Explainable artificial intelligence for field trial analysis

Project Goal

Artificial intelligence is nowadays widely used in the agricultural industry, such as field condition management, yield prediction, pest control and weed control. However, the biggest criticism of an AI application is the black-box process involved in the methodology and it's lack of inferential capabilities.

The aim of the project was to develop a general, transparent framework to interpret the results from machine learning models, numerically and graphically. This framework can provide more valuable insights within the analysis process than traditional statistical methods and can identify patterns, which potentially improve agricultural practice.

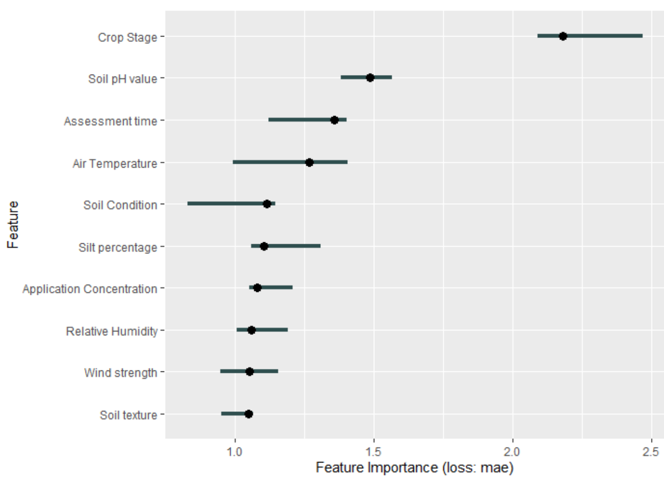
Methods

We focused on the interpretability of the feature importance, the marginal effect for the individual feature and the interaction effect between different features. In order to accomplish the task, the R package iml, lime and DALEX were employed.

Feature importance

The first question from the product manager is "What factor has the strongest influence on the product?". To answer this question, we measure the feature importance by calculating the increase in the model's prediction error after permuting the feature.

The feature importance value describes, when the factor is randomly distributed, how does that affect the total model performance. If the value is close to zero it means that the factor has almost no influence on the model. The higher the value, the more important the factor is.



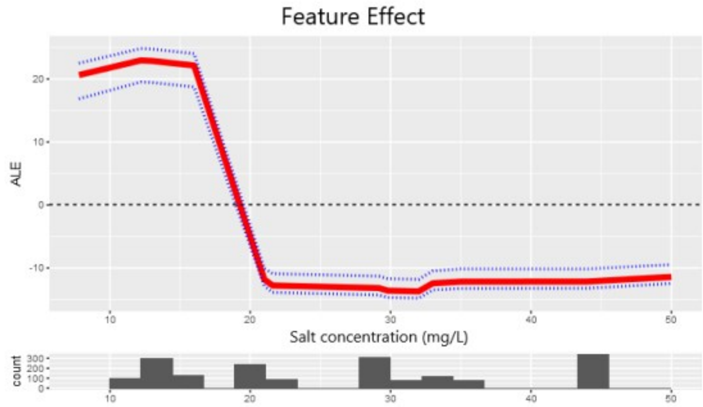
MARGINAL EFFECT

Accumulated Local Effects (ALE) are used for describing the marginal effect, which is the relationship between the target variable and the feature.

The ALE value means: Conditional on a given value, what is the relative effect of changing the feature on the prediction? It is normalized, the value is the difference to the mean prediction.

ALE plot is unbiased — that means it still works when features are correlated. Furthermore, it is fast to compute compared to other methods like partial dependence plot.

In that way, the effect of a particular factor can be evaluated at different levels.

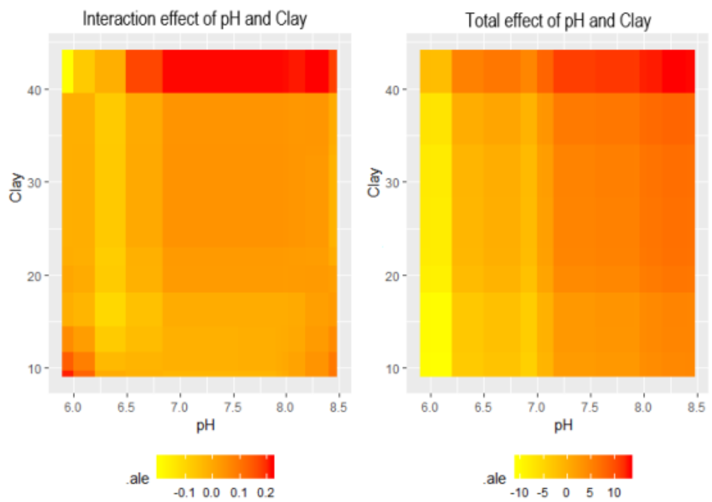


INTERACTION EFFECT

The left heatmap shows the interaction effect (second order effect) of two variables — so how is the effect of one variable on the target influenced by another one.

The right heatmap shows the total effect (interaction effect + individual main effect) of two variables, which gives information about which combination of factor levels lead to good or bad agricultural performance.

By putting these two heatmaps together, the user can identify the high-interacted region and the effect scale between interaction and main effect.



PROJECT OUTCOME

We successfully developed an analysis process for AI model explanation. This process can be applied in various machine learning models such as XGBoosting, random forest, linear model and logistic regression. The explanation is intuitive, easy to understand and the graph can provide further insights from the model. The product manager can make decisions or modify the business strategy based on the gained insights.