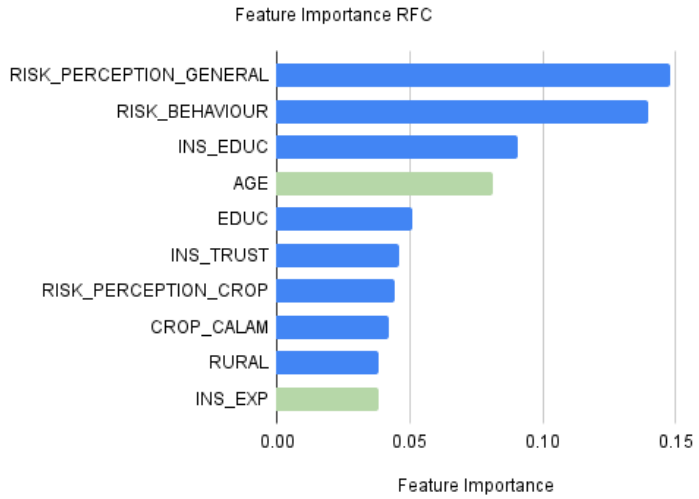

Figure 1. Feature Importance: RFC.

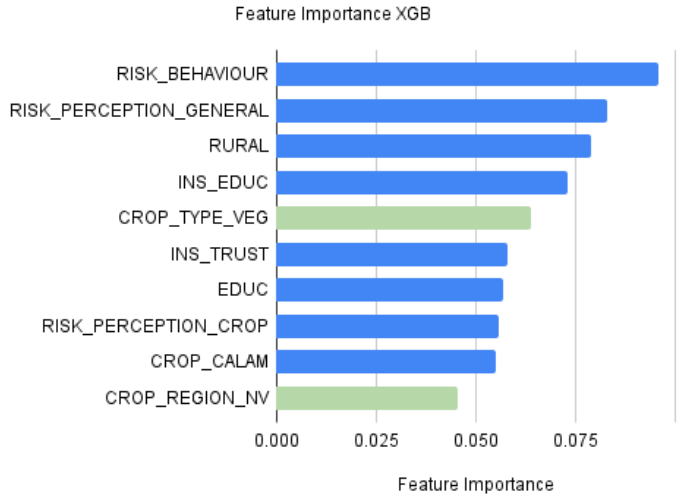

Figure 2. Feature Importance: XGB.

In terms of consistency across models, eight out of the top ten contributing features were common between the two ensemble algorithms. Both identified farmer-level characteristics related to risk as the main two contributing features: the general perception of risk and the likelihood of engaging in risky behaviour. Additionally, the general knowledge about crop insurance was another of the main features highlighted. The perceived risk of losing the crop and the level of damage caused by calamities in the past were also prominent factors. As expected, in line with economic risk theory, this is the key predictor of insurance. Insurers transfer the risk to the insured, and the higher the level of risk they perceive, the more likely they are to take out insurance [49]. In this regard, authors in [50] prove that, as farmers' perception of floods increases or as farmers become more risk-averse, they are more likely to buy crop insurance, but risk-seeking farmers are less likely to purchase crop insurance. Additionally, the general knowledge about crop insurance was another of the main features highlighted (similar to results obtained on other types of insurance [44,45]). Socio-demographic attributes such as the level of education completed by the respondent and the type of residence (rural versus urban) were identified within the main factors, along with the expressed level of trust in insurance companies (see [29,50,51]).

RFC additionally identified the age of the respondent and the perceived quality of the interaction with insurance companies in the past as two other contributing factors. Our results for age are in line with previous research [49,50,52] which showed that older farmers were more likely to purchase crop insurance than younger farmers. On the other hand, XGB results highlighted, on top of the eight main common features, crop-level attributes (e.g., whether it was a vegetable crop, or whether the crop was located in the North-Western region (see [36,43])).

4. Conclusions

In this research, we implemented and evaluated several machine learning models for predicting Romanian farmers' purchase of crop insurance. These models used variables related to crop characteristics, as well as farmer-level variables including socio-demographic attributes, characteristics related to risk perception, as well as insurance knowledge and perception about insurance companies. Among all the models developed, the best-performing ones were the Multi-Layer Perceptron Classifier (MLP) and the Linear Support Vector Classifier (SVC), with an overall accuracy of 0.76 and 0.74, respectively. To identify the main contributing features, tree-based ensemble methods were used, namely Random Forest Classifier (RFC) and eXtreme Gradient Boosting Classifier (XGB). In both approaches, the most important features included the farmer's general perception of risk, their likelihood of engaging in risky behaviour, as well as their level of knowledge about crop insurance, which is supported by the previous research of [49,50].

Crop insurance market actors could use these models to better predict whether farmers would own or not crop insurance, based on crop-level and farmer-level attributes. The farmers that are predicted to own crop insurance can be further targeted and potentially converted into an existing customer base. They can be approached with offers based on more complex crop insurance policies, but much more adapted to their real needs. At the same time, actors on the market may use our results to address the issues that determine the other group of farmers' decision not to get insured and treat these. Additionally, taking into account the most prominent features identified, insurers should first consider the farmers' perception and attitude toward risk, and should potentially invest in increasing the level of knowledge about agricultural insurance. As previously stated in the insurance literature, we emphasize the different importance of the general educational level versus education in the field. The level of knowledge in the insurance sector, in general, and in crop insurance in particular, is of major importance for the farmer's decision and for the future development of the market. An uneducated farmer in the field will not be able to take a feasible decision in respect to purchasing crop insurance. Consequently, we point out the need for market actors to invest in educating their target clients.

The model results indicate that machine learning can be a useful tool for predicting farmers' purchase of crop insurance, and predictions can be further used by insurers for better targeting of potential policyholders. Additionally, we emphasize the different efficiencies of several ML methods in modelling the data. We obtained a similar ranking to other important studies in the field's literature.

Behavioural aspects are non-linear and complex, not only in the crop insurance field, but in any other. The main advantage of machine learning approaches is the fact that they are able to better treat the non-linear relationships that exist in the behavioural field, in contrast with the classical methods that are more conservative and use the linear approach. Consequently, they may lose important information provided by the data. Our results clearly show that machine learning techniques can be used in a very efficient way to predict purchasing decisions with respect to crop insurance. One important advantage is that our methodology can be extended and used in any other sector in which behavioural assessment is required.

The most important limitation of our study, in respect to the part of the field's research which employs classical econometric approaches, is that, for example, we only show the

importance of different variables (features) and rank them accordingly, without showing the direction of the impact. This is an aspect that we intend to treat in future developments.

This research can also be further improved by fragmenting the type of insurance the farmer holds and building a 3-class model (no insurance, standard insurance, extended insurance) instead of the binary one considered in this research. In this way we will contribute to the field's literature by adding new information that can help insurers and policymakers to better target the different groups of farmer clients.

Author Contributions: All authors have read and agreed to the published version of the manuscript. Conceptualization, C.M., D.M., G.-M.M., S.L.D., C.M.D. and A.-A.P.; Data curation, S.L.D., C.M.D. and A.-A.P.; Formal analysis, D.M.; Investigation, C.M., D.M. and G.-M.M.; Methodology, C.M. and D.M.; Project administration, S.L.D.; Resources, G.-M.M. and C.M.D.; Software, D.M.; Supervision, C.M.; Validation, C.M.; Visualization, D.M. and C.M.D.; Writing—original draft, C.M., D.M., G.-M.M. and S.L.D.; Writing—review & editing, C.M., D.M. and C.M.D.

Funding: This work was funded by a grant from the Romanian Ministry of Education and Research, CNCS-UEFISCDI, project number PN-III-P1-1.1-TE-2019-0554, within PNCDI III.

Acknowledgments: This paper is part of the project COST CA19130 FinAI - Fintech and Artificial Intelligence in Finance—Towards a Transparent Financial Industry. We also thank the participants in FINANCECOM2022, 23–24 August 2022, Enschede, The Netherlands, for their valuable feedback and comments.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The tuned hyperparameters can be seen in the table below for each of the implemented models.

Table A1. Model Parameters.

Algorithm	Parameters
LR	C=1, max_iter = 5000, random_state = 42, solver = 'newton-cg'
DT	max_depth = 10, max_features = 'sqrt', min_samples_leaf = 5, random_state = 42
RFC	max_depth = 8, max_features = 'sqrt', n_estimators = 500, random_state = 42
XGB	colsample_bytree=0.4, gamma = 2, max_depth = 2, min_child_weight = 5, random_state = 42, subsample = 0.6
SVC	C = 5, loss = 'hinge', max_iter = 20000, random_state = 42, tol = 0.01
MLP	alpha = 0.01, number_neurons = 100, activation = 'relu', learning_rate_init = 0.01, random_state = 42

Appendix B

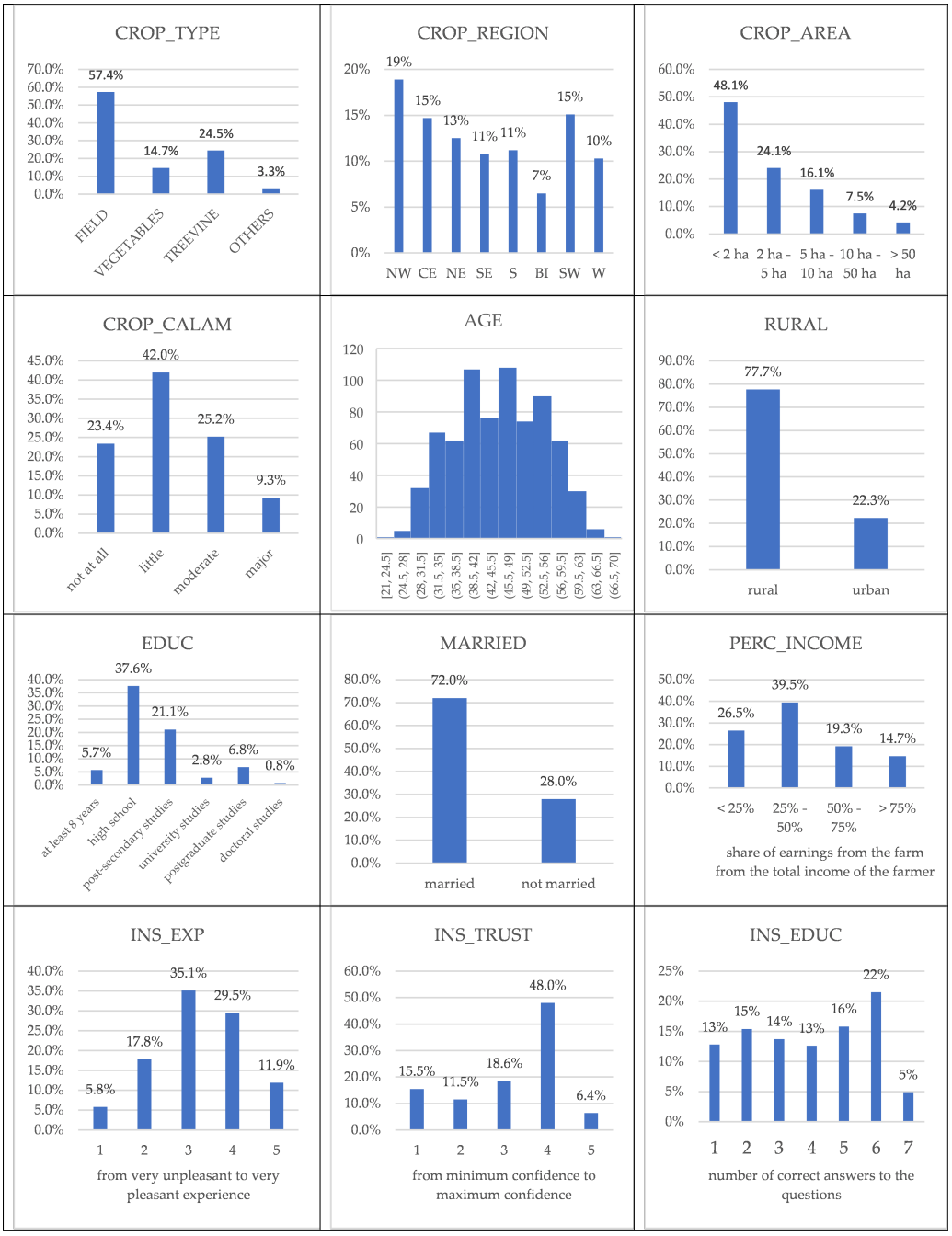

Figure A1. Descriptive Plots.

References

1. European Parliament. Available online: https://www.europarl.europa.eu/news/en/headlines/society/20211118STO17609/eu-agriculture-statistics-subsidies-jobs-production-infographic (accessed on 17 May 2022).
2. Wang, Y.; Xu, W. Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decis. Support Syst.* **2018**, *105*, 87–95.
3. Waghade, S.S.; Karandikar, A.M. A comprehensive study of healthcare fraud detection based on machine learning. *Int. J. Appl. Eng. Res.* **2018**, *13*, 4175–4178.
4. Rukhsar, L.; Bangyal, W.H.; Nisar, K.; Nisar, S. Prediction of insurance fraud detection using machine learning algorithms. *Mehran Univ. Res. J. Eng. Technol.* **2022**, *41*, 33–40.
5. Nguyen, K.A.T.; Nguyen, T.A.T.; Nguelifack, B.M.; Jolly, C.M. Machine Learning Approaches for Predicting Willingness to Pay for Shrimp Insurance in Vietnam. *Mar. Resour. Econ.* **2022**, *37*, 155–182.

6. Biddle, R.; Liu, S.; Tilocca, P.; Xu, G. Automated underwriting in life insurance: Predictions and optimisation. In Proceedings of the Australasian Database Conference, Brisbane, Australia, 14–16 July 2018; pp. 135–146.
7. Hanafy, M.; Ming, R. Classification of the Insureds Using Integrated Machine Learning Algorithms: A Comparative Study. In *Applied Artificial Intelligence*; Taylor & Francis: Abingdon, UK, 2022; pp. 1–32.
8. Boodhun, N.; Jayabalan, M. Risk prediction in life insurance industry using supervised learning algorithms. *Complex Intell. Syst.* **2018**, *4*, 145–154.
9. Gopagoni, D.R.; Lakshmi, P.; Siripurapu, P. Predicting the Sales Conversion Rate of Car Insurance Promotional Calls. In *Rising Threats in Expert Applications and Solutions*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 321–329.
10. Groll, A.; Wasserfuhr, C.; Zeldin, L. Churn modeling of life insurance policies via statistical and machine learning methods—Analysis of important features. *arXiv* **2022**, arXiv:2202.09182.
11. Henckaerts, R.; Côté, M.P.; Antonio, K.; Verbelen, R. Boosting insights in insurance tariff plans with tree-based machine learning methods. *N. Am. Actuar. J.* **2021**, *25*, 255–285.
12. Yang, W.; Sun, S.; Hao, Y.; Wang, S. A novel machine learning-based electricity price forecasting model based on optimal model selection strategy. *Energy* **2022**, *238*, 121989.
13. Li, T.; Qian, Z.; Deng, W.; Zhang, D.; Lu, H.; Wang, S. Forecasting crude oil prices based on variational mode decomposition and random sparse Bayesian learning. *Appl. Soft Comput.* **2021**, *113*, 108032.
14. Lin, B.; Wang, T. Forecasting natural gas supply in China: production peak and import trends. *Energy Policy* **2012**, *49*, 225–233.
15. Devyatkin, D.; Otmakhova, Y. Methods for Mid-Term Forecasting of Crop Export and Production. *Appl. Sci.* **2021**, *11*, 10973.
16. Pabuçcu, H.; Ongan, S.; Ongan, A. Forecasting the movements of Bitcoin prices: an application of machine learning algorithms. *Quant. Financ. Econ.* **2020**, *4*, 679–692.
17. Hanafy, M.; Ming, R. Machine learning approaches for auto insurance big data. *Risks* **2021**, *9*, 42.
18. Maillart, A. Toward an explainable machine learning model for claim frequency: a use case in car insurance pricing with telematics data. *Eur. Actuar. J.* **2021**, *11*, 579–617.
19. Dimri, A.; Paul, A.; Girish, D.; Lee, P.; Afra, S.; Jakubowski, A. A multi-input multi-label claims channeling system using insurance-based language models. *Expert Syst. Appl.* **2022**, *202*, 117166.
20. Kose, I.; Gokturk, M.; Kilic, K. An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. *Appl. Soft Comput.* **2015**, *36*, 283–299.
21. Hsieh, C.Y.; Su, C.C.; Shao, S.C.; Sung, S.F.; Lin, S.J.; Yang, Y.H.K.; Lai, E.C.C. Taiwan's national health insurance research database: past and future. *Clin. Epidemiol.* **2019**, *11*, 349.
22. Mitrova, H.; Madevska Bogdanova, A. Models for Detecting Frauds in Medical Insurance. In Proceedings of the International Conference on ICT Innovations, Skopje, Macedonia, 29 September–1 October 2022; pp. 55–67.
23. Azzone, M.; Barucci, E.; Moncayo, G.G.; Marazzina, D. A machine learning model for lapse prediction in life insurance contracts. *Expert Syst. Appl.* **2022**, *191*, 116261.
24. Wei, C.; Dan, L. Market fluctuation and agricultural insurance forecasting model based on machine learning algorithm of parameter optimization. *J. Intell. Fuzzy Syst.* **2019**, *37*, 6217–6228.
25. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, *177*, 105709.
26. Wu, W.; Wu, X.; Zhang, Y.Y.; Leatham, D. Gaussian process modeling of nonstationary crop yield distributions with applications to crop insurance. *Agric. Financ. Rev.* **2021**, *81*, 767–783.
27. Olila, D.O.; Pambo, K.O. Determinants of farmers' awareness about crop insurance: Evidence from Trans-Nzoia County, Kenya. In Proceedings of the 8th Annual Egerton University International, 26–28 March 2014.
28. Gulseven, O. Estimating the demand factors and willingness to pay for agricultural insurance. *arXiv* **2020**, arXiv:2004.11279.
29. Sihem, E. Economic and socio-cultural determinants of agricultural insurance demand across countries. *J. Saudi Soc. Agric. Sci.* **2019**, *18*, 177–187.
30. Skees, J. Rethinking the role of index insurance: accessing markets for the poor. In Proceedings of the AAAE/AEASA Conference, Westin Grand Hotel, Cape Town, South Africa, 19–23 September 2010; Volume 22.
31. Hill, R.V.; Hoddinott, J.; Kumar, N. Adoption of weather-index insurance: learning from willingness to pay among a panel of households in rural Ethiopia. *Agric. Econ.* **2013**, *44*, 385–398.
32. Binswanger-Mkhize, H.P. Is there too much hype about index-based agricultural insurance? *J. Dev. Stud.* **2012**, *48*, 187–200.
33. Ghahari, A.; Newlands, N.K.; Lyubchich, V.; Gel, Y.R. Deep learning at the interface of agricultural insurance risk and spatio-temporal uncertainty in weather extremes. *N. Am. Actuar. J.* **2019**, *23*, 535–550.
34. Enjolras, G.; Capitanio, F.; Adinolfi, F. The demand for crop insurance: Combined approaches for France and Italy. *Agric. Econ. Rev.* **2012**, *13*, 5–22.
35. Carrer, M.J.; Silveira, R.L.F.d.; Vinholis, M.d.M.B.; De Souza Filho, H.M. Determinants of agricultural insurance adoption: evidence from farmers in the state of São Paulo, Brazil. *RAUSP Manag. J.* **2021**, *55*, 547–566.
36. Lefebvre, M.; Nikolov, D.; Gomez-y Paloma, S.; Chopeva, M. Determinants of insurance adoption among Bulgarian farmers. *Agric. Financ. Rev.* **2014**, *74*, 326–347.
37. Enjolras, G.; Sentis, P. Crop insurance policies and purchases in France. *Agric. Econ.* **2011**, *42*, 475–486.
38. Garrido, A.; Zilberman, D. Revisiting the demand of agricultural insurance: the case of Spain. *Agric. Financ. Rev.* **2008**, *68*, 43–66.

39. Zubor-Nemes, A.; Fogarasi, J.; Molnár, A.; Kemény, G. Farmers' responses to the changes in Hungarian agricultural insurance system. *Agric. Financ. Rev.* **2018**, *78*, 275–288.
40. Trestini, S.; Giampietri, E.; Smiglak-Krajewska, M. Farmer behaviour towards the agricultural risk management tools provided by the CAP: A comparison between Italy and Poland. In Proceedings of the 162nd EAAE Seminar, Budapest, Hungary, 26–27 April 2018.
41. Iyer, P.; Bozzola, M.; Hirsch, S.; Meraner, M.; Finger, R. Measuring farmer risk preferences in Europe: A systematic review. *J. Agric. Econ.* **2020**, *71*, 3–26.
42. Menapace, L.; Colson, G.; Raffaelli, R. A comparison of hypothetical risk attitude elicitation instruments for explaining farmer crop insurance purchases. *Eur. Rev. Agric. Econ.* **2016**, *43*, 113–135.
43. Dragos, S.L.; Mare, C. An econometric approach to factors affecting crop insurance in Romania. *Econ. Manag.* **2014**, *17*, 93–103.
44. Goedde-Menke, M.; Lehmensiek-Starke, M.; Nolte, S. An empirical test of competing hypotheses for the annuity puzzle. *J. Econ. Psychol.* **2014**, *43*, 75–91.
45. Dragos, S.L.; Dragos, C.M.; Muresan, G.M. From intention to decision in purchasing life insurance and private pensions: Different effects of knowledge and behavioural factors. *J. Behav. Exp. Econ.* **2020**, *87*, 101555.
46. Eurostat. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Farms_and_farmland_in_the_European_Union_-_statistics#Farms_in_2016 (accessed on 16 May 2022).
47. Ryman-Tubb, N.F.; Krause, P.; Garn, W. How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. *Eng. Appl. Artif. Intell.* **2018**, *76*, 130–157.
48. Rahimikia, E.; Mohammadi, S.; Rahmani, T.; Ghazanfari, M. Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran. *Int. J. Account. Inf. Syst.* **2017**, *25*, 1–17.
49. Sherrick, B.J.; Barry, P.J.; Ellinger, P.N.; Schnitkey, G.D. Factors influencing farmers' crop insurance decisions. *Am. J. Agric. Econ.* **2004**, *86*, 103–114.
50. Fahad, S.; Wang, J.; Hu, G.; Wang, H.; Yang, X.; Shah, A.A.; Huong, N.T.L.; Bilal, A. Empirical analysis of factors influencing farmers crop insurance decisions in Pakistan: Evidence from Khyber Pakhtunkhwa province. *Land Use Policy* **2018**, *75*, 459–467.
51. Deressa, T.T.; Ringler, C.; Hassan, R.M. *Factors Affecting the Choices of Coping Strategies for Climate Extremes. The Case of Farmers in the Nile Basin of Ethiopia*; IFPRI Discussion Paper; International Food Policy Research Institute: Washington, DC, USA, 2010; Volume 1032.
52. Okoffo, E.D.; Denkyirah, E.K.; Adu, D.T.; Fosu-Mensah, B.Y. A double-hurdle model estimation of cocoa farmers' willingness to pay for crop insurance in Ghana. *SpringerPlus* **2016**, *5*, 873.