A Comparative Study of Machine Learning Approaches for Rice Yield Prediction in Sumatera

Vito Xavier   
School of Computer Science  
Binus UniversityJakarta, Indonesia  
[vito.xavier@binus.ac.id](mailto:vito.xavier@binus.ac.id) Eldwin Felice  
School of Computer Science  
Binus UniversityJakarta, Indonesia  
[eldwin.felice@binus.ac.id](mailto:eldwin.felice@binus.ac.id)   
Axel Valent  
School of Computer Science  
Binus UniversityJakarta, Indonesia  
[axel.valent@binus.ac.id](mailto:axel.valent@binus.ac.id)

*Abstract*—While being the most widely consumed staple food, rice production in Indonesia is relatively low. Therefore, as an effort to maintain food security, this study aims to evaluate different machine learning methods for rice production prediction in Sumatera. Previous studies regarding crop prediction have concluded that nonlinear models such as Random Forest and Neural Networks performed better than linear models. And while some studies in crop prediction experimented with Principal Component Analysis (PCA), most studies used only linear models. The experiments conducted in this study include the application of PCA and PCA on climate variables, using linear machine learning models which include Multiple Linear Regression (MLR), and LASSO Regression. And nonlinear models which include Polynomial Regression, Support Vector Regression (SVR), Random Forest Regression, and Neural Networks using Multi-layer Perceptron Regression. Experiment results show performing PCA on climate variables improves model performance, although performing PCA on all variables may cause significant information loss which result in low accuracies. And while nonlinear models such as SVR generally performed better than linear models with an R2 of 86.8566, they do not perform significantly better than linear models with an R2 of 86.1941, which might allude to the linear relationships in the data. Due to the nature of the data, this study concludes that further data mining and analysis efforts must be performed on agricultural and climate data in Indonesia in order to create a better statistical representation of crop growth and production in Indonesia.

*Keywords—Crop Yield Prediction, Machine Learning, PCA, Multiple Linear Regression, LASSO Regression, Polynomial Regression, Support Vector Regression, Random Forest, Neural Networks, Data Mining*

# Introduction

The popularity of rice as a staple food in Indonesia is not to be contested. This is shown through data from Statista in 2021/2022 [1], which highlights Indonesia as the fifth highest rice consuming country in the world, with a total of 35.6 tons of rice. And in order to fill that demand, Indonesia must procure copious amounts of rice for consumption, in which Indonesia is very blessed with its soil fertility and climate. According to data from the Food and Agriculture Organization (FAO) in 2021 [2], Indonesia ranks third in worldwide rice production, with a total of 54.415.294 tons of rice. And with this amount of production and food security, Indonesia has successfully achieved rice self-sufficiency from 2019-2021, where this achievement was also awarded by the International Rice Research Institute (IRRI) [3]. But with concerns regarding international food crises, climate change, and the increasing world population, it cannot be said with confidence that Indonesia currently has a surplus of rice, as the population is still vulnerable to a decline in rice production. This is shown through very conservative policies regarding the export of rice [4], where on a worldwide scale Indonesia ranks relatively low [5]. Concerning this problem, in accordance with 2 of 17 of the United Nations Sustainable Development Goals, namely zero hunger and climate action, an effort to predict rice production in Indonesia must be undertaken, in order to support and maintain the national food security. In this case, in accordance to the efforts of predicting rice production in Indonesia, a dataset for rice production in the island of Sumatra from the years 1993-2020 has been procured [6], which was compiled and provided by Mr. Ardika Satria on the website Kaggle. The data consists of yearly rice production of every province in the island of Sumatra, obtained from the website of the Central Bureau of Statistics (BPS), and weather data which consists of rainfall, humidity, and average temperature from the website of the Meteorology, Climatology, and Geophysics Council (BMKG).

In this modern world, there are many technological and scientific advancements that allow us to store, process, and make complex calculations. One such calculation technique is machine learning, where computational models are created and trained based on data, to complete a certain task. Which in this case the computational model is intended to perform regression analysis, a statistical calculation model that allows us to estimate a variable, based on a set of other variables. Regression analysis is an important and favored statistical method for estimating data; however, there are many different types of regression with differing use cases, based on the provided data. In the case of crop yield prediction, many studies have implemented different types of regression and machine learning methods, each with differing data types, methods, and conclusions [7]–[16]. Some studies have also proven that linear regression models performed poorly compared to nonlinear regression models [9], [13], [14]. Principal Component Analysis (PCA) may also be performed on regression data, where the dimensionality of the data could be reduced that may create a better representation of the variables, considering the high chance of redundancy with time series data such as climate data. And yet most studies that research the effects of Principal Component Analysis (PCA) on crop yield prediction utilize linear regression models to predict data. [9] Considering the performance of nonlinear regression models, the benefits of PCA on regression analysis, and the possibility of a nonlinear relationship between variables in the data, we intend to utilize and compare popular models from both types of regressions and study the effects of PCA on the data.

This study aims to evaluate and compare different linear and nonlinear regression techniques namely multiple linear regression, LASSO regression, polynomial regression, support vector regression, random forest regression, neural networks, and the effects of PCA on the variables, in order to find the most precise regression model for the provided dataset. This research contains 4 sections, section 1 defines the theories and basis of our experimentation, section 2 describes the materials used in the research, section 3 explains the results of our experiments, and in the last section, the findings of this research and other conclusions will be discussed.

# Related Works

In the field of Crop Yield Prediction using Machine Learning, choosing the correct model, dataset, and features are key to calculate satisfactory results. The provided data must give enough insight for the model to simulate the characteristics of crop growth conditions, and the models should be able to learn the relationships between the different variables. Therefore, to establish a foundation for the experiments that will be performed in this study, a literature review was conducted on recent studies regarding the analysis of machine learning models and methods on crop yield prediction. The literature review revealed various methods of crop yield estimation using Machine Learning, in different parts of the world. Although differences in climate conditions, crop type, and crop management may be significant in our study, these articles have provided some insight into their datasets, feature extraction methods, models, and results, which creates the possibility of adapting and experimenting different methods for this study.

In this field, there were several papers that experimented upon the performance of various feature extraction methods for training and fitting machine learning models. In the paper Selection of Important Features for Optimizing Crop Yield Prediction by Maya Gopal P S, and Bhargavi R. [7], different feature selection algorithms for regression models of crop yield prediction in Tamil Nadu were compared and evaluated, where it was found that forward selection algorithms performed best. And with the importance of irrigation and water supply for the growth of rice, data regarding irrigation and water supply might be important for the prediction of rice

In the paper Sugarcane Crop Yield Forecasting Model Using Supervised Machine Learning by R. A. Medar, et al. [8]. Conducted a unique experiment by predicting the weather conditions and NDVI (Normalized Vegetation Index) using time series forecasting and other parameters. This experiment tries to compare some Machine Learning algorithms to predict the weather conditions and soil attributes such as Support Vector Regression (SVR), LASSO, Naive Bayes, and Decision tree regression. And some Machine Learning algorithms to predict NDVI and crop yield such as Gaussian Process Regression (GPR), Lasso Regression, SVR with RBF kernel, and Kernel Ridge-RBL. The result of this experiment concludes that some attributes can be predicted using other methods such as time series forecasting or prediction using other attributes. Naive Bayes performs better than other 3 algorithms for weather conditions and soil attributes, and SVR-RBF performs better for predicting NDVI and crop yield.

In the paper Mustard Yield Prediction using Machine Learning Approach by A. Goyal and A. Vashisth [9], different machine learning approaches and hybrid models were used to predict mustard yield. Results from this study suggest that feature extraction with Principal Component Analysis (PCA) performed better than Stepwise Multiple Linear Regression (SMLR), with a combination of Support Vector Machine (SVM) and Artificial Neural Network (ANN) being the best performing model, followed by SVM. According to this study, a hybrid model may perform better than individual models if each model can support the shortcomings of the other, but not all combinations of models can achieve better results that the models individually.

And in the paper Enhancing Crop Yield Prediction Utilizing Machine Learning on Satellite-Based Vegetation Health Indices by H. T. Pham, et al. [10], Ensemble Boost Tree with PCA is used in crop yield data and satellite data in the form of Vegetation Condition Index (VCI) and Thermal Condition Index (TCI). Where the research results indicate that machine learning models with PCA have higher accuracy than without PCA. This study has also uncovered that, due to climate data being a type of time-series data, it is very vulnerable to data redundancy. One technique that can overcome this problem is Principal Component Analysis or PCA. However, many of the studies found by H. T. Pham, et al. performed PCA on a linear regression model, knowing there is still the possibility of a nonlinear relationship within the data.

There were also several papers that make comparisons between several Machine Learning models to find out which model has the best performance. In the paper Crop Yield Prediction based on Indian Agriculture using Machine Learning by P. S. Nishant, P. Sai Venkat, B. L. Avinash, and B. Jabber [11]. Comparing LASSO Regression, Elastic Net, Kernel Ridge, and stacked regression that stack all those models for crop yield prediction. The result says that the result for applying stacked regression is better than when those models were applied individually.

In the paper Linear Regression Model to Study the Effects of Weather Variables on Crop Yield in Manipur State by B. Bhattacharyya, et al. [12], linear regression was used on harvest and weather datasets for various types of food in the state of Manipur, India. Through this study, we found that the amount of rice harvested was significantly affected by the harvest area and many climatic parameters, but temperature and relative humidity had a negative effect. And multiple linear regression can produce fairly accurate predictions based on climate data.

In the paper Predicting Crop Yields in Senegal using Machine Learning Methods by A. B. Sarr and B. Sultan [13], linear regression in the form of LASSO Regression is compared to several types of nonlinear regression namely Random Forest, SVM, and Neural Network, on climate data datasets, NDVI satellite data, and yields of several types of food. In this study, it has been proven that in general, types of nonlinear regression such as Random Forest, SVM, and Neural Network provide better accuracy values compared to LASSO which is a type of linear regression.

In the paper Design and Implementation of Crop Yield Prediction Model in Agriculture by S. Gowda and S. Reddy [14], Polynomial Regression, Decision Tree, Random Forest Regression was carried out on climate data datasets, NVDI, satellite data, and crop yields in India. Where it was proven that Polynomial Regression and Random Forest Regression perform better than Decision Tree, therefore Polynomial Regression and Random Forest Regression could be strong candidates for nonlinear regression research.

There are also papers that explain using algorithms such as neural networks and deep learning algorithms. In the paper Machine learning methods for crop yield prediction and climate change impact assessment in agriculture by A. C. Droesch [15], Neural network algorithms such as Semiparametric Neural Networks, Parametric, Fully-nonparametric neural network were used to produce models that use semiparametric variance from deep neural networks, Combining machine learning with domain area knowledge from empirical studies improves predictive skills, while also showing the impact of climate change on agriculture.

In the paper A Hybrid Approach for Crop Yield Prediction using Machine Learning and Deep Learning Algorithms by S. Agarwal and S. Tarar [16], This paper uses SVM, LSTM, and RNN for crop yield prediction which achieved and accuracy of 97%, which is better compared to a previous study which achieved an accuracy of 93% using RF and ANN.

# Research Methodology

## Datasets

This research will use a dataset of Sumatra Rice Plant sourced from Kaggle which was uploaded by Ardika Satria in January 2023 [6]. This dataset contains rice crop production from 8 provinces in Sumatra in 1993-2020. There are 224 rows each consisting of 7 columns which are Province, Year, Production Yield, Production Area, Rainfall, Humidity, and Average Temperature. This research will attempt to predict production yields for the entire island of Sumatera based on province, year, land area, rainfall, humidity, and average temperature.

## Device

This research will be conducted through the utilization of Google Colab and the free GPU accelerator provided by Colab. This research will focus on supervised learning models through the use of the scikit-learn library in training and evaluating Machine Learning models.

## Experiment Methods

In this paper, experimentation will be conducted to find models with the lowest error rate for predicting rice crop yields in Sumatra. From the dataset, six features are chosen in order to predict production result, namely province, year, harvested area, rainfall, humidity, average temperature, and production yield. Where in this experiment, production result will be the dependent variable. After separating the data to its dependent and independent variables, for each machine learning model, three different data treatment methods are evaluated.

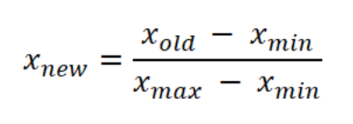


Figure 1. Min-Max Normalization Formula

A picture containing diagram, text, line, parallel

Description automatically generated

Figure 2. First Experiment

As shown in Fig. 2, in the first experiment the data will be normalized using Min-Max Normalization before being used to train the model. This is achieved by subtracting the data value with the minimum value from all data and will be divided by the difference in the maximum and minimum data values, where the results will have a range of 0 to 1. The model will then be validated using 5-fold Cross-Validation, where 5 different validation sets are picked randomly from the entire dataset, with a resulting split of 80:20. For the purposes of our experiment, in order to create comparable results, the randomness of the selection will be fixed using a random state of 100. After the training and validation of each fold, evaluation results from each fold will be averaged to calculate the result.

A picture containing diagram, line, plan, parallel

Description automatically generated

Figure 3. Second Experiment

In the second experiment, after normalization PCA will be performed on climate variables, namely rainfall, humidity, and average temperature, before the training of the model. This will result in a singular normalized PCA feature which will be named Weather Variable, as shown in Fig. 3. The data will then be trained and validated using 5-fold Cross-Validation. And for the third experiment, PCA will be performed on all of the feature variables before training and validation.

A picture containing diagram, line, parallel, plan

Description automatically generated

Figure 4. Third Experiment

## Models and Evaluation

For the three data treatment experiments above, six machine learning models will be tested in this study, namely Multiple Linear Regression, LASSO Regression, Polynomial Regression, Support Vector Regression, Random Forest Regressor, and Neural Networks in the form of Multi-layer Perception (MLP) Regressor. As concluded in [17], considering the random nature of non-deterministic models such as Random Forest Regression and MLP Regression, results from different iterations of model training may vary, therefore the mean of results from different iterations of model training are required in order to capture the non-deterministic nature of Random Forest and Neural Networks. This issue may be solved through the usage of 5-fold Cross-Validation, as for the training of each fold, a new model will be declared with different values, and results will be concluded from the mean of each fold.

Each experiment will be evaluated using 4 metrics, namely R-Squared, Mean Absolute Error, Mean Squared Error, and Root Mean Squared Error. R-Squared represents the proportion of the variance in the dependent variable that explained in the regression model, value of R2 will be less than one. Mean Absolute Error represents the average of the absolute difference between predicted and actual values in datasets. Mean Squared Error measures the variance of the residuals, it represents the average of the squared difference between predicted and actual values and Root Mean Squared Error is the square root of Mean Squared Error and it measured the standard deviation of residuals.

Text

Description automatically generated

Figure 5. R-Squared Formula

Diagram, text, schematic

Description automatically generated with medium confidence

Figure 6. Mean Absolute Error Formula

Diagram, schematic

Description automatically generated

Figure 7. Mean Squared Root Error Formula

Text

Description automatically generated with medium confidence

Figure 8. Root Mean Squared Error Formula

TABLE I. Results from each experiment method and model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Prediction Model** | **Data Treatment** | **R2** | **MAE** | **MSE** | **RMSE** |
| **Multiple Linear Regression** | **Base Model** | 85.8131 | -240,722.6102 | -177,914,284,023.7330 | -420,168.3282 |
| **PCA (Weather Variable)** | 86.1941 | -238,358.1456 | -171,945,109,800.6790 | -413,269.7145 |
| **PCA (All Variables)** | 57.7250 | -557,895.7028 | -524,166,942,599.4360 | -723,434.7196 |
| **LASSO Regression** | **Base Model** | 85.8131 | -240,722.3615 | -177,914,135,899.1690 | -420,168.1683 |
| **PCA (Weather Variable)** | 86.1941 | -238,358.0791 | -171,945,121,429.0440 | -413,269.7355 |
| **PCA (All Variables)** | 57.7250 | -557,895.7358 | -524,166,956,606.3900 | -723,434.7288 |
| **Polynomial Regression** | **Base Model** | 80.6770 | -275,496.8435 | -229,753,331,282.9680 | -474,478.9171 |
| **PCA (Weather Variable)** | 86.7334 | -227,188.5406 | -161,221,083,957.1600 | -399,492.0805 |
| **PCA (All Variables)** | 58.1467 | -568,285.8852 | -518,785,169,175.4930 | -719,503.8680 |
| **Support Vector Regression** | **Base Model** | 86.8405 | -184,049.7637 | -162,867,032,080.8060 | -401,043.4209 |
| **PCA (Weather Variable)** | 86.8566 | -172,203.1117 | -161,166,175,812.7150 | -399,246.0000 |
| **PCA (All Variables)** | 58.0304 | -504,448.5430 | -521,225,809,884.2560 | -716,633.4711 |
| **Random Forest Regression** | **Base Model** | 85.0481 | -190,718.7323 | -169,822,320,918.0480 | -409,866.7045 |
| **PCA (Weather Variable)** | 85.2450 | -184,907.0626 | -174,009,451,879.4930 | -407,901.8711 |
| **PCA (All Variables)** | 59.7636 | -420,655.5262 | -514,146,160,078.2080 | -699,617.1522 |
| **Neural Networks**  **(Multilayer Perceptron**  **Regression)** | **Base Model** | 85.1947 | -236,296.0358 | -184,220,426,352.5940 | -426,286.9542 |
| **PCA (Weather Variable)** | 86.1941 | -238,358.5030 | -171,945,583,585.5220 | -413,270.2918 |
| **PCA (All Variables)** | 57.7250 | -557,895.7119 | -524,166,945,926.3220 | -723,434.7218 |

# Results

Based on the result of our experiments as shown in Table 1, evaluation of different data treatment methods and modelling show very consistent results. Although each model uses different approaches in understanding the data, this reveals that all the experimented models are able to reliably observe underlying patterns and relationships in the data. Results among all experiments and models show that dimension reduction using PCA on climate variables will increase model performance compared to the usage of PCA on all variables and models where PCA is not performed. And although the usage of PCA on a selection of features may increase performance, the application of PCA on all variables show a significant decrease in performance, which may allude to a loss of information from the features as a result of dimensionality reduction. In this case, our findings are in accordance with the results shown in [10].

Model experimentation shows that SVR performed best with an R2 value of 86.8566, followed by Polynomial Regression with a value of 86.7334, and MLR, LASSO Regression, and MLP Regression with a value of 86.1941, in which PCA on climate variables are performed on all of the above experiments. For base models, SVR performed best with an R2 value of 86.8405, and for PCA on all variables, RF Regression performed best with an R2 value of 59.7636.

The consistency of the results from different models show that all models are able to predict underlying patterns and relationships in the data, including linear machine learning models such as MLR and LASSO Regression. From this it can be concluded that there is a linear relationship between the variables in the dataset used for this experiment. And although in accordance to previous studies such as [9], [13], and [14], that concluded nonlinear models such as Support Vector Regression show better performance, the increased complexity of the models may not be beneficial in predicting results from this particular data.

And considering the number of features available, with the linear relationship between variables in the dataset, adding more features into the dataset will not only better represent conditions in the field, but may also be beneficial to some of the machine learning approaches used in this experiment. In the context of rice production, considering the importance of water availability for rice growth, data regarding irrigation such as canal size and availability of tanks or wells, or both may be beneficial in predicting crop yield, as seen in [7].

# Conclusion

The prediction of rice production in Indonesia is greatly beneficial to further the development of the agricultural industry in Indonesia. Although the results of this experiment show promising results, further analysis reveals the linear nature of the data provided by BPS and BMKG through Satria [6]. Therefore, further data mining and analysis for agricultural and climatic data in Indonesia may reveal more about the characteristics of crop growth in Indonesia. And through this, further data and model experimentation may be performed in order to create more accurate models for predicting not only rice, but also other crops grown in Indonesia.

# Acknowledgement

We thank the supervisors and lecturers that helped us throughout the writing process of this study through their constructive suggestions, insights, and comments.

##### References

1. F. Javier, “Cina, Negara Dengan konsumsi Beras Terbanyak di Dunia,” Tempo, 14-Oct-2022. [Online]. Available: https://data.tempo.co/data/1529/cina-negara-dengan-konsumsi-beras-terbanyak-di-dunia. [Accessed: 26-Apr-2023].
2. *Crops and livestock products*, Food and Agriculture Organization of the United Nations, FAOSTAT Statistical Database, 24-Mar-2023. [Online]. Available: https://www.fao.org/faostat/en/#data/QCL/visualize. [Accessed: 26-Apr-2023].
3. I. W. Suarjana, “Indonesia 3 Tahun Tanpa impor Beras Konsumsi,” Dinas Pertanian dan Ketahanan Pangan Provinsi Bali, 18-Aug-2022. [Online]. Available: https://distanpangan.baliprov.go.id/indonesia-3-tahun-tanpa-impor-beras-konsumsi/. [Accessed: 26-Apr-2023].
4. “RI Swasembada 3 Tahun, Kenapa Tolak Ekspor Beras ke China dan Arab?,” CNN Indonesia, 24-Aug-2022. [Online]. Available: https://www.cnnindonesia.com/ekonomi/20220824194945-92-838758/ri-swasembada-3-tahun-kenapa-tolak-ekspor-beras-ke-china-dan-arab. [Accessed: 26-Apr-2023].
5. *Rice in Indonesia*, Observatory of Economic Complexity (OEC). [Online]. Available: <https://oec.world/en/profile/bilateral-product/rice/reporte> r/idn. [Accessed: 27-Apr-2023]
6. A. Satria, “Dataset Tanaman Padi Sumatera, Indonesia”, Kaggle. 11-Dec-2022. [Online]. Available: https://www.fao.org/faostat/en/#data/QCL/visualize. [Accessed: 26-Apr-2023].
7. Maya Gopal P S and Bhargavi R., “Selection of important features for optimizing crop yield prediction,” International Journal of Agricultural and Environmental Information Systems, vol. 10, no. 3, pp. 54–71, Jul. 2019. [Online]. doi: 10.4018/IJAEIS.2019070104. [Accessed: 02-May-2023].
8. R. A. Medar, V. S. Rajpurohit, and A. M. Ambekar, “Sugarcane crop yield forecasting model using supervised machine learning,” International Journal of Intelligent Systems and Applications, vol. 11, no. 8, pp. 11–20, 2019. [Online] Available: <http://www.mecs-press.net/ijisa/ijisa-v11-n8/IJISA-V11-N8-2.pdf> [Accessed: 02-May-2023].
9. A. Goyal and A. Vashisth, “Mustard Yield Prediction using Machine Learning Approach,” Journal of Agricultural Physics, vol. 21, no. 2, pp. 445–456, 2021. [Online]. Available: <http://agrophysics.in/admin/adminjournalpdf/20220811163131872039014/journal-82940449.pdf>. [Accessed: 02-May-2023].
10. H. T. Pham, J. Awange, M. Kuhn, B. V. Nguyen, and L. K. Bui, “Enhancing crop yield prediction utilizing machine learning on satellite-based vegetation health indices,” Sensors, vol. 22, no. 3, p. 719, Jan. 2022. [Online]. doi: <https://doi.org/10.3390/s22030719>. [Accessed: 26-Apr-2023]
11. P. S. Nishant, P. Sai Venkat, B. L. Avinash, and B. Jabber, “Crop yield prediction based on Indian agriculture using machine learning,” 2020 International Conference for Emerging Technology (INCET), 2020. [Online] doi: 10.1109/INCET49848.2020.9154036. [Accessed: 02-May-2023].
12. B. Bhattacharyya, R. Biswas, Sujatha K. and D.Y. Chiphang, “LINEAR REGRESSION MODEL TO STUDY THE EFFECTS OF WEATHER VARIABLES ON CROP YIELD IN MANIPUR STATE,” International Journal of Agricultural and Statistics Sciences, vol. 17, no. 1, Mar. 2021. [Online]. Available: <https://www.researchgate.net/publication/352735410_LINEAR_REGRESSION_MODEL_TO_STUDY_THE_EFFECTS_OF_WEATHER_VARIABLES_ON_CROP_YIELD_IN_MANIPUR_STATE>. [Accessed: 26-Apr-2023].
13. A. B. Sarr and B. Sultan, “Predicting crop yields in Senegal using machine learning methods,” International Journal of Climatology, vol. 43, no. 4, pp. 1817–1838, 2022 [Online]. doi: <https://doi.org/10.1002/joc.7947>. [Accessed: 26-Apr-2023]
14. S. Gowda and S. Reddy, “Design And Implementation Of Crop Yield Prediction Model in Agriculture,” International Journal of Scientific & Technology Research, vol. 8, no.1, pp. 544-549, Jan. 2020. [Online]. Available: <https://www.researchgate.net/publication/344560876_Design_And_Implementation_Of_Crop_Yield_Prediction_Model_In_Agriculture>. [Accessed: 26-Apr-2023]
15. A. C. Droesch, A. Kassahun, and C. Catal, “Machine learning methods for crop yield prediction and climate change impact assessment in agriculture,” Environmental Research Letters, 2018. [Online]. doi: 10.1088/1748-9326/aae159. [Accessed: 26-Apr-2023]
16. S. Agarwal and S. Tarar, “A HYBRID APPROACH FOR CROP YIELD PREDICTION USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS,” Journal of Physics: Conference Series, Jan-2021. [Online]. doi: 10.1088/1742-6596/1714/1/012012. [Accessed: 26-Apr-2023]
17. N. Reimers and I. Gurevych, “Why Comparing Single Performance Scores Does Not Allow to Draw Conclusions About Machine Learning Approaches”, arXiv [cs.LG]. 2018. [Online]. doi: <https://doi.org/10.48550/arXiv.1803.09578>. [Accessed: 29-Apr-2023]