An Experimental Analysis of Machine Learning Algorithms for Maize Yield Prediction

The research question (RQ)

What machine learning technique is suitable for maize yield prediction?

How we address the RQ?

Data used

Table 1: Overview of three datasets used.

Databases	Predictor categories	Data size	Provenance
Crop yield prediction	climate, year, and pesticide	4121	Kaggle
Cover crop and irrigation impacts on weeds and maize yield	cover crop type, irrigation, weed quantity, and water stress characteristics	240	Zenodo
Marked impacts of pollution mitigation on crop yields in China	climatic parameters, and pollution factors	975	Zenodo

Models used

Thirteen models were designed using 70% training data and 30% test data.

- **Classical learning**: SVM, KNN, LR, RR, LASSO, decision trees DT;
- **Ensemble learning:** AdaBoost, XGBoost, GBR, light GBM, ERT, RF, and BR.

Evaluation Metrics

- Rsquare, EVS, MAE, MSE, RMSE, and ME.

Important variables

- Variables tested by permutation techniques.

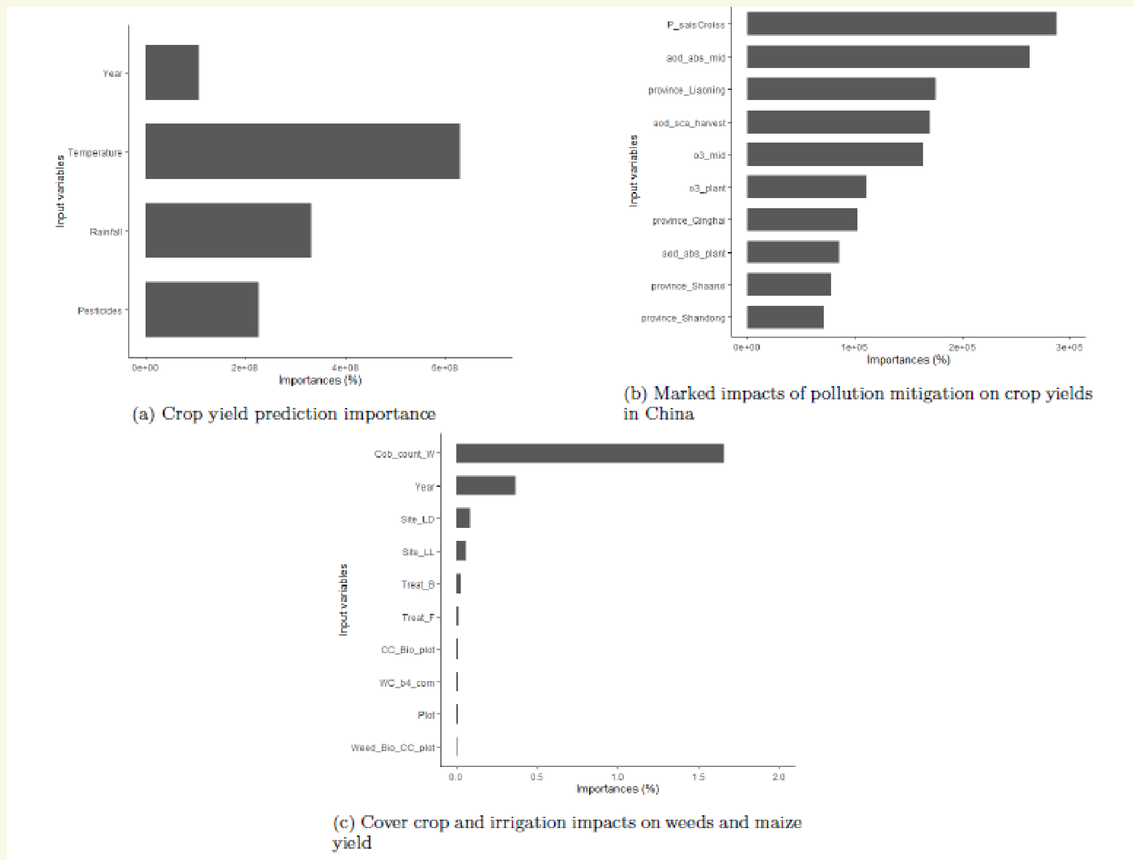
What we found?

Maize yield modeling

Table 2: Performance of the ML models using different evaluation metrics.

Models	Dataset	MAE	MSE	EVS	RMSE	Rsquare	ME	Time
LR	Cover crop and irrigation	0.230	0.094	0.923	0.307	0.920	0.974	0.005
Lasso		0.243	0.105	0.912	0.324	0.911	1.117	0.541
SVR		0.265	0.133	0.894	0.364	0.888	1.289	4.296
KNN		0.934	1.217	0.095	1.103	-0.030	2.380	0.557
RR		0.230	0.094	0.923	0.307	0.920	0.974	0.564
DT		0.361	0.220	0.814	0.469	0.814	1.233	38.123
RF		0.262	0.117	0.902	0.342	0.901	1.097	879.671
GBR		0.250	0.118	0.903	0.344	0.900	1.018	8920.403
GBM		0.300	0.146	0.877	0.382	0.877	1.181	100.562
XGB		0.230	0.088	0.927	0.297	0.925	0.981	2547.399
ADB		0.307	0.153	0.872	0.391	0.871	1.034	329.119
BR		0.319	0.165	0.863	0.407	0.860	1.292	0.913
ERT		0.202	0.074	0.939	0.272	0.937	1.020	324.491

Important variables



(a) Crop yield prediction importance

(b) Marked impacts of pollution mitigation on crop yields in China

(c) Cover crop and irrigation impacts on weeds and maize yield

Figure 1: Important model variables depending on data

Perspectives

ERT model can easily be used to better predict maize yield.

Bibliography

T. R. Shah, K. Prasad, and P. Kumar. Maize—a potential source of human nutrition and health: A review. Cogent Food and Agriculture, 2, 2016. ISSN 23311932. doi: 10.1080/23311932.2016.1166995.

G. P. Souand Tahi[1], Castro G. Hounmènou[1] V. Ratheil Houndji [2,1] & L. Romain Glèlè Kakaï[1],
souandtahi@gmail.com
Laboratoire de Biomathématiques et Estimations Forestières