CYBERNETICS AND PHYSICS, VOL. 12, NO. 3, 2023, 174–181

MACHINE LEARNING FOR CROP YIELD FORECASTING

Bolotbek Biibosunov
Arabaev Kyrgyz State University
Bishkek, Kyrgyzstan
name1@university1.country1

Baratbek Sabitov
KNU named after Jusup Balasagyn
Bishkek, Kyrgyzstan
name1@university1.country1

Saltanat Biibosunova
Arabaev Kyrgyz State University
Bishkek, Kyrgyzstan
name2@university2.country2

Zhamin Sheishenov
Arabaev Kyrgyz State University
Bishkek, Kyrgyzstan
name2@university2.country2

Sharshenbek Zhusupkeldiev
KNU named after Jusup Balasagyn
Bishkek, Kyrgyzstan
name2@university2.country2

Zhyldyz Mamadalieva
KRSU named after B. Yeltsyn
Bishkek, Kyrgyzstan
name2@university2.country2

Article history:
Received 27.08.2023, Accepted 16.09.2023

Abstract

Amid the persistent rise in global population, there has been a heightened focus on food security by academia, governmental initiatives, and international endeavors. Food security serves as a critical pillar in the national security framework, contributing to a nation's sovereignty and self-sufficiency in food supply. To fulfill global requirements for essential food items, there is an imperative need to enhance agricultural efficiency across countries. Concurrently, agricultural practices must align with contemporary quality standards and meet consumer needs, drawing upon an integrated approach to crop cultivation technologies and yield classifications. Methodologies and tools for yield augmentation, grounded in scientific advancements in predictive modeling, are of paramount importance. Investigating the plethora of variables that contribute to optimal crop development, which in turn influences yield, poses significant challenges. Comprehensive inquiries that incorporate cutting-edge scientific and technological methodologies are essential for creating precise yield forecasts. The evolving landscape of yield modeling and prediction has emerged as a technologically sophisticated domain. Advanced methods such as machine learning and deep learning offer robust platforms for addressing crop yield forecasting, particularly when coupled with extensive datasets on environmental variables. A growing body of literature suggests the promising role of computational technologies and machine learning paradigms, inclusive of various forms of remote sensing data, in fine-tuning yield models. Yield prediction models are often characterized by intricate nonlinear equations influenced by a range of factors: seed quality and diversity, soil attributes, climatic variables, fertilizer usage, and other agronomic practices. The impacts of these variables on crop yield are varied, with some exerting greater influence than others. Additionally, crop yield is susceptible to adverse environmental and climatic conditions. While there exists a rich corpus of research on yield forecasting, addressing this issue remains an exigent priority in the agricultural sector.

Key words

Yield, modeling, forecasting, machine learning, crops, regression.

1 Introduction

Before the appearance of contemporary yield forecasting technologies supported scientific approaches, farmers were supported their own experience of growing various crops, which led to different results. as an example, for several years, in Kyrgyzstan, farmers from different regions grew the identical form of crops (for example, potatoes), as a results of an outsized potato harvest, its value on the market fell sharply and this was repeated from year to year, which led to an outflow of farmers from agriculture. In fact, the method of outflow of farmers also requires special attention and study, it's directly associated with the migration process of the population, especially the able-bodied and young. As you recognize, there are many factors that affect crop yields, like the world of sowing plants with proper pretreatment, the effective use of irrigation systems and its improvement, it's necessary to require under consideration weather changes and climate features of the region, support and improvement of agricultural companies for the withdrawal of recent kinds of crops taking into consideration regional characteristics. Proper management

of agrotechnical agriculture for the utilization of fertilizers, to conduct continuous monitoring and timely detection of diseases of agricultural plants. The study of the many complex relationships between traits that affect accurate modelling and forecasting of plant yields has now become possible due to the emergence of powerful machine learning algorithms and deep learning neural network design technologies. Taking into consideration these important factors, as an example, for wheat, this problem was investigated in [Xu et al. 2019]. the employment of reliable, near natural phenomena data currently also plays a key role in obtaining the specified harvest [Filippi et al. 2019, Liakos et al. 2018, Kitchenham et al. 2007]. The category of productivity and its forecasting tasks for agricultural crops refers to a fancy section of forecasting and modelling. In recent years, various authors are conducting research during this area supported machine learning for various cultures [Sujatha et al. 2016, Ying-xue et al. 2017, Everingham et al. 2009, Mola-Yudego et al. 2016] within the works [Paul et al. 2015, Rahman et al. 2015, Kunapuli et al. 2015, Charoen et al. 2019, You et al. 2017, Brown et al. 2017, Everingham et al. 2019], classification and regression problems were studied employing a style of machine learning algorithms for yield problems. Page 1 of 2 the fundamentals of a random forest in relevancy applied problems are investigated within the works [Breiman 2001, Breiman 1996]. The foremost complete review of the research literature for this area is given within the work [Thomas et al. 2020] the most purpose of this work is to use machine learning methods for agricultural tasks. In the study of yield modelling and forecasting, machine learning algorithms multivariate analysis (LR), Lasso regression (Lasso R), stochastic gradient descent (SGD), decision tree (RT) were employed in this work, which give good results for several agricultural tasks. At the identical time, the algorithms of K – nearest neighbors (KNN) [Romero et all, 2013], random forest (RF) [Mola-Yudego et al.2016] are important and effective in modelling and forecasting yields, where it's required to research many main influencing factors, the support vector machine (SVR) [Girish et al. 2018] and gradient boosting variants (GBR) [Huber et al. 2022].

2 Methods

In this study, data were gathered from five distinct districts within the Issyk-Kul region of Kyrgyzstan. These data encapsulate variables such as the types and amounts of fertilizers employed (i.e., nitrogen, phosphorus, and potassium), local climatic conditions (i.e., temperature, humidity, and precipitation rates), soil acidity levels, as well as potato yield metrics specific to each district. To facilitate analysis using Python libraries, this collected dataset was organized into a consolidated .csv file for each district under consideration. A salient aspect influencing the construction of our yield model pertains to the attrition rate among farmers. This phenomenon is analogously studied in the context of customer churn in the telecommunications sector [Kashnitsky , 2017]. Clearly, optimal yields cannot be realized without consistent farmer engagement. Within the Issyk-Kul region, an in-depth analysis was conducted to assess the rate of farmer attrition. The findings are presented as histograms in Figure 1.

As elucidated in Figure 1, farmer attrition manifests variably depending on multiple elements. Notably, this trend is acute among farmers in agrarian settings who lack personal agricultural machinery. In such contexts, farmers often resort to equipment rentals, which surprisingly correlates with higher rates of attrition as evident in Figure 1(b) and 1(c). One plausible explanation is that the unregulated and substantial costs incurred from machinery rental create points of contention, leading to dissatisfaction among agricultural producers and, ultimately, their departure from the field. Another vital determinant of crop yield is the prevalence of weed species. From seed sowing to harvest, it is essential to safeguard crops from not just weeds, but also pests, diseases, and adverse climatic conditions like droughts and floods. The growing incidence of extreme weather events is altering traditional growing seasons and affecting water availability, thereby exacerbating the proliferation of diverse weed types, pests, and fungi, all of which can negatively impact yields. Effective strategies are crucial for protecting crops from these threats, including technologies aimed at promoting robust plant health. In this study, we examine the specific impact of various weed classifications on agricultural crops within the five districts of the region under investigation. The analysis led to the development of a dataset categorizing weeds into annual, perennial, and parasitic types, each with distinct detrimental effects on crop yield. For instance, it was ascertained that annual weeds predominantly affect tuberous plants, as illustrated in Figure 2.

When collecting data , the features of each of the five districts were taken into account , and for the convenience of working with Python libraries , the data was presented in the form .csv files. All yield data are subject to the normal distribution law. The correlation matrix Fig.3. the database under study is distributed as follows, where N, P and K mean pesticides: nitrogen, phosphorus and potash fertilizers, the independent weather variables temperature, humidity, rainfall mean temperature, humidity and precipitation, respectively. The pH variable means the acidity of the soil of the studied regions. We can see from the correlation matrix that some independent factor data admit a small multicollinearity and even show its absence. This distribution allows us to investigate the tasks of forecasting the dependent variable yield. In generalization, the multiple dependent variable harvest –yield, in our case, does not correlate with other independent variables. All our studied factor data obey the normal distribution law Figure 4.

In this paper, multiple linear regression has one target

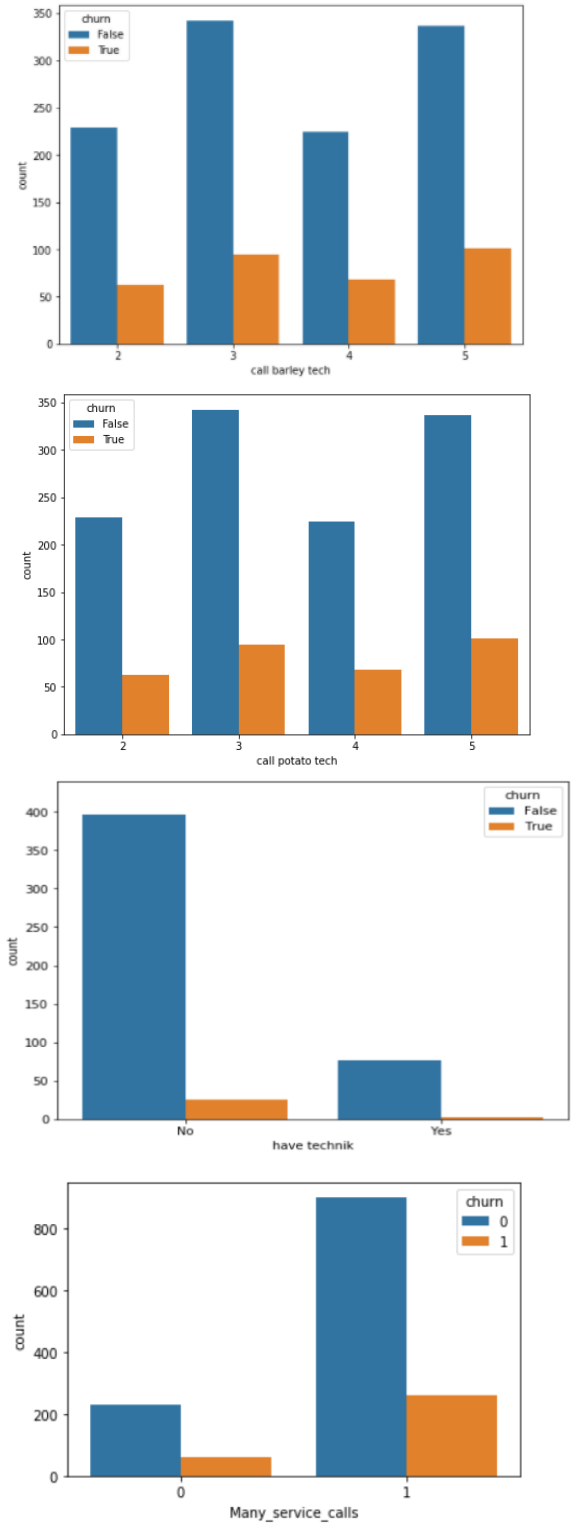

Figure 1. Outflow of farmers (0 - no outflow, 1- there is an outflow) depending on the cases: a) when using agricultural machinery; b), c) renting agricultural machinery for growing barley and potatoes; d) many other services using

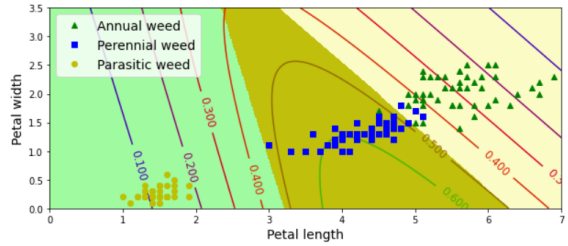

Figure 2. The results of the classification of weed plants into three classes: annual, perennial and parasitic

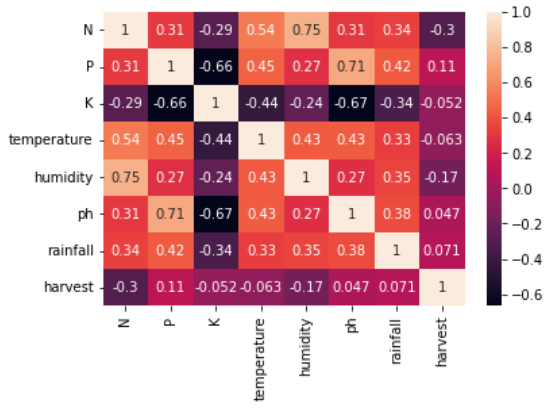

Figure 3. The results of the construction of the correlation matrix of the predicted factor -yield and other influencing factors

variable y and two or more independent variables x. This is an extension of simple linear regression, since more than one predictor variable is required to predict the target variable. Assumptions for multiple linear regression: 1. There must be a linear relationship between the target and predictor variables. 2. The regression residuals should be normally distributed. 3. MLR assumes little or no multicollinearity (correlation between independent variables) in the data. The dependent variable yield as an initial representation in the form of a multiple regression model has the following form:

$$Y = \omega_0 + \omega_1 X_1 + \omega_2 X_2 + ... + \omega_p X_{p+\varepsilon} \qquad (1)$$

where Y – the dependent variable yield, X_p – predictor variable components of the multiple regression, ω_p – unknown coefficients, and ε – model error. It is easy to see that the vector:

$$\vec{\omega} \qquad (2)$$

is an algorithm for determining the unknown coefficients ω_i in (1) and the minimum of the next problem for the quadratic functional:

$$L = \Sigma(y_i - \hat{y}_i)^2 \Sigma(y_i - \hat{\omega_0} - \hat{\omega_0} X_i 1 - \hat{\omega_2} X_i 2 - \\ ... - \hat{\omega_p} X_i p)^2 \to \min \qquad (3)$$

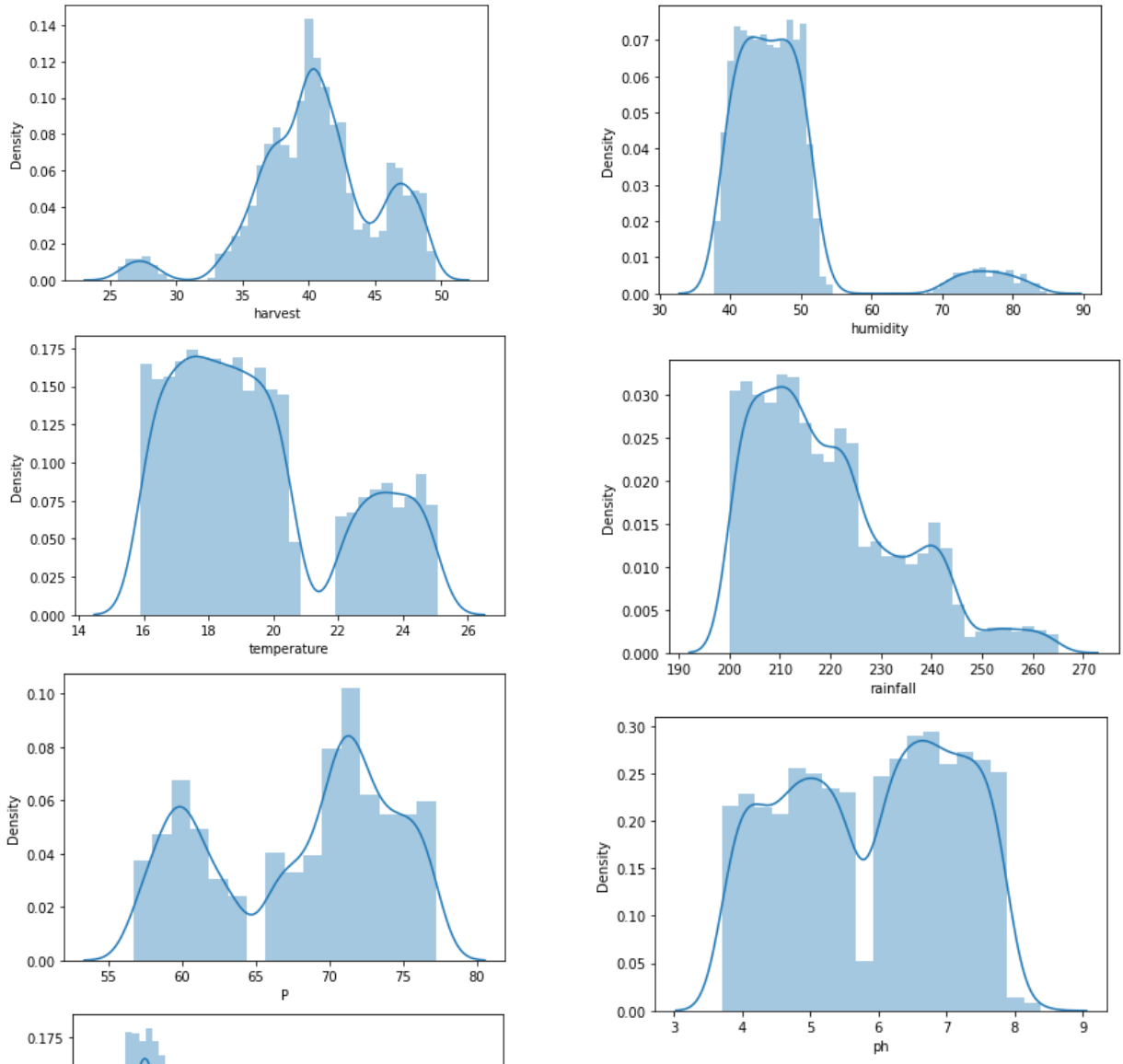

Figure 4. Distribution of data, for example, the predicted factor of yield and temperature according to the normal law

where $\hat{y}_i$ are the predicted values for the exact values and summation y_i is performed for $i = 1, n$. Generally speaking, calculations using the algorithm (2) lead to over-trained models and have no generalizing properties, losing meaning for the test data. The model created by this algorithm will interpolate the data as much as desired instead of extrapolating when testing the data. To solve this problem, some regularizing function R($\vec{\omega}$) is introduced. Let us now formulate the problem of determining the weight rations $\vec{\omega}$ with the transformed error functional in the form (4):

$$\Omega(X, \vec{y}, \vec{\omega}) = L(X, \vec{y}, \vec{\omega}) + \lambda R(\vec{\omega}) \Rightarrow \min \quad (4)$$

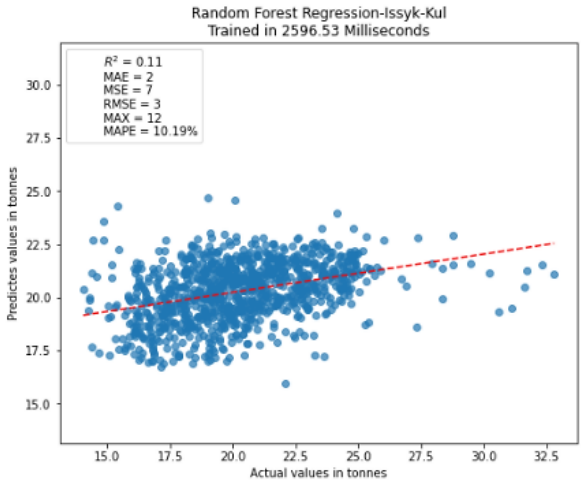

Figure 7. Accumulations of data around the regression line when using a random forest

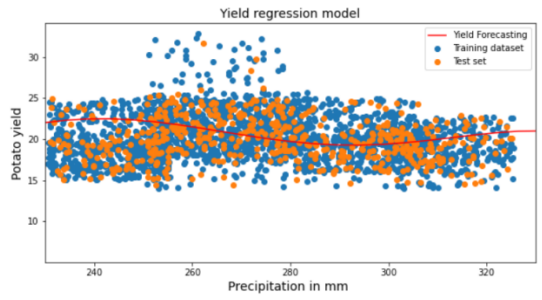

Figure 5. Regression yield prediction obtained using a regularizing algorithm

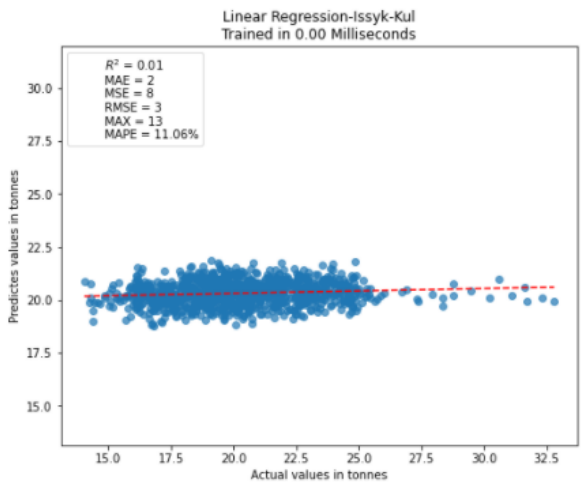

Figure 6. Summary result of multiple regression and data clustering around a multiple factor regression line

where λ is called the regularization ratio, Ω — the unit matrix.

3 Results and discussions

Below are the results obtained by applying the regularizing algorithms of the problem (1), (4). The results obtained when creating models for one-dimensional and multiple regression of yield forecasting in relation to the studied database by region are shown in Figure 5 and Figure 6.

Here is the result of the regression equation for a specific case when the yield depends on nitrogen fertilizer, temperature and precipitation:

$$harvest = 18.196024931744926 + 0.007244397628626693 * N + 0.05654913521333674 * t°C + 0.0002257625611471747 5 * rainfall \quad (5)$$

The use of the random forest algorithm is shown in Figure 7.

Regarding the results obtained using the Lasso algorithm for predicting yield problems, the algorithm for this method is described below for the overfitted models. The error function that is used to find the regression ratios can be used as the LASSO operator. In this method, which is also a regularization method, the coefficients on correlated features are excluded, i.e., their coefficient is set to zero by this method [James et al. 2013]. A so-called penalty term (ω_i) is added to the error function of the linear regression model in LASSO, which can reduce the coefficients to zero (L1 regularization). The LASSO loss function, in this case, is as follows [Tibshirani et al. 1996]:

$$L = \Sigma(y_i - \hat{y_i})^2 + \alpha \Sigma V \omega_i V \quad (6)$$

where in (6), is the parameter responsible for the convergence rate, which must be determined before performing the training task. The effectiveness of this algorithm and the forecast results for our case are shown in Figure 9 and in Table 1. The paper investigates the process of constructing a model using visualization of a model based on a decision tree. In this case, to expand the study of yield for several plants, a transformed database is used, taking into account the labels entered below:

$$label = \{'potato' : 1, 'maize' : 2, 'barley' : 3, 'blackcurrant' : 4, 'apricot' : 5, 'alfalfa' : 6, 'apple' : 7, 'pear' : 8, 'cherry' : 9, 'corn' : 10\} \quad (7)$$

We derive the equation and the calculated data of the free term and coefficients in the study of multiple linear regression for the database under study. The equation of multiple linear regression of yield from independent signs of fertilizers, temperature, humidity, soil acidity and precipitation with updated data based on the coefficients obtained has the form:

Table 1. Performance of the main machine learning algorithms

Estimates / ML Algorithm	MAE	R^2 Score	RMSE
Lasso Regression - all parameters	0,7237154	0,0351341	0,8865015
Lasso Regression - selected parameters	0,0393173	0,9980028	0,0403324
SVR – all parameters	0,6776377	0,1411921	0,6776377
SVR - selected parameters	0,05250301	0,9958232	0,0583266
Random Forest Regression – selected parameters	0,0003116	0,9999996	0,0005596
Random Forest Regression – all parameters	0,6847263	0,0750289	0,8679807
Gradient Descent Algorithm – all parameters	0,7329265	0,0047446	0,9003539
Gradient Descent Algorithm – selected parameters	0,5885566	0,3576148	0,7233421

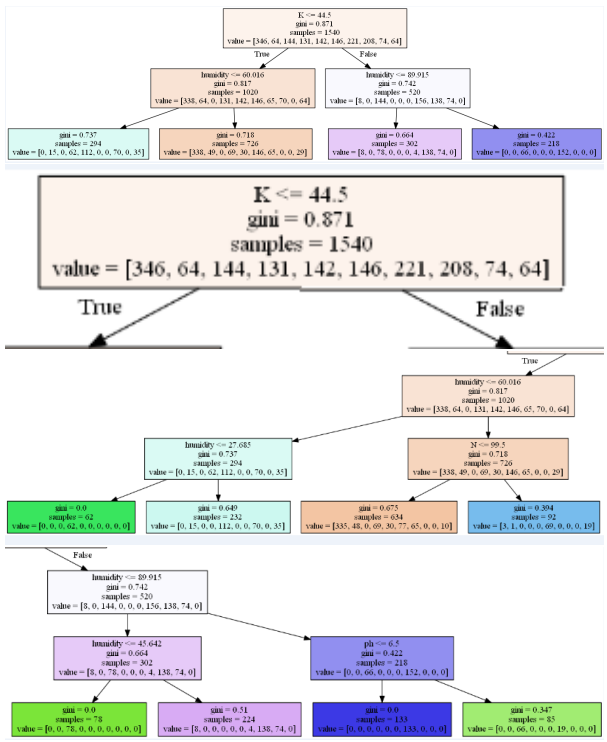

Figure 8. Graphical representation of the model using a decision tree when 1) max depth=2, 2) max depth=3

$$harvest = 0.31813690723082944 + 0.08049244 * N - 0.11737906 * P + 0.01466049 * K + 0.10570846 * t°C - 0.06511367 * humidity - 0.01802083 * ph + 0.01505539 * rainfall \quad (8)$$

Equation (7) is the simplest regression equation — the yield model. Now let's calculate the estimates of the multiple regression prediction model in various metrics.

$$\begin{aligned} &MeanAbsoluteError(MAE) = 0.116981, \\ &MeanSquareError(MSE) = 0.020748, \\ &RootMeanSquareError(RMSE) = 0.144041, \end{aligned} \quad (9)$$

Now let us see the results of using more advanced machine learning algorithms. Below are the prediction results, in the form, which are obtained using the following four basic algorithms with parameter adjustments.

From Table 1, we can see that all MAE are close to zero. Ideally, we should have MAE equal to zero. Some of the database prediction results from Table 1 are shown in Figures 9 and 10. The results and comparative analysis of the performance of models for potato yield showed that the accuracy of the gradient descent algorithm is lower than other algorithms. In the calculations given below, various machine learning technologies are used everywhere for the retrained models. Below are the results of using machine learning algorithms as components of an ensemble to study regression models, as well as their accuracy estimates based on yield prediction. The obtained results of yield regression using machine learning algorithms are shown in Table 2, and the visualization of yield calculations is shown in the figures (Figures 9 and 10). Table 2 shows that gradient boosting algorithms with MAPE =10.14% and random forest with MAPE =10.19% give good results. The method of support vectors MAPE =10.12% turned out to be the leader of the forecast.

A significant role, in all algorithms, for evaluating the accuracy of the model is played by the choice of algorithm parameters. Cases with a choice of all parameters and with a partial choice of parameters are considered. The analysis of the obtained calculations carried out using a machine learning algorithm: the support vector machine method, Lasso regression showed satisfactory results. The results of yield forecasting showed that Random Forest, Lasso regression and SVR algorithms with selected coefficients are the most accurate forecasts. When using ensemble algorithms, or as it is called Lazy Prediction, Lazy Prediction for crop yield forecasting tasks with several ensemble components, advanced results are obtained in the form of figure11. It is appropriate to note here that the constituent algorithms can be any. Below, using this library with a special selection of

```
 21%|██        | 9/42 [00:02<00:10,  3.12it/s]GammaRegressor model failed to execute
Some value(s) of y are out of the valid range for family GammaDistribution
 76%|███████   | 32/42 [00:19<00:06,  1.55it/s]PoissonRegressor model failed to execute
Some value(s) of y are out of the valid range for family PoissonDistribution
100%|██████████| 42/42 [00:26<00:00,  1.61it/s]
```

Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
NuSVR	0.08	0.14	0.84	1.58
SVR	0.07	0.13	0.84	1.68
HistGradientBoostingRegressor	0.06	0.13	0.84	0.62
LGBMRegressor	0.06	0.13	0.84	0.20
RandomForestRegressor	0.03	0.10	0.86	3.61
MLPRegressor	0.02	0.09	0.86	5.50
BaggingRegressor	0.02	0.09	0.86	0.38
ExtraTreesRegressor	-0.01	0.06	0.87	1.35
HuberRegressor	-0.04	0.04	0.89	0.05
SGDRegressor	-0.04	0.04	0.89	0.03

Figure 10. Results of applying the ensemble method to yield prediction problems.

Table 2. Results of model evaluations obtained by machine learning algorithms

Estimates / ML Algorithm	R^2	MAE	MSE	RMSE	MAX	MAPE in %
Linear Regression	0.01	2	8	3	13	11.06
Decision Tree Regression	-0.61	3	12	4	16	13.49
Stochostic Gradient Descent Regression	0	2	8	3	13	11.12
K – Nearest Neighbour 5	0.03	2	8	3	12	10.58
SVR	0.1	2	9	3	13	10.12
Gradient Boosting Regression	0.12	2	7	3	12	10.14

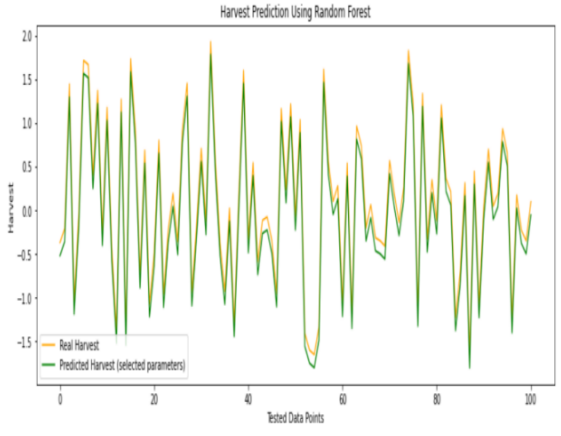

Figure 9. Performance analysis of the random forest algorithm for yield prediction

machine learning algorithm parameters, we calculated the performance of the 10 most powerful known classifications or regression models on several performance matrices. Calculations on ensemble algorithms for regression problems, which are given below figure11 calculated in Google Colaboratory.

4 Conclusion

The paper shows the effectiveness of using machine learning algorithms in modeling and forecasting agricultural tasks. Studies on the initial data for five districts of the Issyk-Kul region on the yield of potatoes and other crops, using machine learning, have shown the effectiveness of identifying the main characteristics for building models and forecasts. An important process in forecasting yields is investigated - the outflow of farmers, introduced by the authors, as one of the important factors in forecasting yields. The strength of forecasting the outflow of farmers is the identification of the potential for growing crops — a human resource. Another important factor studied by the authors is the fight against weeds for the entire growing season of the plant. The problem of classification of weed plants by plant type into annual, perennial and parasitic based on the length and width of the leaves of the plant is considered. It is shown that annual weeds are the main influencing factor for root-bearing plants. The results are obtained in the form of equations (5), (7) of multiple regression of yield from independent factors and their evaluation (8). Table 1 shows the calculations constructed using machine learning algorithms, which gave the results of the deviation of accurate and forecast data based on regression models. Fig. 6 and figure 7 summarize the results of applying machine learning algorithms for regression problems, using the example of a random forest as a graphical representation as an indicator of potato yield for the region as a whole. In this case, the intensity of data accumulation around the regression line, as we see, is the greatest. As can be seen from Table 2, the best performance for predicting yield was obtained when evaluating accuracy, the average absolute error of MAPE = 10.19% for the random forest algorithm. Other algorithms such as the support vector machine, gradient boosting, and the nearest neighbor method are close with MAPE values to the result of a random forest. Figure 11 shows the results of applying the ensemble method to yield forecasting problems, here the results for the ten best ensemble models are selected and obtained. The yields of the most common crops that are grown in these areas are also studied and a model is obtained using a decision tree with different choices of tree parameters and their depths Figure 8. On Figure 9 using the example of a random forest, the results of model performance are obtained and estimates of model accuracy are given. The performance of the Lasso method algorithm for yield prediction is visualized in Fig.10. The application of advanced machine learning algorithms to the problems of forecasting the yield of various crops, such as the method of support vectors, K –nearest neighbors, gradient boosting options and random forest are shown in Figure 2, Figure 3. For comparative analysis, model accuracy estimates

were compared with the results of multiple regression. The use of machine learning is to identify some hidden features, factors that affect the yield for the selected region. For further study of the region, it is necessary to collect data with a more extended range of factors to include in the model. At the same time, climate change, various accidental natural phenomena such as hail precipitation, a sharp increase in daytime or a decrease in night-time temperatures–frosts, increased solar activity and risks associated with prolonged abnormally hot days in the summer, which destroys the soil leads to erosion of sowing areas, the effect on yield, natural phenomena such as frequent mudflows or the fight against low water, which are not uncommon for this region. All these factors lead to a loss of yield on a colossal scale. According to the authors, these studies should use deep learning technologies in the future, as shown in [Rao et al. 2019], [Taherei-Ghazvinei et al. 2018].

References

Ahamed, A.T.M.S., Mahmood, N.T., Hossain, N., Kabir, M.T., Das, K., Rahman, F., and Rahman, R.M. (2015). Applying data mining techniques to predict the annual yield of major crops and recommend planting different crops in different districts in Bangladesh. *In: 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2015 –proceedings. http://doi.org/10.1109/SNPD.2015.7176185.*.

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

Brown, G. (2017). Ensemble Learning. In: Summit, C., and Webb, G.I. (Eds.). In *Encyclopedia of Machine Learning and Data Mining. Boston, MA: Springer US, pp. 393–402.*

Charoen-Ung, P., and Mittrapiyanuruk, P. (2019). Sugarcane yield grade prediction using random forest with forward feature selection and hyperparameter tuning. *In: IC2IT 2018: Recent Advances in Information and Communication Technology 2018, pp. 33–42.*

Everingham, Y.L., Smyth, C.W., Inman-Bamber N.G. (2009). Ensemble data mining approaches to forecast regional sugarcane crop production. Agricultural and Forest Meteorology. *149, 689–696.*

Filippi P., Jones, E.J., Wimalathunge, N.S., Somarathna, P.D.S.N., Pozza, L.E., Ugbaje, S.U., Jephcott, T.G., Paterson, S.E., Whelan, B.M., and Bishop, T.F.A. (2019). Ensemble data mining approaches to forecast regional sugarcane crop production. Agricultural and Forest Meteorology. *149, 689–696.*

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

Kashnitsky, Y. Open Machine Learning Course. *https://github.com/Yorko/mlcourse.ai. Accessed on 2 May 2022.*

Kitchenham, B., Charters, S., Budgen, D., Brereton, P., Turner, M., Linkman, S., and Visaggio, G. Guidelines for Performing Systematic Literature Reviews in Software Engineering. *https://userpages.unikoblenz.de/ laemmel/esecourse/slides/slr.pdf. Accessed 2 May 2022.*

Liakos, K.G., Busato, P., Moshou, D., Pearson, S., and Bochtis D. Machine learning in agriculture: a review. *Sensors (Switzerland), 18, 2674. https://doi.org/10.3390/s18082674.*

Mola-Yudego, B., Rahlf, J., Astrup, R., and Dimitriou, I. Spatial yield estimates of fast-growing willow plantations for energy based on climatic variables in northern Europe. *GCB Bioenergy, 8, 1093–1105. https://doi.org/10.1111/gcbb.12332.*

Paul, M., Vishwakarma, S.K., and Verma, A. Analysis of soil behavior and prediction of crop yield using data mining approach. *In: 2015 International Conference on Computational Intelligence and Communication Networks (CICN), pp. 766–771. https://doi.org/10.1109/CICN.2015.156.*

Rao, T., and Manasa, S. Artificial Neural Networks for soil quality and crop yield prediction using machine learning. *International Journal on Future Revolution in Computer Science and Communication Engineering, 5, pp. 57–60.*

Sujatha, R., and Isakki, P. A study on crop yield forecasting using classification techniques. *In: 2016 International Conference on Computing Technologies and Data Engineering, ICCTIDE 2016. https://doi.org/10.1109/ICCTIDE.2016.7725357.*

Taheri-Qazvini, P., Hassanpour-Darvishi, H., Mousavi, A., Yusof, K.W., Ali zamir, M., Shamshirband, and Chau, S.K. Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. *In: Engineering Applications of Computational Fluid Mechanics, 12, 738–749. https://doi.org/10.1080/19942060.2018.1526119.*

Van Klompenburg, T., Kassahun, A., and Catal, C. Crop yield prediction using machine learning: A systematic literature review. *In: Computers and Electronics in Agriculture, 177, 105709. https://doi.org/10.1016/j.compag.2020.105709.*

Ying-xue, S., Huang, X., and Li-jiao, Y. Support vector machine-based open crop model (SBOCM): Case of rice production in China. *In: Saudi Journal of Biological Sciences, 24, 537–547. https://doi.org/10.1016/j.sjbs.2017.01.024.*

You, J., Li, X., Low, M., Lobell, D., and Ermon, S. Deep Gaussian process for crop yield prediction based on remote sensing data. *In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), 2559-2565* .