Adaptive Crop Yield Forecasting Using Statistical Learning Algorithms

Team 4 - Vaishak Muralidharan, Kaushal Shivaprakashan, Vamsi Sai Garapati

Introduction: [Q1]

As the world gets hungrier, so does it need better ways of agriculture. Yield forecasting informs resource use, planting, and improves food security. Hence, this paper seeks to develop an improved model of machine learning in crop yield forecasting, considering climate, soil, and resources. Hence, the regression and classification techniques will perform feature selection automatically, to be used by the model in adapting to regional conditions. Precise and reliable crop yield forecasts will thus be seen in productivity and sustainable development among farmers and people involved in agribusiness. The statistical learning method will identify such patterns where the proposed solution enjoys strong support in data-driven insight

Impact and Relevance and Evaluation: [Q4, Q5, Q9]

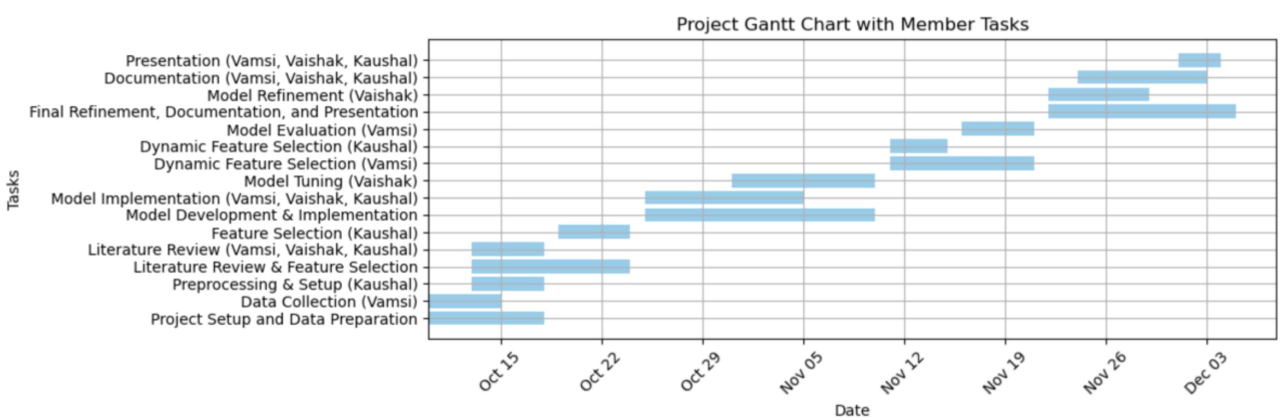
This will make the project highly valuable to farmers, agricultural businesses, and policymakers through correct forecasts of crop yield, hence helping them in resource management and cost reduction with enhanced food security. Researchers will benefit from the insights that come from the influencing factors of yield to help in the development of better agricultural practices and policies. The project will be also measuring success with midterm evaluations using key performance metrics-RMSE and R-squared-and real-world validation with agricultural partners for practical applicability. It shall further be assured that the model optimizes resource use qualitatively. Field trials would become the final exam in which the reliability of the model across diverse conditions and impacts on decision-making for the stakeholders is to be demonstrated.

Risks and Payoffs: [Q6]

There could otherwise be data integrity, systemic regression, or availability issues that may develop and ultimately limit the effectiveness of this model. The other risks include overfitting, which reduces the capability for generalization in varied regions. On the upside, this has very high power for proper forecast techniques on crop yield, enhances agricultural decision-making, cuts costs in resource use optimization, and contributes toward food security. This can effectively be mitigated to create a strong, resilient, adaptive framework that enables farmers and stakeholders in agriculture with data-informed insights to make better decisions.

Cost, Timeline and contribution [Q7, Q8]

There are no costs involved in this project as the dataset is freely available online from kaggle.



Literature survey: [Q2, Q3]

Crop yield prediction plays an important role in responding to climate change and developing sustainable agriculture. Machine learning techniques, such as ensemble learning and artificial neural networks, predict crop yield effectively and help in optimization of resources and making better decisions. Many research studies have explored these approaches, hence providing useful insight into the development of models that can be used for making a prediction.

Various studies stress the point that quality data and feature selection dynamically will bring the effective crop yield forecasts. Islam et al. emphasize that estimation of Yp and Yw requires highly resolved detailed data on weather, soil, and management practices [1]. Another research, which is based on an ANN-based approach, explains that climatic variables like temperature, rainfall, and nutrients in the soil will be predictors for crop yields [2]. The papers also draw on dynamic feature selection so as to make the models adaptive across different regions [1, 4]. Several studies discuss the effects of climatic variables on crop yield.

Papers related to climate change impacts on land suitability for agriculture and the usage of LSTM networks for classification prove that models can indeed predict land-use change due to climate variability such as changes in precipitation and temperature [3]. While these models emphasize long-term land suitability, the insight from climate data may well be applicable in short-term crop yield forecasting, especially with regression models, which can allow for real-time adaptability [2,5]. Along with those, ensemble techniques like XGBoost are very effective in predicting crop yields from complex data.

The paper about crop yield prediction in Saudi Arabia provides evidence that XGBoost, Random Forest, and KNN were used to predict the yield based on variables such as temperature, rainfall, and pesticides [1, 5]. The XGBoost showed the best performance with an R^2 score of 0.9745 [4]. These ensemble methods can improve your project by better tuning the predictions for regional climate and soil variations [4]. On the other hand, some studies note limitations in their generalizability.

Ensemble models and ANNs achieve high accuracy but struggle to adapt to new regions or datasets without major recalibration [3,4]. The study on wheat yield prediction displays several limitations, where machine learning models like Random Forest and Boosted Trees are overfitting and not generalizing well [5]. Again, cross-validation and dynamic feature selection will render the model flexible over different regions and conditions [7]. This seriously restricts its application in changing environments with fluctuating climate and resources, since most of them do not have real-time feature selection.

Papers on climate change effects using models like KNN and SVM emphasize that real-time adaptability is necessary to place emphasis on the main predictors, such as soil quality and rainfall, considering the changing environmental conditions [9].

This again would improve the responsiveness for your project in different agricultural conditions, reduce the risks of overfitting, and improve the wide-area prediction accuracy of models [4,6]. These studies have finally indicated the crucial role that machine learning can play in predicting crop yield. Dynamic feature selection and ensemble methods can be implemented in your project toward enhancement of the existing models for better predictions under various conditions. This needs to ensure resource optimization and reduce overfitting while supporting sustainable agriculture in view of climate change.

Data Set:

https://buffalo.box.com/s/ildyy5p0vrwh62iyy0r0m4l0stvcumht

References:

[1] Islam, R., & Masum, A. K. M. (2024). Crop yield prediction through machine learning: A path towards sustainable agriculture and climate resilience in Saudi Arabia. AIMS Agriculture and Food, 9(4), 980–1003. https://doi.org/10.3934/agrfood.2024053

[2] Islam, A., Schuenemann, J., & Webber, H. (2023). How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. Global Food Security, 43, 100659. https://doi.org/10.1016/j.gfs.2023.100659

[3] Chen, T., & Wang, L. (2023). Climate change impact on agricultural land suitability: A machine learning-based Eurasia case study. arXiv preprint. https://doi.org/10.48550/arXiv.2310.15912

[4] Dahikar, S. S., & Rode, S. V. (2014). Agricultural crop yield prediction using artificial neural network approach. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 2(1), 683-686.

[5] Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P. (2015). Crop selection method to maximize crop yield rate using machine learning technique. International Journal of Computer Applications, 116(23), 1–5. https://doi.org/10.5120/20525-3238

[6] Tejeswani, P., Lakshmi, P., Pallavi, T., Purna, P. N. B., Ashesh, K., & Vara Prasad, P. V. (2024). Analysing the effect of climate change on crop yield using machine learning techniques. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4802729

[7] Iqbal, N., Shahzad, M. U., Sherif, E. M., Rashid, J., Le, T. V., Tariq, M. U., & Ghani, A. (2024). Analysis of wheat-yield prediction using machine learning models under climate change scenarios. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1234567

[8] Kang, Y., Khan, S., & Ma, X. (2009). Climate change impacts on crop yield, crop water productivity, and food security – A review. Agricultural Water Management, 97(4), 523–531. https://doi.org/10.1016/j.agwat.2009.08.005

[9] Mishra, S., Mishra, D., & Santra, G. H. (2020). Applications of machine learning techniques in agricultural crop production: A review paper. Biosc.Biotech.Res.Comm., 13(4), 1857–1862. https://doi.org/10.21786/bbrc/13.4/54