Adaptive Crop Yield Forecasting Using Statistical Learning Algorithms

Team 003

VAMSI SAI GARAPATI, VAISHAK MURALIDHARAN , KAUSHAL SHIVAPRAKASHAN

Introduction

The increasing need for accurate crop yield prediction is driven by the challenges of modern agriculture, such as climate change, resource optimization, and food security. This project aims to forecast crop yields using machine learning models, leveraging statistical learning algorithms to analyze climatic, soil, and resource datasets.

Methods

We utilized machine learning algorithms and statistical models for crop yield forecasting. The Kaggle Agriculture Crop Yield dataset, which includes climate, soil, and resource data, formed the basis of this study.

Dataset Acquisition and Preprocessing

- Handling missing values, encoding categorical variables, and scaling.

Exploratory Data Analysis

- Correlation analysis, multicollinearity checks, and feature visualization.

Feature Selection:

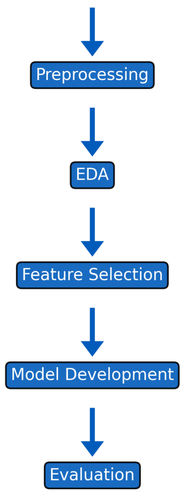
- Techniques like Forward, backward, Random Forest and Lasso Regression to identify key predictors.

Model Development

- Models tested included Lasso, Ridge, Random Forest, and Logistic Regression.

Model Evaluation

- Metrics used included MSPE, Accuracy, Sensitivity, Specificity, and Confusion Matrix.



Data Analysis

content

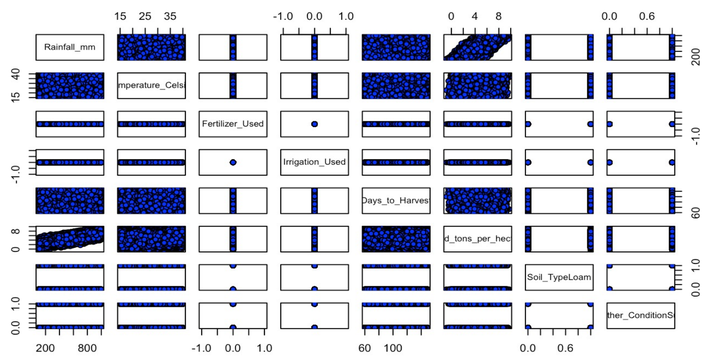
•Insights from EDA:Strong correlation between *Rainfall_mm* and crop yield (r = 0.76).
•Positive association between *Temperature_Celsius* and *Yield_tons_per_hectare* (r = 0.08).
•Outliers detected in *Rainfall_mm* and *Temperature_Celsius*, managed using IQR-based filtering.
•Multicollinearity analysis showed no significant redundancy among variables (VIF < 2 for all features).

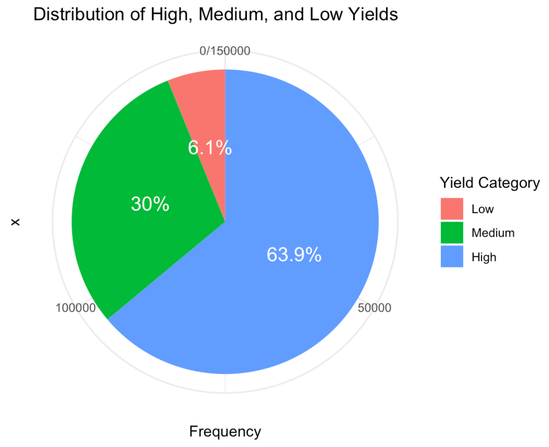
Distribution Analysis

Majority of the data falls within the medium-yield range (2-4 tons per hectare).

Categorical Variables:

Soil_TypeLoam and CropWheat were most frequent in the dataset.





Model Evaluations and Results

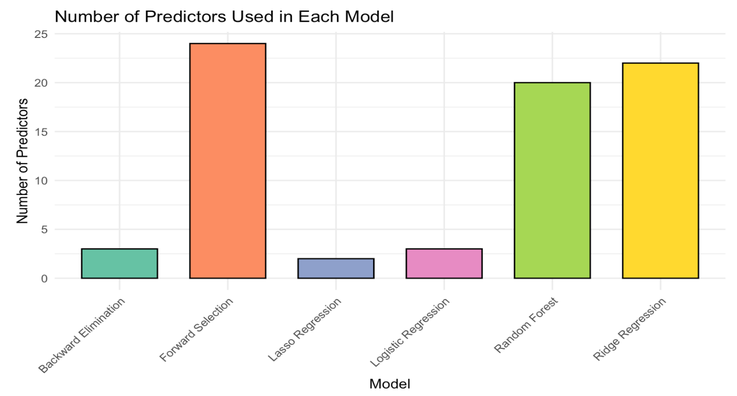
Overview:

After the data analysis phase, various models were evaluated to identify the best-performing algorithm for predicting crop yield. The evaluation was based on key metrics such as **Mean Squared Prediction Error (MSPE)** for regression models and **accuracy**, **sensitivity**, and **specificity** for classification models.

Regression Models and Metrics comparison

Model	MSPE	R-Squared	BIC
Lasso	1.1774	0.5906	360120.5
Ridge	1.1855	0.5905	360130.1
Random Forest	1.2415	0.5655	N/A
Forward	1.1860	0.5905	360156.2
Backward	1.2606	0.5482	360178.6

No of Predictors finalized for each model



Classification Models and Metrics comparison

Metric	Logistic regression
Accuracy	99.99%
Sensitivity	99.99%
Specificity	99.99%
Balanced Accuracy	99.99%
Confusion Matrix	TP:14,999,FP:1, TN:14,998 FN:2

Conclusion

The project "Adaptive Crop Yield Forecasting Using Statistical Learning Algorithms" demonstrated effective crop yield prediction and classification using machine learning models. Key findings include:

Regression Results:

- Lasso Regression achieved the best performance with an MSPE of 1.1774 and the lowest BIC (360120.5), making it the most suitable model for predicting crop yield, Significant predictors: Rainfall_mm and Temperature_Celsius

Classification Insights

- Logistic Regression showed near-perfect metrics (99.99% accuracy, sensitivity, and specificity) due to strong correlations in the dataset. However, this unusually high accuracy suggests the need for further validation on diverse datasets.

Insights

- Rainfall_mm and Temperature_Celsius were the most critical predictors across all models, highlighting their strong influence on crop yield for both regression and classification tasks.

Future work

- validate models on additional datasets and explore advanced algorithms for broader applicability.

References

1. Islam, R., & Masum, A. K. M. (2024). *Crop yield prediction through machine learning*. AIMS Agriculture and Food, 9(4), 980–1003.
2.Chen, T., & Wang, L. (2023). *Climate change impact on agricultural land suitability*. arXiv preprint.
3.Dahikar, S. S., & Rode, S. V. (2014). *Agricultural crop yield prediction using neural networks*. IJIR in Electrical, Electronics, Instrumentation, 2(1), 683–686.
4.Kumar, R., et al. (2015). *Crop selection to maximize yield using ML techniques*. Int. Journal of Computer Applications, 116(23), 1–5.
5.Tejeswani, P., et al. (2024). *Effect of climate change on crop yield via ML*. SSRN Electronic Journal
6.. Iqbal, N., et al. (2024). *Wheat-yield prediction under climate change scenarios*. SSRN Electronic Journal.

University at Buffalo The State University of New York

EAS 503 group project – team003
- vamsisai@buffalo.edu
- kshivapr@buffalo.edu
- vaishakm@buffalo.edu