

Integrating numerical methods and machine learning to optimize agricultural land use

Assemgul Tynykulova¹, Ayagoz Mukhanova¹, Ainagul Mukhomedyarova², Zhanar Alimova³,
Nurbolat Tasbolatuly⁴, Ulmeken Smailova⁵, Mira Kaldarova⁶, Marat Tynykulov⁷

¹Department of Information Systems, L. N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan

²Institute of Agricultural Technology, West Kazakhstan Agrarian-Technical University, Uralsk, Republic of Kazakhstan

³Faculty of Computer Science, Toraighyrov University, Pavlodar, Republic of Kazakhstan

⁴Higher School of Information Technology and Engineering, Astana International University, Astana, Republic of Kazakhstan

⁵Center of Excellence of Autonomous Educational Organization Nazarbayev Intellectual Schools, Astana, Republic of Kazakhstan

⁶Higher School of Information Technology and Engineering, Astana International University, Astana, Republic of Kazakhstan

⁷Department of Biotechnology and Microbiology, Faculty of Natural Sciences, L. N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan

Article Info

Article history:

Received Apr 16, 2024

Revised Jun 8, 2024

Accepted Jun 16, 2024

Keywords:

Dimensionality reduction

Factor analysis

Feature selection

Machine learning models

Numerical methods

Principal component analysis

ABSTRACT

In the current context, optimizing the utilization of agricultural land resources is increasingly vital for production intensification. This study presents a methodological approach employing numerical methods and machine learning algorithms to analyze and forecast land use optimality. The objective is to develop effective models and tools facilitating rational and sustainable agricultural land resource management, ultimately enhancing productivity and economic efficiency. The research employs data dimensionality reduction techniques such as principal component analysis and factor analysis (FA) to extract key factors from multidimensional land data. The simplex method is utilized to optimize resource allocation among crops while considering constraints. Machine learning algorithms including extreme gradient boosting (XGBoost), support vector machine (SVM), and light gradient boosting machine (LightGBM) are employed to predict optimal land use and yield with high accuracy and efficiency. Analysis reveals significant differences in model performance, with LightGBM achieving the highest accuracy of 99.98%, followed by XGBoost at 95.99%, and SVM at 43.65%. These findings underscore the importance of selecting appropriate algorithms for agronomic data tasks. The study's outcomes offer valuable insights for formulating agricultural practice recommendations and land management strategies, integrable into decision support systems for the agricultural sector, thereby enhancing productivity and production efficiency.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ayagoz Mukhanova

Department of Information Systems, L. N. Gumilyov Eurasian National University

010000 Astana, Republic of Kazakhstan

Email: ayagoz.mukhanova.83@mail.ru

1. INTRODUCTION

In the modern world, where the planet's population is constantly growing and natural resources are limited, the task of ensuring food security is becoming increasingly urgent. One of the key directions for solving this problem is the optimization of agricultural land use, which involves efficiently using land resources [1]–[3] to increase productivity [4] while simultaneously reducing the negative impact on the environment. Advances in information technology and machine learning are opening up new opportunities

for agriculture, making it possible to analyze large volumes of data to optimize land use decisions. Numerical methods such as principal component analysis (PCA), factor analysis (FA), and simplex method are traditionally used to solve various optimization and data analysis problems. They make it possible to identify relationships and patterns in large data sets, which is especially important in the context of the complexity and multidimensionality of agricultural production. On the other hand, machine learning algorithms, including extreme gradient boosting (XGBoost) [5]–[7], support vector machine (SVM) [8]–[10], and light gradient boosting machine (LightGBM) [11]–[13], demonstrate an outstanding ability to process and analyze data to predict outcomes and automate decision making.

The integration of numerical methods and machine learning is a promising direction in the field of agricultural land use. This approach allows not only to optimize the use of land and water resources, but also to predict yields, analyze risks associated with climate change, plant diseases, and pests, and formulate recommendations for improving agricultural practices. In this article, we present an integrated approach to the analysis and optimization of agricultural land use, based on the consistent application of numerical methods and machine learning algorithms [14], [15]. Starting with the application of PCA to the original data set to reduce dimensionality and identify key variable factors. Then, using FA, we further consolidate the data, identifying hidden factors that can have a significant impact on agricultural productivity. The next step is to apply the Simplex method [16], [17] to determine the optimal allocation of resources in order to maximize yield under given budget constraints. The resulting factors and optimized values are integrated [18], [19] as features to train predictive machine learning models, including LightGBM, XGBoost, and SVM, resulting in highly accurate predictions of crop yields and land use efficiency.

The optimization problem in the agricultural sector is a multidimensional problem that includes various aspects: from climatic conditions and soil conditions to economic indicators and market trends. It is especially important to pay attention to economic parameters, such as the cost of fertilizers and land taxes, which directly affect profitability. Additionally, factors such as the sum of temperatures and rainfall play a key role in determining yield potential. As a result, it is necessary to use analytical methods to process and analyze this data in order to make informed decisions [20]–[22]. At the intersection of data and forecasting, our work explores the potential of a variety of analytical tools and models. The key challenge is not only to extract useful information from a wealth of data but also to create predictive models that can adapt to changing conditions and provide operational recommendations for optimal resource management [23]–[25].

The research focuses on finding the optimal balance between accuracy, speed, and interpretability of results, which is especially important for solving applied problems in agronomy [26]–[28]. We strive to ensure that our models are flexible enough that they can be effectively implemented in real-life agricultural practices, where decisions often need to be made under conditions of uncertainty and time constraints. Thus, our research aims to develop a valid and validated methodology that can serve as a basis for the development of sustainable agricultural systems. We pay special attention not only to the scientific rigor of the study but also to the practical significance of the results obtained to ensure their applicability in a wide range of agricultural conditions. This includes different types of crops, climate zones, and levels of technology in agricultural production [29]–[31]. As a result, the presented approach can serve not only the academic community but also practicing agronomists seeking innovative development of agriculture.

2. METHOD

Integrating numerical methods such as PCA and FA with machine learning provides powerful analytical tools for predicting and understanding data. Instead of using these methods solely to reduce the dimensionality of data, we view them as a way to complement and expand the feature space for training models. PCA and FA allow us to extract additional features that may be relevant for predicting target variables. For example, instead of considering only the original variables, we can add new components extracted using PCA and FA to the model to improve the performance of machine learning models.

PCA searches for linear combinations of input features that most effectively describe the variation in the data. These principal components were used as additional features for training the models. To calculate PC, you first need to calculate the covariance matrix of the original features. Let X be a data matrix where each row represents one observation and each column represents one feature. After calculating the covariance matrix Σ , we calculate its eigenvectors v_i and eigenvalues λ_i . PCs are found as linear combinations of the original features. Let's say we have a data matrix X of size $n \times p$, where n is the number of observations and p is the number of features. Mathematically, PCA is based on finding the eigenvectors and eigenvalues of the covariance matrix of the original data (1). Calculate the covariance matrix Σ :

$$\Sigma = \frac{1}{n} (X - \bar{X})^T (X - \bar{X}) \quad (1)$$

where \bar{X} is the vector of average values of features. We find the eigenvectors v_1, v_2, \dots, v_n and eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ from the (2).

$$\Sigma v_i = \lambda_i v_i \quad (2)$$

Select the k principal components with the largest eigenvalues so that they explain most of the variance in the data. We project the original data X onto a new feature space formed by the principal components. Select the k principal components with the largest eigenvalues so that they explain most of the variance in the data. We project the original data X onto a new feature space formed by the principal components.

When choosing the number of principal components for a model, several factors should be considered. First, the explained variance of the data allows the components to be quantified in such a way as to retain a certain percentage of the total variability, for example, 95%. Second, the Kaiser eigenvalue criterion indicates the need to retain only those components that have eigenvalues greater than one, which helps identify significant components. In addition, the optimal number of components can be determined based on the specific requirements of the model, possibly through experiments on test data. The eigenvectors are ordered in descending order of their respective eigenvalues, where the first principal component $PC1$ explains the most variability in the data, $PC2$ the next largest variability, and so on. Thus, if we want to keep only k principal components, we can select only the first k eigenvectors and calculate only their corresponding principal components, for example, $PC1, PC2, PC3, PC4, PC5$. These principal components can be used in further data analysis or as features to train machine learning models. The factor analysis method is used to identify the latent factors underlying the observed variables. These latent factors explain the correlations between observed variables. Our factor analysis model has m latent factors. Mathematically, FA models the observed variables X as (3).

$$X = \Lambda F + \varepsilon \quad (3)$$

where X is a data matrix of dimension $n \times p$, where n is the number of observations, and p is the number of original features. F is a matrix of latent factors of dimension $n \times 2$ (for two factors), where n is the number of observations. Λ is a $p \times 2$ factor loading matrix containing coefficients for each original feature and each latent factor. ε is an $n \times p$ error matrix representing the random error or noise in the data. Equation (2) describes a linear model in which the original data X is decomposed into the product of a factor loadings matrix Λ and a latent factor matrix F , with the addition of an error matrix ε . This model seeks to minimize the approximation error between the observed data and its predicted values based on the latent factors.

The selection of two latent factors to be calculated in a factor analysis method can be justified on the basis of various factors, such as assumptions about the structure of the data, empirical observations, or analysis of the results of preliminary studies. In this case, the choice of two factors was probably made based on an analysis of data variability and model requirements. The factor analysis method seeks to identify the latent factors that best explain the observed variability in the original data. Estimating a factor analysis model takes into account the overall variability in the data and attempts to separate it into common and unique. The two principal components can be selected to retain the largest portion of the total variance explained by the latent factors. To calculate the values of latent factors using factor analysis, methods for estimating model parameters, such as the least squares method or the maximum likelihood method, are often used. These methods help find values of latent factors that best explain the variability in the original data. For example, the least squares method can be used to estimate model parameters by minimizing the sum of squared residuals between the original data and its predicted values based on latent factors. Therefore, two latent factor scores were calculated based on the factor analysis method to best describe the variability in the original data and provide a more interpretable model.

Machine learning offers a range of techniques that can analyze complex relationships in data and make predictions based on learned patterns. The main advantage of such algorithms is their ability to adapt to new information and improve their predictive abilities as new data is processed. For the agricultural sector, this means the ability to predict yields taking into account changes in climatic conditions and economic parameters. Among the algorithms used, it is worth noting gradient boosting methods such as LightGBM and XGBoost, have proven their effectiveness in processing large data sets with high prediction accuracy. Additionally, SVM is used for classification and regression in cases where the data structure is complex and requires a non-linear approach to class separation. XGBoost is an implementation of the gradient boosting algorithm. It creates an ensemble model by sequentially adding new models to correct errors made by previous models as shown in Figure 1. Each new model corrects errors, making predictions more and more accurate. XGBoost uses gradient descent to minimize the loss function, optimizing the parameters of decision trees at each stage.

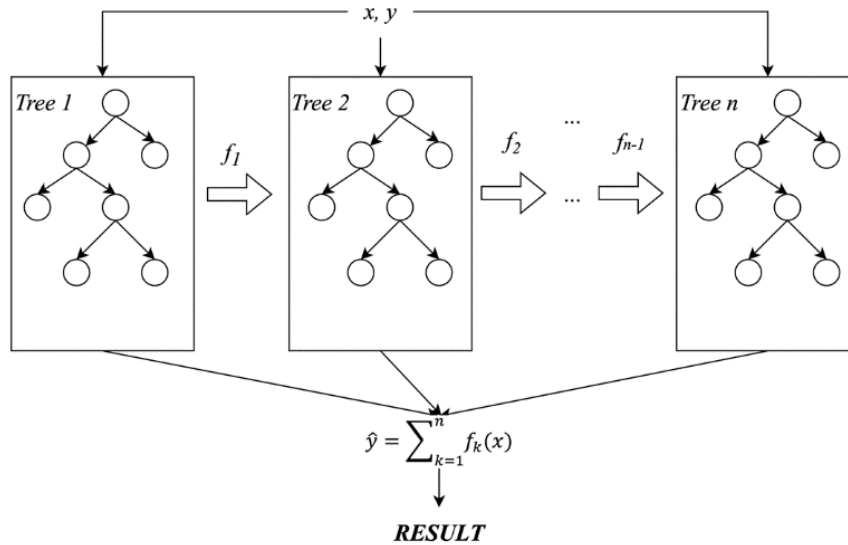


Figure 1. The architecture of the XGBoost method

SVM is designed to solve classification and regression problems. The basic idea is to find a hyperplane in feature space that best divides the data into classes. In the case of linearly inseparable data, SVM uses so-called kernels to move to a higher-dimensional space where the data becomes separable. SVM seeks to maximize the gap between classes by relying on support vectors that represent the most difficult data points to classify as shown in Figure 2.

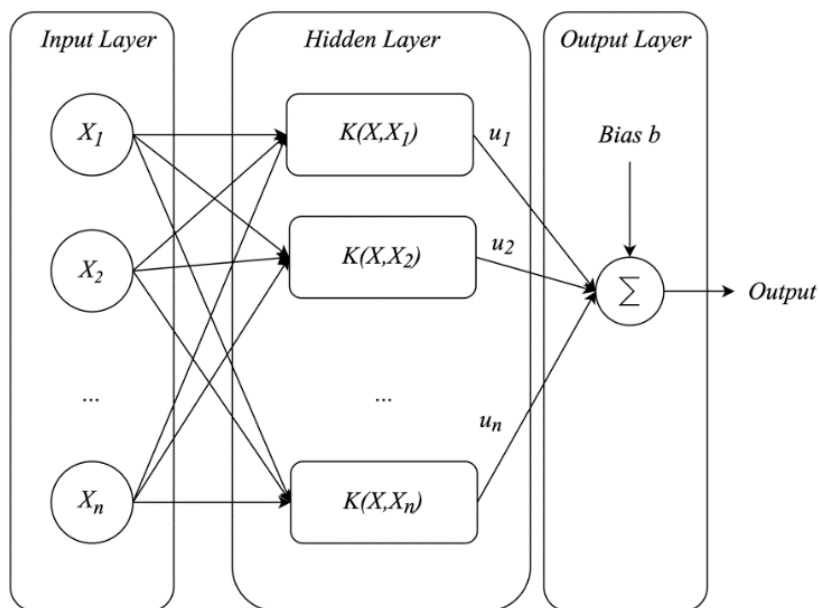


Figure 2. The architecture of the SVM method

Unlike XGBoost, the SVM training process is not iterative in the sense that it does not build the model in stages, adding new models to correct errors. SVM solves the optimization problem to find the best-separating hyperplane immediately. After determining the support vectors and adjusting the model parameters, the training process is completed. LightGBM is a gradient-boosting library developed and maintained by Microsoft. It is an efficient and scalable implementation of gradient-boosting algorithms optimized for working with large data sets and high learning rates in Figure 3.

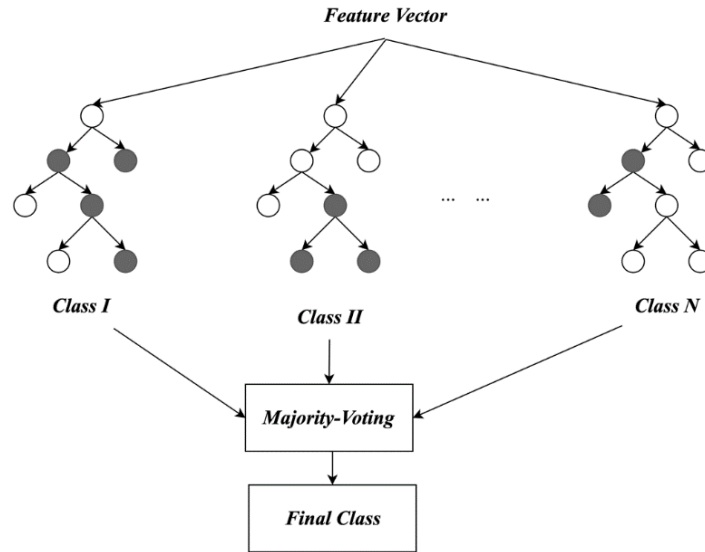


Figure 3. The architecture of the LightGBM method

LightGBM has several key features that contribute to its effectiveness in machine learning tasks. A notable feature is the use of histograms to build decision trees, which speed up the process by replacing traditional data sorting. In addition, LightGBM applies a leaf algorithm for constructing trees, expanding nodes with the largest reduction in the loss function. This results in deeper and more complex trees, potentially increasing model accuracy. The framework automatically processes categorical features without requiring their prior conversion to a numeric format. Particularly important is the high speed of learning and prediction due to the use of histogram methods and leaf partitioning, which is useful when working with large amounts of data. The model provides a wide range of parameters for regularization and optimization, making it easier to fine-tune and prevent overfitting. In addition, LightGBM has built-in support for cross-validation and early stopping, making it easier to select the optimal number of trees and prevent overfitting. Visualization of the learning process, including learning curves and feature importance, further facilitates the analysis and interpretation of simulation results.

3. RESULTS AND DISCUSSION

In this work, to optimize agricultural land use, a model was developed in Figure 4, combining methods of data dimensionality reduction, numerical optimization, and predictive analysis for efficient land use in agriculture. In the first step, we used PCA to transform the original data set and identify the main components influencing land productivity. These components, representing the main drivers of variation, were then added to the data set as new features. The next step was to use factor analysis, which further reduced the dimensionality of the data by identifying latent variables, or factors, that have strong predictive potential. These factors were also included in the model as new features.

In a process of sequential learning based on the provided data, a classification model was first developed to determine the type of crop using features such as the sum of temperatures, precipitation, previous crop, planting area, pH level, sowing date, use of nitrogen, phosphorus, potassium, manure, availability pests, cadastral value, land tax and degree of mineralization. Next, based on these same features plus the predicted crop type from the first model, a regression model was trained to predict the gross yield, thus allowing us to take into account both the initial and received data on the crop type. The final step was training a third regression model to predict costs, where the input data was the original features, the results of previous models, as well as additional analytical data, such as the results of the principal component analysis and factor analysis. This approach to sequential learning made it possible to effectively take into account the relationships between various aspects of agricultural production, significantly increasing the accuracy and quality of forecasts at each stage.

To solve the resource optimization problem, the simplex method was used, the goal of which was to maximize productivity under given budget constraints. Within the framework of agricultural optimization, the goal is usually to maximize economic profit or yield, with an emphasis on increasing gross yield, which is formulated as the problem of maximizing the total gross yield (4).

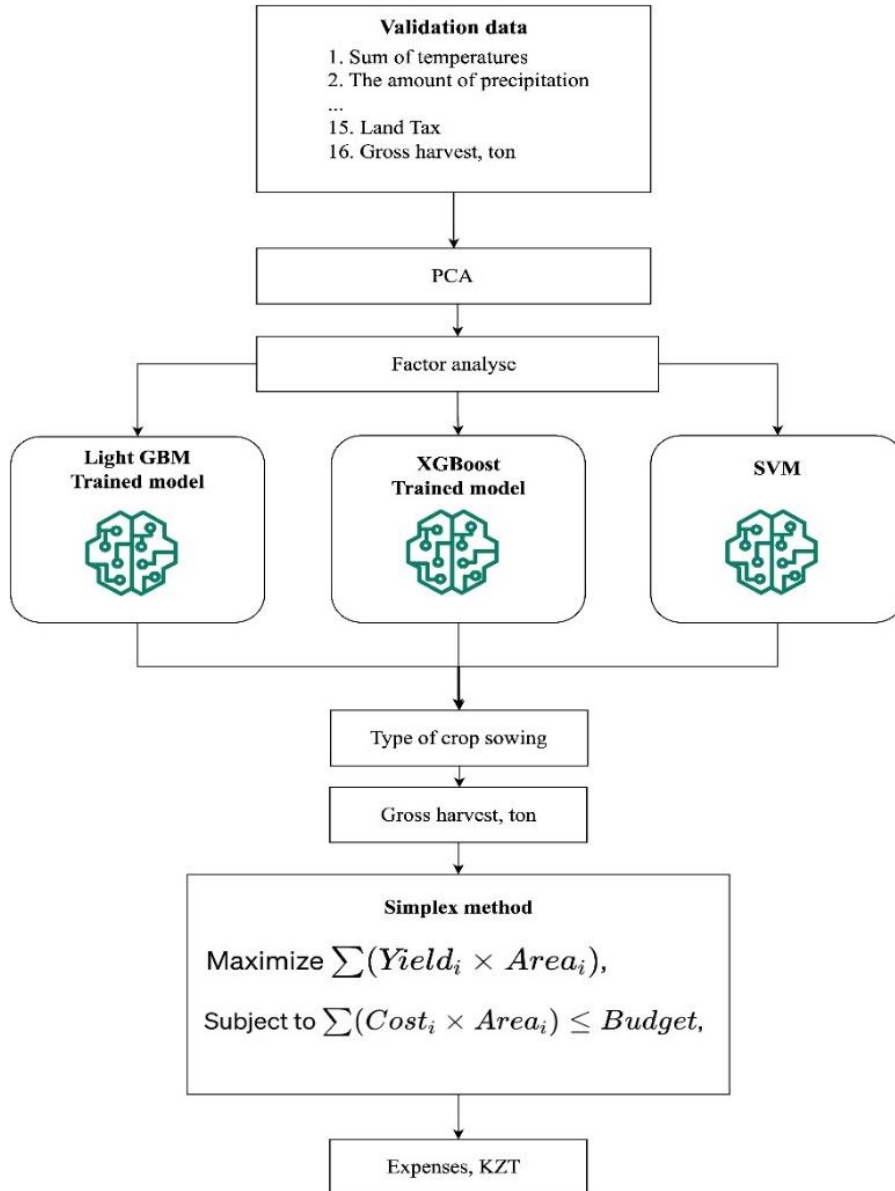


Figure 4. Model for predicting the optimal use of agricultural land

$$\Sigma(Yield_i * Area_i) \tag{4}$$

where *Yield* is the yield per unit *Area_i*, and *Area* is the planted area for crop *i*. Constraints include economic frameworks expressed in terms of (5).

$$\Sigma(Cost_i * Area_i) \leq Budget \tag{5}$$

where *Cost* is the cost of inputs per unit, including the cost of fertilizers, and labor, and other operating costs, and *Budget* is the total available budget, as well as agronomic constraints for the optimal selection of crops that maintain crop rotation and soil fertility, which ensures not only economic efficiency but and promotes the long-term sustainability and productivity of agricultural systems, reflecting integrated goals related to financial benefit, environmental sustainability, and agronomic requirements? The obtained optimal values were integrated into the model in the form of additional features. In the final stage, these features were used to train three different machine learning algorithms: LightGBM, XGBoost, and SVM. Each of these models was trained on a classification or regression task, depending on the land use purpose, and was evaluated based on its accuracy. LightGBM achieved the highest accuracy of 99.98%, showing that it is particularly effective for this data set and task. XGBoost also showed a high accuracy of 95.99%, making it a valuable

tool for predictive modeling in agricultural applications. SVM showed a significantly lower accuracy of 43.65%, which may indicate that this method is less suitable for processing this type of data or requires additional parameter tuning. Thus, an integrated model incorporating numerical methods and machine learning shows significant potential for optimizing agricultural land use, offering tools for more efficient and sustainable management of agricultural resources.

In the course of the research, data on land use and productivity in a certain agricultural area for the period from 2021 to 2023 were analyzed. In 2021, wheat was sown on an area of 18,500 hectares, resulting in a gross harvest of 320,050 tons, at a total cost of 6,000,000 tenge. The next year, 2022, peas were sown on an area of 16,200 hectares, which resulted in a gross harvest of 262,440 tons at a cost of 3,500,000 tenge. In 2023, the analyzed area was again used for growing wheat on an area of 19,370 hectares, which made it possible to harvest a crop of 321,542 tons at a financial cost of 6,500,000 tenge. Based on data for the specified period and the application of our developed model for predicting optimal land use, the following forecasts were made for the 2024 agricultural season. The first class of crops is represented by barley, which, according to the model, demonstrates a success rate of 91.22%, with a projected gross yield estimated at 320,050.03125 tons with costs equal to 5,999,999.5 tenge. The second class is wheat with a probability of successful cultivation of 1.48%, an estimated gross yield of 316,438.90625 tons, and costs of 5,999,309.5 tenge. The third class of crops includes vegetables, where the probability of achieving the desired indicators is 0.98%, and the gross harvest is predicted at 320,053.90625 tons for 5,375,918.5 tenge. Thus, the presented data and model calculations allow us to conclude that there is a high probability of efficient use of land resources when sowing barley in the agronomic season of 2024, taking into account the ratio of expected yield and costs.

Optimal selection of crop passives, aimed at selecting those types of crops that provide the best conditions for the next season, taking into account crop rotation and soil improvement, is a key aspect of agricultural optimization and can be integrated into the optimization model through additional constraints. These limits ensure that agronomic principles of crop rotation are followed to prevent soil degradation and maintain soil fertility, allowing agronomists and agricultural managers to make decisions that balance economic efficiency with long-term sustainability and productivity of the land. This approach requires the formulation of the objective function and constraints in such a way that they reflect the complex goals of agricultural activity, including not only financial benefits but also environmental sustainability and agrotechnical requirements, thereby ensuring integrated and sustainable development of the agro-industrial sector.

4. CONCLUSION

Our research demonstrated how the integration of principal component analysis (PCA), Factor analysis (FA), and machine learning can significantly enhance forecasting and optimization in agricultural land use. A three-year study revealed critical relationships between land use, productivity, and economic performance, forming the foundation for creating a predictive model. These methods simplified multivariate data by identifying underlying components and latent factors, thereby improving the data quality for subsequent analysis. The simplex method was used to find the optimal land use strategy, ensuring a balance between gross yield and costs.

The application of machine learning algorithms, such as LightGBM and XGBoost, showed high accuracy in yield prediction, with LightGBM achieving the highest accuracy of 99.98% and XGBoost at 95.99%. The SVM algorithm, while providing valuable insights into complex data aspects, showed lower accuracy at 43.65%, indicating the need for additional parameter tuning or its lower suitability for this data type. The developed model provides valuable recommendations for agronomic practice and has significant potential for sustainable agricultural sector development, offering tools for more efficient and sustainable management of agricultural resources, ultimately enhancing productivity and economic efficiency.

ACKNOWLEDGEMENTS

This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP19677451).




REFERENCES

- [1] A. Singh, "Judicious and optimal use of water and land resources for long-term agricultural sustainability," *Resources, Conservation & Recycling Advances*, vol. 13, May 2022, doi: 10.1016/j.rcradv.2022.200067.
- [2] A. Sultonov and M. Tursunov, "Problems of optimal use of water resources for crop irrigation," *AIP Conference Proceedings*, vol. 2789, no. 1, 2023, doi: 10.1063/5.0145804.




- [3] M. Li, Q. Fu, V. P. Singh, D. Liu, T. Li, and Y. Zhou, "Managing agricultural water and land resources with tradeoff between economic, environmental, and social considerations: A multi-objective non-linear optimization model under uncertainty," *Agricultural Systems*, vol. 178, Feb. 2020, doi: 10.1016/j.agry.2019.102685.
- [4] M. Yessenova *et al.*, "The effectiveness of methods and algorithms for detecting and isolating factors that negatively affect the growth of crops," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, pp. 1669–1679, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1669-1679.
- [5] J. Tussupov *et al.*, "Analysis of formal concepts for verification of pests and diseases of crops using machine learning methods," *IEEE Access*, vol. 12, pp. 19902–19910, 2024, doi: 10.1109/ACCESS.2024.3361046.
- [6] Y. Meng, N. Yang, Z. Qian, and G. Zhang, "What makes an online review more helpful: an interpretation framework using XGBoost and SHAP values," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 16, no. 3, pp. 466–490, Nov. 2020, doi: 10.3390/jtaer16030029.
- [7] S. S. Azmi and S. Baliga, "An overview of boosting decision tree algorithms utilizing AdaBoost and XGBoost boosting strategies," *International Research Journal of Engineering and Technology*, vol. 7, no. 5, pp. 6867–6870, 2020.
- [8] Q. Wen, Y. Wang, H. Zhang, and Z. Li, "Retraction note: application of ARIMA and SVM mixed model in agricultural management under the background of intellectual agriculture," *Cluster Computing*, vol. 26, no. S1, pp. 167–167, Apr. 2023, doi: 10.1007/s10586-022-03925-4.
- [9] M. Al-Naeem, M. M. Hafizur Rahman, A. Banerjee, and A. Sufian, "Support vector machine-based energy efficient management of UAV locations for aerial monitoring of crops over large agriculture lands," *Sustainability*, vol. 15, no. 8, Apr. 2023, doi: 10.3390/su15086421.
- [10] G. Wang, H. Zheng, and X. Li, "ResNeXt-SVM: a novel strawberry appearance quality identification method based on ResNeXt network and support vector machine," *Journal of Food Measurement and Characterization*, vol. 17, no. 5, pp. 4345–4356, Oct. 2023, doi: 10.1007/s11694-023-01959-9.
- [11] Q. Cao, Y. Wu, J. Yang, and J. Yin, "Greenhouse temperature prediction based on time-series features and LightGBM," *Applied Sciences*, vol. 13, no. 3, Jan. 2023, doi: 10.3390/app13031610.
- [12] J. Feng, Y. Shi, Y. Su, W. Mu, and D. Tian, "Evaluation method of agricultural production technical efficiency based on Borderline-SMOTE and LightGBM," in *3rd International Conference on Management Science and Software Engineering (ICMSSE 2023)*, 2024, pp. 586–592.
- [13] Y. Chen, J. Jia, C. Wu, L. Ramirez-Granada, and G. Li, "Estimation on total phosphorus of agriculture soil in China: a new sight with comparison of model learning methods," *Journal of Soils and Sediments*, vol. 23, no. 2, pp. 998–1007, Feb. 2023, doi: 10.1007/s11368-022-03374-x.
- [14] S. Hazra, S. Karforma, A. Bandyopadhyay, S. Chakraborty, and D. Chakraborty, "Prediction of crop yield using machine learning approaches for agricultural data," *Authorea Preprints*, 2023.
- [15] K. Alimhan, N. Tasbolatuly, and A. Yerdanova, "Global output tracking control for high-order non-linear systems with time-varying delays," *Journal of Theoretical and Applied Information Technology*, vol. 99, no. 13, pp. 3337–3352, 2021.
- [16] K. Alimhan, N. Otsuka, M. N. Kalimoldayev, and N. Tasbolat, "Practical output tracking for a class of uncertain nonlinear time-delay systems via state feedback," *MATEC Web of Conferences*, vol. 189, Aug. 2018, doi: 10.1051/mateconf/201818910027.
- [17] H. Setiawan and N. Hidayat, "Implementation linear programming with method simplex and POM-QM for windows on SMEs in the city of TARAKAN," *Proceeding of The Borneo International Conference of Management, Accounting and Economy*, vol. 1, no. 1, pp. 95–110, 2023.
- [18] S. Saghii, A. Nazemi, S. Effati, and M. Ranjbar, "Simplex algorithm for hesitant fuzzy linear programming problem with hesitant cost coefficient," *Iranian Journal of Fuzzy Systems*, vol. 20, no. 1, pp. 137–152, 2023.
- [19] S. Nahar, M. Akter, and M. A. Alim, "Solving multi-objective linear programming problem by statistical averaging method with the help of fuzzy programming method," *American Journal of Operations Research*, vol. 13, no. 02, pp. 19–32, 2023, doi: 10.4236/ajor.2023.132002.
- [20] A. A. Ismailova, A. K. Zhamangara, S. K. Sagnayeva, G. D. Kaziyeva, A. Abakumov, and S. Y. Park, "Technologies of information monitoring biogens lakes of Kazakhstan," *News of the National Academy of Sciences of the Republic of Kazakhstan - Series of Geology and Technical Sciences*, no. 4, pp. 69–73, 2018.
- [21] E. Popova *et al.*, "Sustainable production of Ajuga bioactive metabolites using cell culture technologies: a review," *Nutrients*, vol. 15, no. 5, Mar. 2023, doi: 10.3390/nu15051246.
- [22] G. Taubayev *et al.*, "Machine learning algorithms and classification of textures," *Journal of Theoretical and Applied Information Technology*, vol. 98, no. 23, pp. 3854–3866, 2020.
- [23] J. Tussupov *et al.*, "Applying machine learning to improve a texture type image," *Eastern-European Journal of Enterprise Technologies*, vol. 122, no. 23, 2023.
- [24] M. Yessenova *et al.*, "Identification of factors that negatively affect the growth of agricultural crops by methods of orthogonal transformations," *Eastern-European Journal of Enterprise Technologies*, vol. 3, no. 2 (117), pp. 39–47, Jun. 2022, doi: 10.15587/1729-4061.2022.257431.
- [25] M. Yessenova *et al.*, "The applicability of informative textural features for the detection of factors negatively influencing the growth of wheat on aerial images," *Eastern-European Journal of Enterprise Technologies*, vol. 4, no. 2 (118), pp. 51–58, Aug. 2022, doi: 10.15587/1729-4061.2022.263433.
- [26] M. Yessenova *et al.*, "Features of growth of agricultural crops and factors negatively affecting their growth," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 1, pp. 625–632, Apr. 2023, doi: 10.11591/ijeecs.v30.i1.pp625-632.
- [27] M. Nenov, I. Dimitrov, V. Lozanova, I. Gerasimova, and N. Traikov, "Optimization of agrotechnical units in technologies for grain production in the Sofia area," *Bulgarian Journal of Soil Science, Agrochemistry and Ecology*, 2020, vol. 54, no. 3, pp. 28–40.
- [28] S. Pospelov and K. Panchenko, "Optimization of agrotechnical practices of growing plants of the Malva genus," *Agrology*, vol. 5, no. 4, pp. 110–115, 2022.
- [29] D. T. Jumanov, J. B. Kuziboyev, and L. A. Izzatullayev, "The impact of agrotechnical factors on the productivity of cotton in the context of optimization," *BIO Web of Conferences*, vol. 71, Nov. 2023, doi: 10.1051/bioconf/20237101078.
- [30] R. I. Mukhamediev *et al.*, "Coverage path planning optimization of heterogeneous UAVs group for precision agriculture," *IEEE Access*, vol. 11, pp. 5789–5803, 2023, doi: 10.1109/ACCESS.2023.3235207.
- [31] A. Rath and P. C. Swain, "Optimal allocation of agricultural land for crop planning in Hirakud canal command area using swarm intelligence techniques," *ISH Journal of Hydraulic Engineering*, vol. 27, no. 1, pp. 38–50, Jan. 2021, doi: 10.1080/09715010.2018.1508375.

BIOGRAPHIES OF AUTHORS






Assemgul Tynykulova    has teaching experience at leading universities in Kazakhstan. The main research interests of the master Tynykulova A.S. are optimization methods, decision-making, expert systems. Over the past 5 years, he has published 10 scientific articles, including 1 in a journal included in the Scopus database. Currently, master of science Tynykulova A.S. works as a senior lecturer at the Higher School of Information Technology and Engineering Astana International University, Astana. She can be contacted at email: asem_110981@mail.ru.






Ayagoz Mukhanova    received her PhD in 2015 in information systems from L.N. Gumilyov Eurasian National University, Kazakhstan. Currently, she is an associate professor of the Department of Information Systems at the same university. Her research interests include artificial intelligence and decision making. She can be contacted at email: ayagoz198302@mail.ru.






Ainagul Mukhomedyarova    PhD agricultural sciences Ainagul Mukhomedyarova has 20 years of teaching experience in Zhangir Khan West Kazakhstan Agrarian-Technical University. The main scientific interests of PhD agricultural sciences A. Mukhomedyarova are in the fields of agriculture, crop production, farming, agrochemistry and forestry. Over the last 5 years she has published 9 scientific articles, including 1 in journals included in Scopus database. Currently, A. Mukhomedyarova works as an acting associate professor, PhD in agricultural sciences at the West Kazakhstan Agrarian-Technical University. She can be contacted at email: aina25111980@mail.ru.






Zhanar Alimova    has more than 20 years of university teaching experience. Her main scientific interests are in the field of information systems, fuzzy logic and chain of cause-effect relationships in data mining. For the last 5 years she has published more than 20 scientific papers, including 3 in journals included in Scopus database, 2 textbooks and 1 certificate of state registration of intellectual property rights (computer program). Currently, Zh. Alimova is a postdoctoral researcher at Toraighyrov University, Pavlodar, Kazakhstan. She is the head of a grant research project. She can be contacted at email: alimovazhanars@gmail.com.






Nurbolat Tasbolatuly    in 2020 defended his doctoral thesis in computing engineering and software at Al-Farabi Kazakh National University. From 2012 to 2020, he worked as senior researcher at the Institute of Information and Computational Technologies in Almaty, Kazakhstan. Since December 2020, he has been working at the Higher School of Information technologies and engineering of the Astana International University as deputy dean. His main research interests include the theory of nonlinear control and computer simulation. He has more than 30 scientific articles, including 6 works in the rating publications Web of Science and Scopus, H-index – 2. He can be contacted at email: tasbolatuly@gmail.com.






Ulmeken Smailova    is an experienced senior manager of the “Center of Excellence” of the Autonomous Educational Organization “Nazarbayev Intellectual Schools”, located on Hussein bin Talal Street in Astana, Kazakhstan (postal code: 010000). Candidate of Physical and Mathematical Sciences, 13/05/16 - Application of computer technology, mathematical research and mathematical methods in scientific research. Associate Professor in the specialty 05.13.00 - “Informatics, computer technology and management”. Associate Professor in the specialty 05.13.00 - “Informatics, computer technology and management”. Has more than 70 scientific and methodological articles. She can be contacted by email samilova_tarsu@mail.ru.



Mira Kaldarova    graduated in 2023 from the doctoral program in the educational programs - D094 Information technology from Department of Computer Engineering and Software Seifullin Kazakh Agro Technical Research University. Currently, she is a senior lecture, Higher School of Information Technology and Engineering Astana International University Kabanbay Batyra ave, 8, Astana, Republic of Kazakhstan. Her research interests include artificial intelligence, machine learning, mathematical modeling and data mining. She can be contacted at email: kmiraj8206@gmail.com.



Marat Tynykulov    Candidate of agricultural sciences, associate professor M.K. Tynykulov has teaching experience at leading universities in Kazakhstan. Scientific interests: agricultural industry, agroecology, technology for the production of biofertilizers and biofuels. Over the past 5 years, he has published 30 scientific articles, including 6 in a journal included in the Scopus database. Currently, candidate of agricultural sciences Tynykulov M.K. works as an associate professor at the LN Gumilyov Eurasian National University, Astana. She can be contacted at email: tynykulov@list.ru.