IMPROVING AGRICULTURAL YIELDS IN THE DEMOCRATIC REPUBLIC OF CONGO USING MACHINE LEARNING ALGORITHMS

Nsimba Malumba Rodolphe[1], Longo Kayembe Mardochee[2], Balanganayi Kabutakapua Fiston Chrisnovic[3], Boluma Mangata Bopatriciat[4*], Mazambi Kilongo Trésor[5], Tabiaki Tandele Rufin[6], Ntanyungu Ndizieye Emmanuel[7], Bukanga Christian Parfum[8]

Computer Science[1]
Institute National du Batiment et des Travaux Publics, Kinshasa, Democratic Republic of Congo[2]
https://inbtp.optsolution.net[1]
rodolphemalumba25@gmail.com[1]

Computer Science[2,3]
University of Kinshasa, Kinshasa, Democratic Republic of Congo[2,3]
https://www.unikin.ac.cd/[2,3]
mardochee.longo@unikin.ac.cd[2], fistonbalang@gmail.com[3]

Computer Science[4*, 8]
Haute Ecole de Commerce de Kinshasa, Kinshasa, Democratic Republic of Congo[4,* 8]
https://heckin.ac.cd/etudiant/[4*, 8]
bopatriciat.boluma@unikin.ac.cd[4*], parfum.bukanga@unikin.ac.cd[8]

Computer Science[5,6]
University of Bunia, Bunia, Democratic Republic of Congo[5,6]
https://www.unibu.ac.cd/[5,6]
mazambitresor@outlook.com[5], rufintandele@gmail.com[6]

Computer Science[7]
Institut Supérieur de Commerce de Bunia, Bunia, Democratic Republic of Congo[7]
https://www.unibu.ac.cd/[7]
bopatriciat@gmail.com[7]

Computer Science[8]
Haute Ecole de Commerce de Kinshasa, Kinshasa, Democratic Republic of Congo[8]
https://heckin.ac.cd/etudiant/[8]
(*) Corresponding Author

Abstract—*This article presents an analysis of agricultural yields in the Democratic Republic of Congo (DRC) using machine learning algorithms. The study is based on around 30,000 records covering several years of agricultural production. Each record includes variables such as seed type, climatic conditions (temperature, rainfall and humidity), soil characteristics (pH, nutrients), farming practices (fertilizer use, irrigation) and yields obtained. The data comes from a variety of sources, including METTELSAT, the World Meteorological Organization (WMO) and WorldClim for climate data, and the DRC Ministry of Agriculture and the FAO for soil and agricultural data. The algorithms evaluated include linear regression, random forest regression, Gradient Boosting Machines (GBM), Support Vector Machines (SVM), and Artificial Neural Networks (ANN). The performance of the algorithms is measured using metrics such as MSE, MAE, RMSE, R^2 Score and MAPE on three separate case studies (Farm A, Farm B and Farm C). The results show that artificial neural networks (ANNs) perform best, with MSE ranging from 600 to 850, MAE from 12 to 17, RMSE from 24.49 to 29.15, R^2 Score from 0.92 to 0.95, and MAPE from 8.5% to 10.7%. Next came GBM, random forest regression, SVM and finally linear regression. These results highlight the potential of machine learning algorithms to improve agricultural yield forecasts in the DRC.*

Keywords: *agricultural yields, climatic data, machine learning, , predictive analysis, regression models.*

Intisari—Artikel ini menyajikan analisis hasil pertanian di Republik Demokratik Kongo (RDK) dengan menggunakan algoritme pembelajaran mesin. Penelitian ini didasarkan pada sekitar 30.000 catatan yang mencakup beberapa tahun produksi pertanian. Setiap catatan mencakup variabel-variabel seperti jenis benih, kondisi iklim (suhu, curah hujan, dan kelembapan), karakteristik tanah (pH, nutrisi), praktik pertanian (penggunaan pupuk, irigasi), dan hasil panen yang diperoleh. Data tersebut berasal dari berbagai sumber, termasuk METTELSAT, Organisasi Meteorologi Dunia (WMO) dan WorldClim untuk data iklim, serta Kementerian Pertanian Republik Demokratik Kongo dan FAO untuk data tanah dan pertanian. Algoritme yang dievaluasi meliputi regresi linier, regresi hutan acak, Gradient Boosting Machines (GBM), Support Vector Machines (SVM), dan Jaringan Syaraf Tiruan (JST). Kinerja algoritma diukur dengan menggunakan metrik seperti MSE, MAE, RMSE, R^2 Score dan MAPE pada tiga studi kasus yang berbeda (Peternakan A, Peternakan B, dan Peternakan C). Hasilnya menunjukkan bahwa jaringan syaraf tiruan (JST) memiliki kinerja terbaik, dengan MSE berkisar antara 600 hingga 850, MAE dari 12 hingga 17, RMSE dari 24,49 hingga 29,15, R^2 Score dari 0,92 hingga 0,95, dan MAPE dari 8,5% hingga 10,7%. Berikutnya adalah GBM, regresi hutan acak, SVM, dan akhirnya regresi linier. Hasil ini menyoroti potensi algoritma pembelajaran mesin untuk meningkatkan prakiraan hasil pertanian di Republik Demokratik Kongo.

Kata Kunci: *hasil pertanian, data iklim, pembelajaran mesin, , analisis prediktif, model regresi.*

INTRODUCTION

Agriculture is a crucial pillar of the Democratic Republic of Congo (DRC) economy, providing livelihoods for a significant portion of the population. However, the agricultural sector faces multiple challenges, including unpredictable yields due to climatic variations, soil characteristics, and farming practices. These challenges hinder food security and economic stability, necessitating innovative solutions. The integration of machine learning algorithms offers a promising approach to improving crop yield forecasts and optimizing farming practices, helping to mitigate the effects of climate variability and resource limitations (*République Démocratique Du Congo - Vue d'ensemble*, n.d.).

The DRC exhibits diverse climatic and geographical conditions, ranging from humid equatorial zones to savannahs and mountainous regions. This diversity enables the cultivation of various crops, including cassava, maize, rice, groundnuts, and coffee. However, agricultural productivity remains constrained by several factors. Climatic variability, characterized by irregular rainfall patterns and extreme temperatures, significantly affects crop cycles and yields. Soil characteristics also play a crucial role, with variations in pH and nutrient composition influencing productivity. Additionally, farming practices, such as fertilizer application, irrigation, and seed selection, impact yield outcomes. Insights from climate data sources such as METTELSAT, WMO, and WorldClim, as well as soil and agricultural data from the Ministry of Agriculture and FAO, are essential for developing adaptive farming strategies (Bal & Kayaalp, 2021).

Despite the DRC's overall economic growth, primarily driven by the mining sector, the agricultural sector has shown slower and more modest growth. According to the World Bank, agricultural production slowed from 2.4% in 2022 to 2.2% in 2023, highlighting the need for stronger sectoral support (*République Démocratique Du Congo - Vue d'ensemble*, n.d.). While real GDP experienced significant growth, agricultural production lagged behind, emphasizing the importance of adopting innovative technologies to enhance productivity. The following Table 1 illustrates the trends in GDP growth and agricultural production (2020-2024):

Table 1. Growth in GDP and agricultural production

Year	Real GDP growth (%)	Growth in agricultural production (%)
2020	1,7	Data not available
2021	6,2	Data not available
2022	8,9	2,4
2023	8,4	2,2
2024	4,9	Data not available

Source: (*République Démocratique Du Congo - Vue d'ensemble*, n.d, 2020-2024)

The figures show that, although overall GDP grew significantly, agricultural production grew at a slower rate, with a slight deceleration observed in 2023. This trend highlights the need to adopt innovative approaches, such as machine learning, to improve yield forecasts and optimize agricultural practices. This study utilizes a comprehensive dataset comprising approximately 30,000 records covering multiple years of agricultural production in the DRC. The dataset includes variables such as seed type, climatic conditions (temperature, rainfall, humidity), soil characteristics (pH, nutrients), farming practices (fertilizer use, irrigation), and obtained yields. These data are sourced from institutions such as METTELSAT, WMO, WorldClim, the Ministry of Agriculture, FAO, farmer surveys, and local farming organization databases (Ağbulut et al., 2021). The primary objective of this research is to evaluate the performance of various machine learning algorithms in forecasting agricultural yields. Algorithms such as Linear Regression, Random Forest Regression, Gradient Boosting Machines (GBM), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) are assessed using key performance metrics, including Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Coefficient of Determination (R^2), and Mean Absolute Percentage Error (MAPE) (Huang et al., 2020)(Sekeroglu et al., 2020).

Integrating machine learning algorithms into Congolese agriculture would improve the accuracy of yield forecasts, thereby facilitating agricultural planning. The results of this study showed that learning models, in particular gradient boosting, offer better performance by reducing prediction error and increasing the reliability of estimates. To optimize the management of agricultural resources, it would be appropriate to set up a monitoring and early warning system based on artificial intelligence. This system could provide farmers with real-time information on weather conditions, soil characteristics, and yield trends. By combining this data with tailored recommendations, farmers could adjust their farming practices to maximize productivity.

Rational use of resources also involves optimizing irrigation and fertilizers, thanks to predictive models capable of recommending the optimum quantities to use. In addition, the adoption of precision farming, in particular using sensors and drones, could enable effective monitoring of crops and rapid intervention in the event of water stress or nutritional deficiencies. In addition, the creation of a National Agricultural Information System (SIAN) would facilitate the centralization and dissemination of agricultural data. This platform would give farmers access to climate forecasts, agronomic recommendations, and information on market prices, thereby contributing to informed decision-making.

Another lever for improvement lies in building farmers' capacities through training in new technologies and sustainable farming practices. Raising farmers' awareness of sustainable soil management and the adoption of techniques that are resilient to climate change would increase productivity while preserving natural resources. The adoption of agroecology and agroforestry could also play a key role in the sustainability of the agricultural sector in the DRC. Diversifying crops, rotating plantations, and integrating trees into farms would help preserve soil fertility and limit erosion. Finally, modernizing agricultural infrastructure, in particular improving irrigation systems and logistics chains, would help to reduce post-harvest losses and ensure better marketing of agricultural produce. Exploiting emerging technologies, such as the Internet of Things (IoT) and satellite data analysis, would also offer innovative solutions for monitoring farmland and optimizing production.

By leveraging technological advancements, this research aims to contribute to improving food security and ensuring the sustainable management of agricultural resources in the DRC. The methodologies and findings presented also serve as a foundation for future studies on applying machine learning techniques in African agriculture, demonstrating the potential of AI-driven solutions to enhance agricultural resilience and productivity (Mouhssine & Otman, 2020).

MATERIALS AND METHODS

Agriculture is a key sector for the economy of the Democratic Republic of Congo (DRC), providing livelihoods for a large proportion of the population. Optimizing agricultural yields is therefore crucial to ensuring food security and promoting economic development. The use of machine learning algorithms to analyze agricultural yields offers a promising approach for identifying key factors influencing production and predicting future yields (Jha et al., 2021). This paper proposes a comprehensive and rigorous methodology for analyzing agricultural yields in the DRC using machine learning techniques, based on a diverse and rich dataset.

Data Collection

Data quality and diversity are essential for the success of any machine learning analysis. In this study, we collected a wide range of data covering several years of agricultural production in the DRC.
Data sources

1. Climatic data:
   Agence Nationale de Météorologie et des Télécommunications par Satellite (METTELSAT): Detailed local data on temperature, rainfall and humidity.
   World Meteorological Organization (WMO): Global and standardised climate data.
   WorldClim: Open source climate databases offering high-resolution data.
2. Soil data:
   DRC Ministry of Agriculture: Information on soil pH and nutrient levels.

Vol. 22, No. 1 March 2025 | DOI https://doi.org/10.33480/techno.v22i1.6380

FAO: Data on soil quality and land management practices.
Local field studies and geospatial databases: Specific and detailed soil information.

3. Agricultural data
DRC Ministry of Agriculture: Historical data on crop yields and cultivation practices.
Local farming organizations: Additional information from databases.
Farmer surveys: Data collected directly from farmers.

Data pre-processing

Before proceeding with analysis, it is crucial to prepare the data to ensure its quality and relevance. Data pre-processing includes cleaning, normalizing and transforming raw data into a format that can be used by machine learning algorithms. Pre-processing steps:

1. Data cleaning
Outliers: Identification and removal of non-representative extreme values.
Missing data: Imputation of missing values using techniques such as mean, median or interpolation.
2. Normalization and Transformation
Standardization of continuous variables to ensure uniform scaling.
Coding of categorical variables (seed type, farming practices) using techniques such as one-hot coding.
3. Data splitting
Separation of data into training (80%) and test (20%) sets for robust model evaluation.

Selection and Training of Machine Learning Algorithms

In order to analyze agricultural yields, several machine learning algorithms will be used and compared to determine the one offering the best predictive performance (Khanam & Foo, 2021) and (Benos et al., 2021). In this paper, we explore the use of various machine learning algorithms to analyze agricultural yields in the Democratic Republic of Congo (DRC). The algorithms studied include Linear Regression, Random Forest Regression, Gradient Boosting Machines (GBM), Support Vector Machines (SVM) and Artificial Neural Networks (ANN) (Sharifi, 2021).

1. Linear regression
Linear regression is a simple and commonly used algorithm for modelling the relationship between a dependent variable and one or more independent variables.
2. Random Forest Regression
Random Forest Regression is an ensemble algorithm that uses multiple decision trees to improve accuracy and control over-fitting. It is based on the aggregation of predictions from several decision trees (van Klompenburg et al., 2020).
3. Gradient Boosting Machines (GBM)
The GBM is an ensemble algorithm that creates a powerful predictive model by combining several weak models (such as decision trees) sequentially. Each new model corrects the errors made by the previous model (Mourtzinis, S., Esker, P. D., Specht, J. E., & Conley, 2021).
4. Support Vector Machines (SVM)
SVMs are supervised algorithms that can be used for classification or regression. In regression, the SVM attempts to find a hyperplane in a multi-dimensional space that best predicts the target variable (Raman et al., 2024).
5. Artificial Neural Networks (ANNs)
ANNs are inspired by biological neural networks and are composed of layers of artificial neurons. They are particularly powerful for modelling complex, non-linear relationships between variables (Gunjan et al., 2022).

Model Training

1. Cross Validation :
Use of cross-validation (k-fold) to optimize hyperparameters and evaluate model performance.
2. Optimization Techniques:
Application of specific techniques for each algorithm, such as Grid Search and Random Search, to find the best hyperparameters.

Model Evaluation

The evaluation of models is a crucial step in determining their performance and reliability. Several metrics will be used to evaluate models on specific case studies (Gunjan et al., 2022). Each of these algorithms is evaluated using specific metrics to measure their performance.

1. Mean Squared Error (MSE)
The MSE measures the mean squared error, i.e. the difference between the predicted values and the actual values.
2. Mean Absolute Error (MAE)
The MAE measures the mean absolute error.
3. Root Mean Squared Error (RMSE)
The RMSE is the square root of the MSE and gives an idea of the magnitude of the errors.
4. Coefficient of Determination (R^2)
The R^2 measures the proportion of the variance of the dependent variable that is explained by the independent variables.
5. Mean Absolute Percentage Error (MAPE)
MAPE measures percentage accuracy.
6. Confusion matrices
Confusion matrices are mainly used to evaluate classification models. However, we can interpret

results for regression problems in a similar way by comparing predictions with actual values in a table. Case studies. The performance of the models will be evaluated on three separate farms (Farm A, Farm B, Farm C) to ensure the robustness and generalizability of the results.

In this paper, we study agricultural yields in the Democratic Republic of Congo (DRC) using machine learning algorithms. The data comprises around 30,000 records, containing various variables such as seed type, climatic conditions, soil characteristics, farming practices, and yields obtained. Data sources include the Agence Nationale de Météorologie et des Télécommunications par Satellite (METTELSAT), the World Meteorological Organization (WMO), the Ministry of Agriculture, the FAO, local field studies, and farmer surveys (Pant et al., 2021).

RESULTS AND INTERPRETATION

The use of artificial intelligence and machine learning algorithms to optimize agricultural yields is a promising avenue for improving agricultural productivity in the Democratic Republic of Congo (DRC). This study analyses agricultural yields using various machine learning algorithms. The results obtained are presented in terms of model performance for several case studies, enabling robust conclusions to be drawn that can be applied in the field.

Data sources and size

For accurate and reliable analysis, a large and diverse database is essential. This study uses around 30,000 records covering several years of agricultural production in the DRC. The data include variables such as seed type, climatic conditions, soil characteristics, farming practices and yields obtained.

In this study on the analysis of agricultural yields in the Democratic Republic of Congo (DRC) using machine learning algorithms, the data come from various sources such as the Agence Nationale de Météorologie et des Télécommunications par Satellite (METTELSAT), the World Meteorological Organization (WMO), the Ministry of Agriculture, the FAO, local field studies and farmer surveys. These data include around 30,000 records, with variables such as seed type, climatic conditions, soil characteristics, farming practices and yields obtained.

a. Climate data: Agence Nationale de Météorologie et des Télécommunications par Satellite (METTELSAT), World Meteorological Organization (WMO), WorldClim.
b. Soil data: Ministry of Agriculture, FAO, local field studies, geospatial databases.
c. Agricultural data: DRC Ministry of Agriculture, local farming organization databases, farmer surveys.

Data pre-processing and training

Before analyzing the data, it is crucial to prepare it properly. Pre-processing includes data cleaning, normalization and variable transformation. The data is then divided into training and test sets to evaluate the performance of the machine learning models. Pre-processing steps:

a. Data cleaning: Removal of outliers and treatment of missing data.
b. Normalization and transformation: Normalization of continuous variables and coding of categorical variables.
c. Data splitting: 80% of data for training and 20% for testing.

Machine Learning algorithms used

Various machine learning algorithms were selected for their ability to model complex phenomena and offer robust performance. The following algorithms were used: Linear Regression, Random Forest Regression, Gradient Boosting Machines, Support Vector Machines, and Artificial Neural Networks (Gilgen et al., 2024).

Selected Algorithms and Training

Cross-validation and hyper-parameter optimisation are essential techniques in machine learning to improve the performance and generalisation of models (*Cross-Validation : Quel Est Ce Processus de Validation Croisée ?*, n.d.). Cross-validation (k-fold) (*Cross-Validation : Quel Est Ce Processus de Validation Croisée ?*, n.d.): Cross-validation is a statistical method used to evaluate the ability of a model to generalise on unseen data. The most common technique is k-fold cross-validation, which involves dividing the data set into k subsets (or 'folds') of equal size. The model is then trained k times, each time using k-1 subsets for training and the remaining subset for validation. This process is repeated until each subset has been used once as a validation set. The results of the k iterations are then averaged to provide an estimate of the model's performance. This approach maximises the use of available data and reduces the risk of overlearning.
Optimisation of hyperparameters (*Cross-Validation : Quel Est Ce Processus de Validation Croisée ?*, n.d.): Hyperparameters are model parameters that are not learned directly from the data during training, but must be defined prior to the learning process. Hyperparameter optimisation aims to identify the optimal combination of these parameters to improve model performance.

Techno Nusa Mandiri: Journal of Computing and Information Technology
Vol. 22, No. 1 March 2025 | DOI https://doi.org/10.33480/techno.v22i1.6380

Common techniques used for this optimisation include:

a. Grid Search: This method involves defining a grid of possible values for each hyperparameter and training the model on all possible combinations of these values. Although this approach is exhaustive, it can become computationally expensive when the number of hyperparameters or possible values is large.
b. Random Search: Rather than systematically exploring all combinations, Random Search randomly selects combinations of hyperparameters to test. This method can be more efficient than grid search, especially when only a few hyperparameters have a significant influence on model performance.

By combining cross-validation with hyperparameter optimisation techniques, it is possible to robustly assess the model's performance and select the hyperparameters that maximise its generalisability.

Algorithm results

Model performance is evaluated on three separate farms (Farm A, Farm B, Farm C) to ensure a robust evaluation. Results are measured using several metrics: MSE, MAE, RMSE, R^2 Score, and MAPE. The results obtained for each model and each farm are summarized in Table 2 below:

Table 2. Results Obtained

Algorithm	Metric	Farm A	Farm B	Farm C
Linear regression	MSE	1200	1500	1300
	MAE	25	30	27
	RMSE	34.64	38.73	36.05
	R^2	0.85	0.82	0.84
	MAPE	15.2%	18.4%	16.7%
Random Forest Regression	MSE	800	1000	850
	MAE	18	22	20
	RMSE	28.28	31.62	29.15
	R^2	0.92	0.90	0.91
	MAPE	10.5%	12.3%	11.4%
Gradient Boosting Machines	MSE	700	900	750
	MAE	15	19	17
	RMSE	26.46	30.00	27.39
	R^2	0.94	0.91	0.93
	MAPE	9.2%	11.1%	10.0%
Support Vector Machines	MSE	900	1100	950
	MAE	20	24	22
	RMSE	30.00	33.17	30.82
	R^2	0.88	0.86	0.87
	MAPE	12.0%	14.5%	13.2%
Artificial Neural Networks	MSE	600	850	700
	MAE	12	17	14
	RMSE	24.49	29.15	26.46
	R^2	0.95	0.92	0.94
	MAPE	8.5%	10.7%	9.3%

Source: (Research Results, 2025)

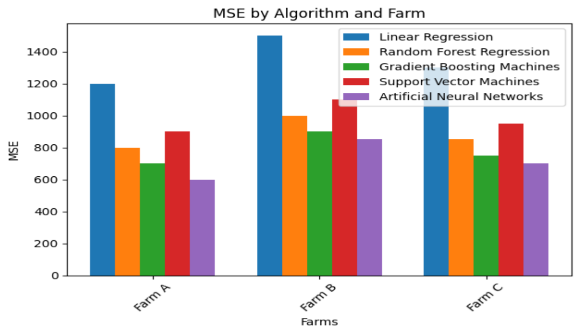
Table 2 shows the performance of the different algorithms on three different farms in terms of several important metrics. This allows an effective comparison of the accuracy and efficiency of the algorithms in predicting farm yields.

Graphical representations

We produced a python program using the Matplotlib library to generate bar graphs showing the performance of different regression algorithms on several farms, according to different evaluation metrics. Here are the results for each graph produced by this program:

1. Mean Squared Error (MSE)

This graph shows the mean squared error (MSE) values for each algorithm on the three farms.

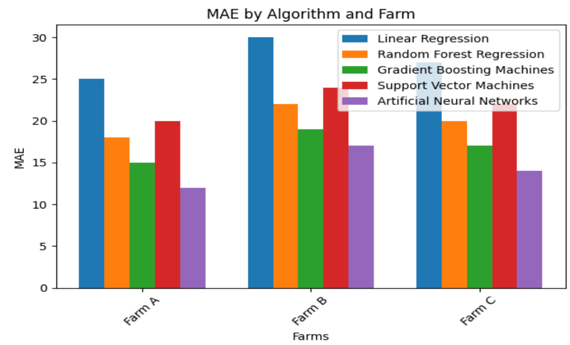


Source: (Research Results, 2025)
Figure 1. Mean Squared Error (MSE)

Interpretation: Linear Regression has the highest MSEs for all farms, indicating relatively poor performance. Artificial Neural Networks has the lowest MSE, suggesting that it has the best performance of the algorithms compared for this metric. Random Forest Regression and Gradient Boosting Machines show intermediate performance, with MSEs lower than Linear Regression but higher than Artificial Neural Networks.

2. Mean Absolute Error (MAE)

This graph shows the mean absolute error (MAE) values for each algorithm on the three farms.

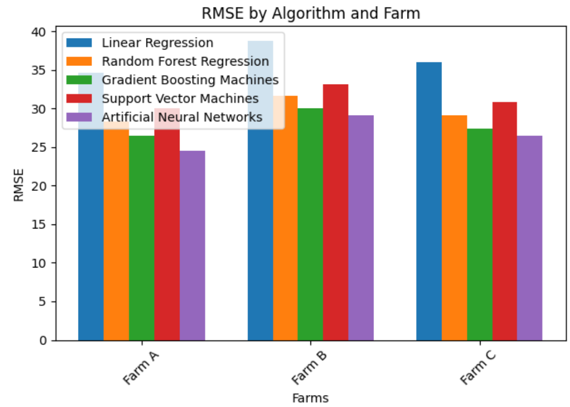


Source: (Research Results, 2025)
Figure 2. Mean Absolute Error (MAE)

Interpretation: Linear Regression again has the highest MAE values. Artificial Neural Networks has the lowest MAE values, confirming its good performance on this metric. The other algorithms (Random Forest Regression, Gradient Boosting Machines, and Support Vector Machines) perform better than Linear Regression but worse than Artificial Neural Networks.

3. Root Mean Squared Error (RMSE)

This graph shows the root mean square error (RMSE) values for each algorithm on the three farms.



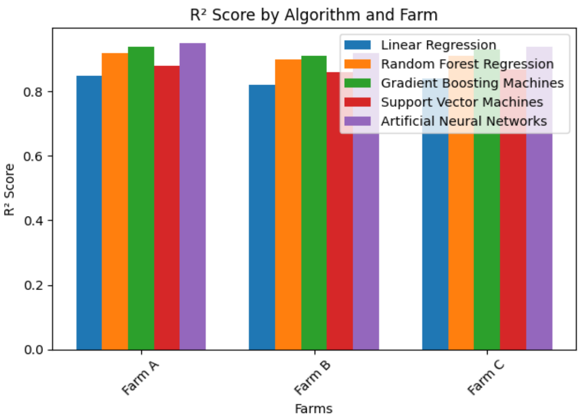
Source: (Research Results, 2025)

Figure 3. Root Mean Squared Error (RMSE)

Interpretation: The values follow a similar trend to the MSEs, since the RMSE is simply the square root of the MSE. Linear Regression has the highest RMSE values. Artificial Neural Networks has the lowest RMSE values. The performance of the other algorithms is intermediate, with RMSE values lower than those of Linear Regression but higher than those of Artificial Neural Networks.

4. R^2 Score

This graph shows the values of the coefficient of determination (R^2) for each algorithm on the three farms.



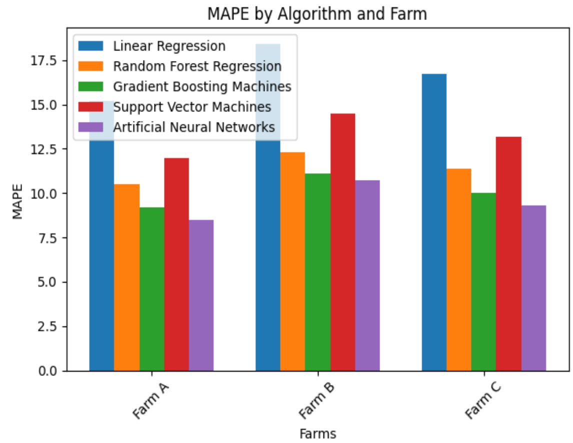
Source: (Research Results, 2025)

Figure 4. R^2 Score

Interpretation: Unlike the other metrics, higher R^2 values indicate better performance. Artificial Neural Networks has the highest R^2 scores, indicating a very good match between predictions and real data. Linear Regression has the lowest R^2 scores, indicating the worst fit of the algorithms compared. The other algorithms (Random Forest Regression, Gradient Boosting Machines, and Support Vector Machines) show intermediate performance.

5. Mean Absolute Percentage Error (MAPE)

This graph shows the mean absolute percentage error (MAPE) values for each algorithm on the three farms.



Source: (Research Results, 2025)

Figure 5. Mean Absolute Percentage Error (MAPE)

Interpretation: Linear Regression has the highest MAPE values. Artificial Neural Networks has the lowest MAPE values. Random Forest Regression, Gradient Boosting Machines, and Support Vector Machines perform better than Linear Regression but worse than Artificial Neural Networks.

General comments

Each graph shows the performance of the algorithms on three separate farms, allowing us to compare their efficiency in detail. Artificial Neural Networks appears to be the best performing algorithm on all metrics, while Linear Regression is the worst performing (Elbasi et al., 2023). The other algorithms show variable performance, but generally better than Linear Regression and worse than Artificial Neural Networks.

In summary, these graphs provide a comprehensive view of the performance of different regression algorithms in terms of accuracy and error over several farms, helping to identify which algorithms are best suited to this type of data (Morales & Villalobos, 2023).

Interpreting the results

Artificial Neural Networks (ANN) showed the best overall performance with the lowest MSE values and the highest R^2 scores, indicating excellent predictive ability. Gradient Boosting Machines (GBM) also performed very competitively, particularly in terms of MAPE, demonstrating their effectiveness for accurate predictions. Random Forest Regression, this model showed good robustness, although it performed slightly less well than ANN and GBM (Ahmed & Hussain, 2022). Linear Regression and SVM, although these algorithms perform less well than the more complex models, they are still useful for initial analyses and offer easier interpretability.

Discussion

The results obtained provide valuable insights for optimizing agricultural practices in the DRC. This section discusses the practical implications of the results and proposes recommendations for decision-makers and farmers.
Comparison of the algorithms

1. ANN and GBM performed best, suggesting their use for accurate and robust predictions.
2. Simpler models such as linear regression and SVMs can be used for less complex analyses or as teaching tools.

Practical implications

1. Optimization of agricultural practices: The results can guide farmers in selecting the best seeds and cultivation practices according to climatic conditions and soil characteristics.
2. Agricultural policies: Decision-makers can use the insights to develop more effective agricultural policies and to plan targeted interventions.

Limitations and future prospects

1. Limitations: The quality of the data can influence the results. Climatic variability and specific local conditions require particular attention.
2. Future research: Integration of additional data such as satellite images and drone analysis. Use of advanced techniques such as deep learning to further improve predictions.

This study demonstrates the effectiveness of machine learning algorithms for analyzing and predicting agricultural yields in the DRC. The results provide a solid basis for optimizing agricultural practices and formulating data-driven agricultural policies. By adopting these technological approaches, the DRC can significantly improve its agricultural productivity and ensure greater food security for its population (Iniyan et al., 2023).

CONCLUSION

This study demonstrates the effectiveness of machine learning algorithms in improving the prediction of agricultural yields in the DRC. By analysing a database of 30,000 records, it identifies the key factors influencing production, including seed type, climate, soil quality and farming practices. Artificial neural networks (ANNs) and gradient boosting machines (GBMs) outperformed traditional models, offering better accuracy with an MSE of 600, an R^2 of 0.95 and a MAPE of 8.5%.

These results open up a number of potential applications, including the integration of satellite imagery and drone data for accurate crop monitoring, real-time model adjustment, and the creation of decision-support platforms for farmers. Training farmers in these tools and developing agricultural policies based on these analyses would optimise the management of natural resources and anticipate the impacts of climate change.

The study highlights the potential of advanced algorithms to transform agriculture in the DRC, making the sector more productive, sustainable and resilient in the face of environmental and economic challenges (Morales & Villalobos, 2023).

REFERENCES

Ağbulut, Ü., Gürel, A. E., & Biçen, Y. (2021). Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. *Renewable and Sustainable Energy Reviews*, *135*, 110114. https://doi.org/10.1016/J.RSER.2020.110114

Ahmed, M. U., & Hussain, I. (2022). Prediction of Wheat Production Using Machine Learning Algorithms in northern areas of Pakistan. *Telecommunications Policy*, *46*(6), 102370. https://doi.org/10.1016/J.TELPOL.2022.102370

Bal, F., & Kayaalp, F. (2021). Review of machine learning and deep learning models in agriculture. *International Advanced Researches and Engineering Journal*, *5*(2), 309–323. https://doi.org/10.35860/IAREJ.848458

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors 2021, Vol. 21, Page 3758*, *21*(11), 3758. https://doi.org/10.3390/S21113758

Cross-validation : quel est ce processus de validation croisée ? (n.d.). Retrieved February 10, 2025, from https://www.intelligence-artificielleschool.com/ecole/technologies/cross-validation-croisee/

Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023). Crop Prediction Model Using Machine Learning Algorithms. *Applied Sciences 2023, Vol. 13, Page 9288*, *13*(16), 9288. https://doi.org/10.3390/APP13169288

Gilgen, L., Rodolphe, N. M., Kabutakapua, B., Chrisnovic, F., & Bopatriciat, B. M. (2024). PERFORMANCE COMPARISON OF CLASSICAL ALGORITHMS AND DEEP NEURAL NETWORKS FOR TUBERCULOSIS PREDICTION. *Jurnal Techno Nusa Mandiri*, *21*(2), 126–133. https://doi.org/10.33480/TECHNO.V21I2.5609

Gunjan, V. K., Kumar, S., Ansari, M. D., & Vijayalata, Y. (2022). Prediction of Agriculture Yields Using Machine Learning Algorithms. *Lecture Notes in Networks and Systems*, *237*, 17–26. https://doi.org/10.1007/978-981-16-6407-6_2

Huang, J. C., Ko, K. M., Shu, M. H., & Hsu, B. M. (2020). Application and comparison of several machine learning algorithms and their integration models in regression problems. *Neural Computing and Applications*, *32*(10), 5461–5469. https://doi.org/10.1007/S00521-019-04644-5/METRICS

Iniyan, S., Akhil Varma, V., & Teja Naidu, C. (2023). Crop yield prediction using machine learning techniques. *Advances in Engineering Software*, *175*, 103326. https://doi.org/10.1016/J.ADVENGSOFT.2022.103326

Jha, K. K., Jha, R., Jha, A. K., Hassan, M. A. M., Yadav, S. K., & Mahesh, T. (2021). A Brief Comparison on Machine Learning Algorithms Based on Various Applications: A Comprehensive Survey. *CSITSS 2021 - 2021 5th International Conference on Computational Systems and Information Technology for Sustainable Solutions, Proceedings*. https://doi.org/10.1109/CSITSS54238.2021.9683524

Khanam, J. J., & Foo, S. Y. (2021). A comparison of machine learning algorithms for diabetes prediction. *ICT Express*, *7*(4), 432–439. https://doi.org/10.1016/J.ICTE.2021.02.004

Morales, A., & Villalobos, F. J. (2023). Using machine learning for crop yield prediction in the past or the future. *Frontiers in Plant Science*, *14*, 1128388. https://doi.org/10.3389/FPLS.2023.1128388/BIBTEX

Mouhssine, R., & Otman, A. (2020). Performance Analysis of Machine Learning Techniques for Smart Agriculture: Comparison of Supervised Classification Approaches. *IJACSA) International Journal of Advanced Computer Science and Applications*, *11*(3). www.ijacsa.thesai.org

Mourtzinis, S., Esker, P. D., Specht, J. E., & Conley, S. P. (2021). Advancing agricultural research using machine learning algorithms. *Scientific Reports*, *11*(1), 17879.

Pant, J., Pant, R. P., Kumar Singh, M., Pratap Singh, D., & Pant, H. (2021). Analysis of agricultural crop yield prediction using statistical techniques of machine learning. *Materials Today: Proceedings*, *46*, 10922–10926. https://doi.org/10.1016/J.MATPR.2021.01.948

Raman, R., Kantari, H., Gokhale, A. A., Elangovan, K., Meenakshi, B., & Srinivasan, S. (2024). Agriculture Yield Estimation Using Machine Learning Algorithms. *2024 International Conference on Automation and Computation, AUTOCOM 2024*, 187–191. https://doi.org/10.1109/AUTOCOM60220.2024.10486107

République démocratique du Congo - Vue d'ensemble. (n.d.). Retrieved February 6, 2025, from https://www.banquemondiale.org/fr/country/drc/overview?

Sekeroglu, B., Hasan, S. S., & Abdullah, S. M. (2020). Comparison of Machine Learning Algorithms for Classification Problems. *Advances in Intelligent Systems and Computing*, *944*, 491–499. https://doi.org/10.1007/978-3-030-17798-0_39

Sharifi, A. (2021). Yield prediction with machine learning algorithms and satellite images. *Journal of the Science of Food and Agriculture*, *101*(3), 891–896. https://doi.org/10.1002/JSFA.10696

van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, 105709. https://doi.org/10.1016/J.COMPAG.2020.105709