Review of: "Adoption of Machine Learning Methods for Crop Yield Prediction-based Smart Agriculture and Sustainable Growth of Crop Yield Production – Case Study in Jordan"

Ekta Hooda

Potential competing interests: No potential competing interests to declare.

Adoption of Machine Learning Methods for Crop Yield Prediction-based Smart Agriculture and Sustainable Growth of Crop Yield Production – Case Study in Jordan

The paper explores the adoption of nine machine learning algorithms for the purpose of crop yield prediction and their potential to support sustainable growth in Jordan. The study provides a thorough comparison between the various techniques and uses extensive datasets from reliable sources.

- The authors have mentioned in the first objective that these predictions are helpful to farmers but never explained the “how.” Please add that too.
- The authors should mention if there were any missing values and, if there were, how they were handled. Also, they should account for outliers in the dataset.
- Future implications of the work should have been discussed.
- The results indicate that Random Forest and Lasso regression performed the best with the lowest MSE and R-squared values.
- The techniques used in the article cannot be called Machine Learning algorithms. Some of them are ML, but others are not.
- The models have underlying assumptions. Have they been checked or tested? Has the data been checked for a trend initially before choosing the techniques?
- On page 6/22, it's written after point 9 that ANN outperformed all models, but previously, ANN hasn’t been mentioned anywhere, or is it a reference? Then it should be mentioned and written accordingly.
- The study uses yearly data from 1999 to 2019, i.e., 20 observations for the year. Isn't it a small dataset for ML models? Please look for literature and check the minimum number of observations to use these models.
- For various models under section 5, Training ML models, there is no explanation of the equation variables. Also, the equations should have been written using the original variables under study, not the random 1, 2, 3…
- In figure 3, please mention the title of the y-axis as well and explain the cause of the difference in the pattern of lentil, chickpea, and vetch.
- In table 3, the adjusted R-square should also be included. That might impact the results.
- In 14/22, the second paragraph, mention clearly which model has 99.4% variation. The discussion could be improved. It needs to be thoroughly revised.

- Figure 6 is based on which model? Please mention.