



# ERKLÄRUNG KÜNSTLICHER INTELLIGENZ FÜR DIE ANALYSE VON FELDVERSUCHEN

## PROJEKTZIEL

Künstliche Intelligenz ist heute in der Landwirtschaft weit verbreitet, z.B. für das Zustandsmanagement von Feldern, die Ertragsvorhersage sowie Analysen zur Schädlings- und Unkrautbekämpfungen. Die größten Kritikpunkte an KI-Anwendungen sind jedoch häufig, dass die Modelle wie eine Black-Box funktionieren und nicht für neue Daten verallgemeinert werden können.

Ziel des Projekts war die Entwicklung eines allgemeinen, transparenten Frameworks für die analytische und grafische Interpretation der Ergebnisse von Machine Learning Modellen. Dieses Framework kann wertvollere Einblicke in den Analyseprozess liefern als traditionelle statistische Methoden und kann Muster erkennen, welche die landwirtschaftliche Praxis potenziell verbessern.

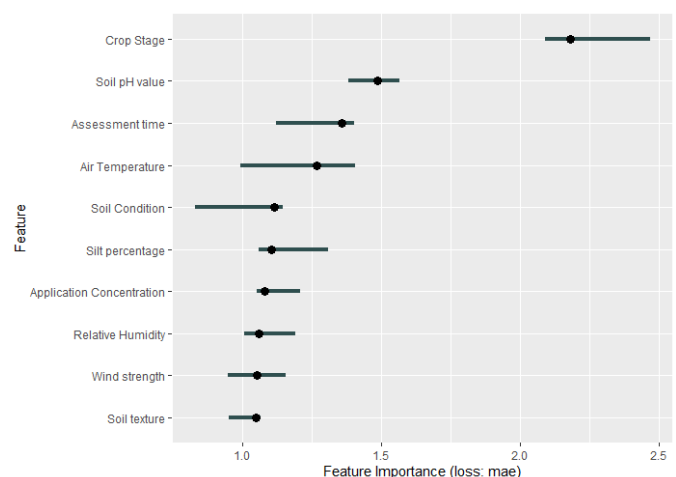
## ANGEWANDTE METHODEN

Wir konzentrierten uns auf die Interpretierbarkeit der Feature Importance, den Effekt der einzelnen Features und die Interaktionseffekte zwischen verschiedenen Features. Um dies umzusetzen, wurden die R-Pakete `iml`, `lime` und `DALEX` eingesetzt.

## FEATURE IMPORTANCE

Die erste Frage eines Produktmanagers lautet: "Welcher Faktor hat den stärksten Einfluss auf das Produkt?" Um diese Frage zu beantworten, messen wir die Feature Importance des Modells. Dies geschieht durch die Berechnung der Zunahme des Vorhersagefehlers ohne das jeweilige Feature.

Der Wert für die Feature Importance beschreibt, bei zufälliger Verteilung des Faktors, wie er sich auf die Gesamtleistung des Modells auswirkt. Wenn der Wert nahe Null liegt, bedeutet dies, dass der Faktor fast keinen Einfluss auf das Modell hat. Je höher der Wert, desto wichtiger ist der Faktor.

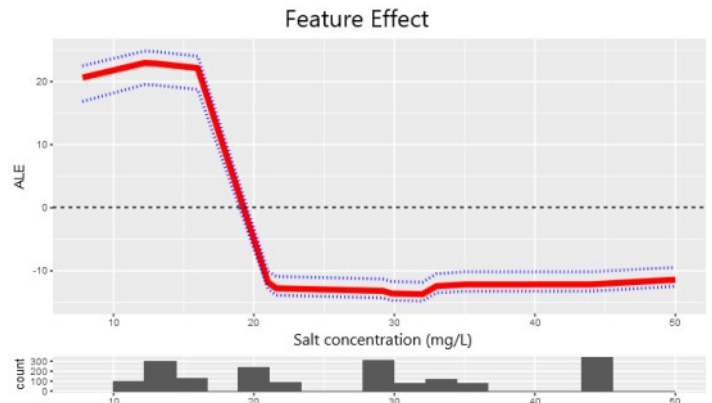


### MARGINALER EFFEKT

Accumulated Local Effects (ALE) werden zur Beschreibung des marginalen Effekts verwendet, der das Verhältnis zwischen der Zielvariablen und dem Feature darstellt.

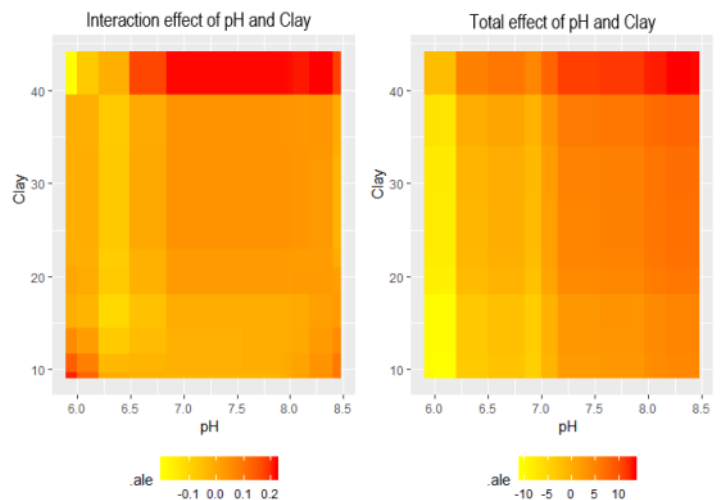
Der ALE-Wert beantwortet: Bedingt durch einen gegebenen Wert, was ist der relative Effekt einer Änderung des Features auf die Vorhersage? Der Wert ist normalisiert und zeigt die Differenz zur mittleren Vorhersage.

Die ALE-Darstellung ist unverzerrt - das heißt, sie funktioniert noch, wenn die Features korreliert sind. Außerdem ist sie im Vergleich zu anderen Methoden wie der partiellen Abhängigkeitsdarstellung schnell zu berechnen. Auf diese Weise kann die Wirkung eines bestimmten Faktors auf verschiedenen Ebenen bewertet werden.



### INTERAKTIONSEFFEKT

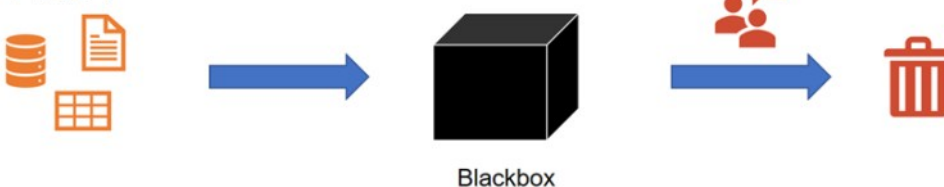
Die linke Heatmap zeigt den Interaktionseffekt (Effekt zweiter Ordnung) zweier Variablen. Er zeigt wie der Effekt eines Features auf die Zielvariable durch ein anderes Feature beeinflusst wird. Die rechte Heatmap zeigt den Gesamteffekt (Interaktionseffekt + individueller Haupteffekt) von zwei Variablen, was Aufschluss darüber gibt, welche Kombination von Faktorstufen zu einer guten oder schlechten landwirtschaftlichen Leistung führt. Wenn diese beiden Heatmaps zusammengesetzt werden, kann der Benutzer stark interagierende Regionen und die Wirkungsskala zwischen Interaktions- und Haupteffekt identifizieren.



### PROJEKTERGEBNIS

Es wurde erfolgreich ein Analyseprozess für die Erklärung künstlicher Intelligenz entwickelt. Dieser Prozess kann in verschiedenen Machine Learning Modellen wie XGBoosting, Random Forests, linearen Modellen und logistischen Regressionen angewandt werden. Die Erklärung ist intuitiv und leicht verständlich, und die Grafiken können weitere Erkenntnisse aus dem Modell liefern. Der Produktmanager kann nun auf Grundlage der gewonnenen Erkenntnisse Entscheidungen treffen und sie in die Geschäftsstrategien einfließen lassen.

#### Früher



#### Heute

