SAHaRA II · BORNO STATE, NIGERIA

Evaluation unter Zeitdruck

Die Verknüpfung von Data Science, Machine Learning und reproduzierbaren Workflows in einer humanitären Wirkungsstudie

Dr. rer. pol. Eduardo Wirthmann Ferreira
movimentar GmbH

movimentar
PROJECT MANAGEMENT & DATA SCIENCE

SAHaRA II • BORNO STATE, NIGERIA

Von humanitären Bedarfen zu mehr Resilienz

Vereinfachte Interventionslogik

Integrierte humanitäre Hilfe in schwer zugänglichen Gebieten

HUMANITÄRE AUSGANGSLAGE

Konflikt, Vertreibung und anhaltende Krisen führen zu:

- Hohe Vulnerabilität von IDPs, Rückkehrenden und Aufnahmegemeinden
- Ernährungsunsicherheit und Druck auf Lebensgrundlagen
- Eingeschränkter Zugang zu sicherem Wasser und Sanitärversorgung (WASH)
- Schutzrisiken und unsichere Lebensbedingungen

INTEGRIERTE SAHaRA II INTERVENTION

Umgesetzt von einem Konsortium von Partnerorganisationen

Ernährungssicherung & Lebensgrundlagen (FSL)
- Cash-/Nahrungsmittelhilfe
- Unterstützung von Lebensgrundlagen
- Marktbezogene Maßnahmen

WASH
- Wasserinfrastruktur
- Sanitärversorgung
- Hygieneförderung

Protection
- Schutzmaßnahmen
- Sensibilisierung (z. B. SGBV)
- Gemeindebeteiligung
- Unterstützungs- und Überweisungswege

ERGEBNISSE (OUTCOMES)

Verbesserter Zugang, Nutzung und Sicherheit für betroffene Menschen

- **Verbesserte Ernährungssicherheit und weniger negative Coping-Strategien**
- **Besserer Zugang zu Wasser, Sanitärversorgung und Hygiene sowie sicherere Nutzung**
- **Stärkung von Protection und Sicherheit für besonders schutzbedürftige Personen**

LANGFRISTIGE WIRKUNG

Beitrag zu

Höhere Resilienz und verbessertes Wohlbefinden

für von der Krise betroffene Gemeinden in Borno State

ZENTRALE INDIKATOREN IN DIESER ANALYSE

- Food Consumption Score (FCS)
- Coping Strategy Index (CSI)
- Ausgewählte WASH- und Protection-Indikatoren

SAHaRA PARTNERORGANISATIONEN – aktuelle Konsortiumskonstellation

Caritas Nigeria · NEYIF · Caritas Germany · CAFOD Catholic Agency for Overseas Development · JOPI Maiduguri Diocese · CRS Catholic Relief Services

Mit Unterstützung der EU

Kofinanziert von der Europäischen Union Humanitäre Hilfe

Hinweis: Dies ist eine vereinfachte Interventionslogik für Präsentationszwecke und nicht die vollständige Theory of Change des Projekts.
Die dargestellte Interventionslogik bezieht sich auf SAHaRA II; die Logos zeigen die aktuelle Konsortiumskonstellation.

Die Panel-Frage: Schnell oder schlampig?

Kurzes Evaluations-Zeitfenster

Jahre operativer Daten

Hohe Erwartungen an Evidenz

OECD/DAC-Kriterien strukturieren das Urteil — sie ersetzen kein Forschungsdesign.

Evaluatives Urteil

Relevanz · Kohärenz · Effektivität · Effizienz · Impact · Nachhaltigkeit

Evidenzbasis

Qualitative Evidenz · Monitoring · Kontext · Quantitative Wirkungsanalyse

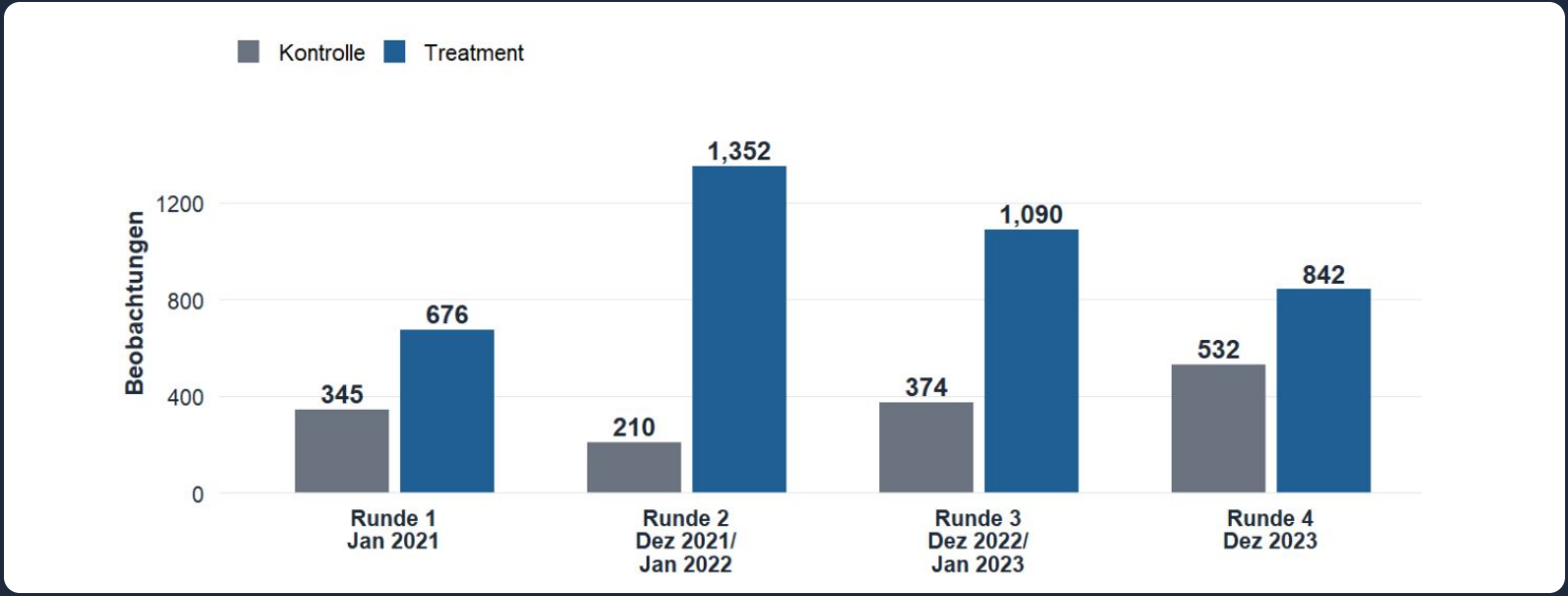
Kernerkenntnis

Tempo entsteht nicht durch weniger Prüfung, sondern durch weniger Rekonstruktion.

DATENBASIS & DESIGN

SAHaRA II: Von der Gesamtevaluation zur Wirkungsanalyse

5.421 Beobachtungen im zusammengeführten Analysedatensatz · wiederholte Querschnitte, kein Haushaltspanel



METHODIK & ANALYSE

Der tatsächliche Analyse-Stack

01 **1. Datenaufbereitung & EDA**
Cleaning, Harmonisierung, Verteilungen, Missingness

02 **2. Random Forest**
datengetriebenes Prädiktor-Screening bei sehr vielen Variablen

03 **3. PSM / Matching**
Propensity Scores, Nearest-Neighbour-Matching (Ähnliche Treatment- und Vergleichsfälle zuordnen · Vergleichbarkeit prüfen)

04 **4. DiD / Regression**
Treatment vs. Vergleich im Laufe der Zeit / Exposition

05 **5. Sensitivität & evaluatives Urteil**
alternative Spezifikationen, Triangulation, Grenzen

💡 **Merksatz:** Machine Learning hat das Screening unterstützt. Die Wirkungsaussage wurde aus dem Vergleichsdesign, der statistischen Modellierung und dem evaluativen Urteil getroffen.

METHODIK & SAMPLING

Wie entstanden Treatment und Vergleich?

01 1. Stichprobe im Dorf

- Dörfer als operative Auswahleinheit
- Haushaltszahl ungefähr proportional zur geschätzten Dorfbevölkerung
- Zufallsbewegung ab einem zentralen Punkt
- **Interview mit jedem dritten Haushalt**, bis die Sollzahl erreicht war
- **Ziel: systematische Erfassung**

02 2. Vulnerabilität & Targeting

Needs-Assessment-Indikatoren flossen in Vulnerabilitätsindex ein:

- Food Security (FCS)
- Coping Strategy Index (CSI)
- Zugang zu Wasser und Seife
- Schutzindikatoren

💡 **Kernprinzip:** Höhere Vulnerabilität ➔ höhere Wahrscheinlichkeit der Projektpriorisierung.

METHODIK & ANALYSE

Wie wurde beobachtete Vergleichbarkeit geprüft?

01 1. Propensity Score Matching (PSM)

Schätzung der Treatment-Wahrscheinlichkeit anhand beobachteter Merkmale.

Prozessschritte:

1. Propensity Scores schätzen
2. Ähnliche Treatment- und Vergleichsbeobachtungen einander zuordnen
3. Vergleichbarkeit nach Matching prüfen
4. Nicht gematchte Beobachtungen dokumentieren

02 2. Diagnostik & Kriterien

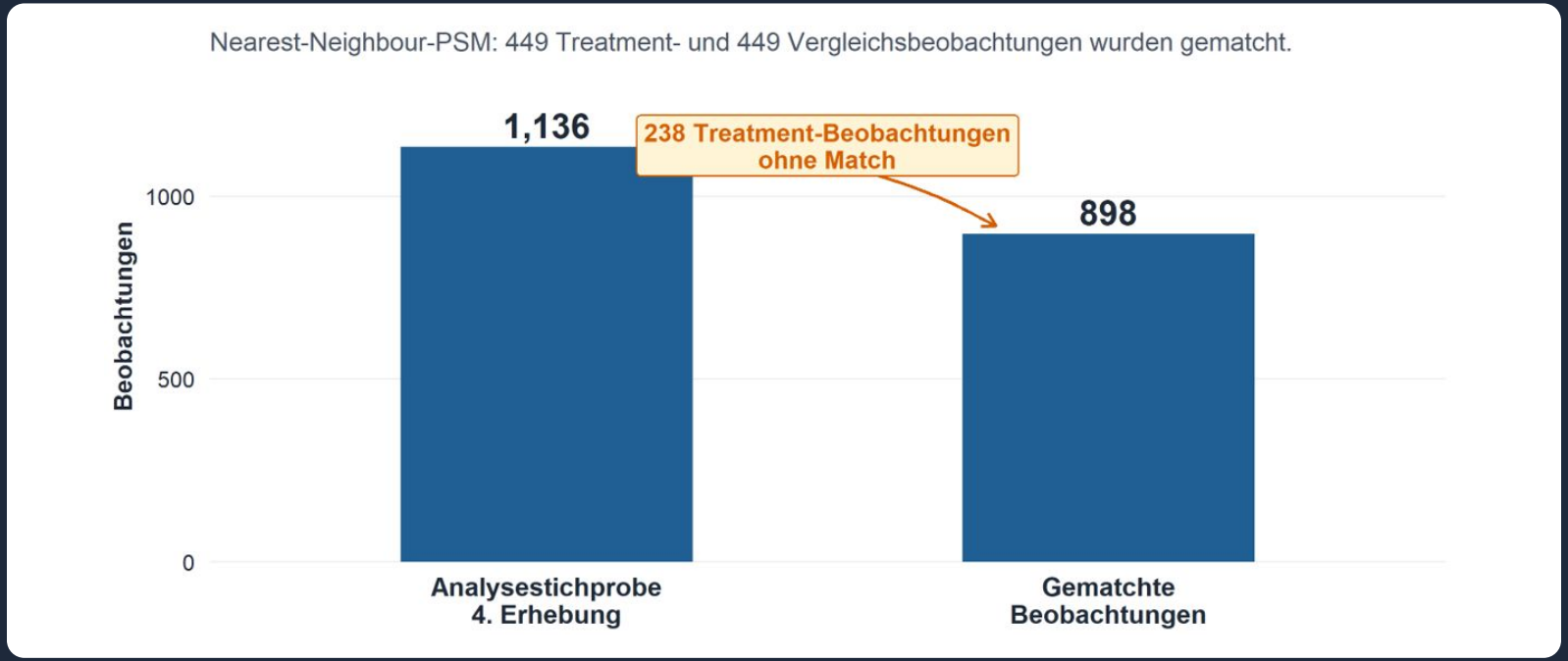
Folgende Diagnoseparameter wurden systematisch ausgewertet:

- Standardisierte Mittelwertdifferenzen
- Varianzverhältnisse
- Unterschiede in den empirischen Verteilungen
- Zahl gematchter und nicht gematchter Fälle

⚠ **Propensity Score Matching ≠ Randomisierung:**
Es verbessert die Vergleichbarkeit nur hinsichtlich der **beobachteten und modellierten Merkmale**.

METHODIK & ANALYSE

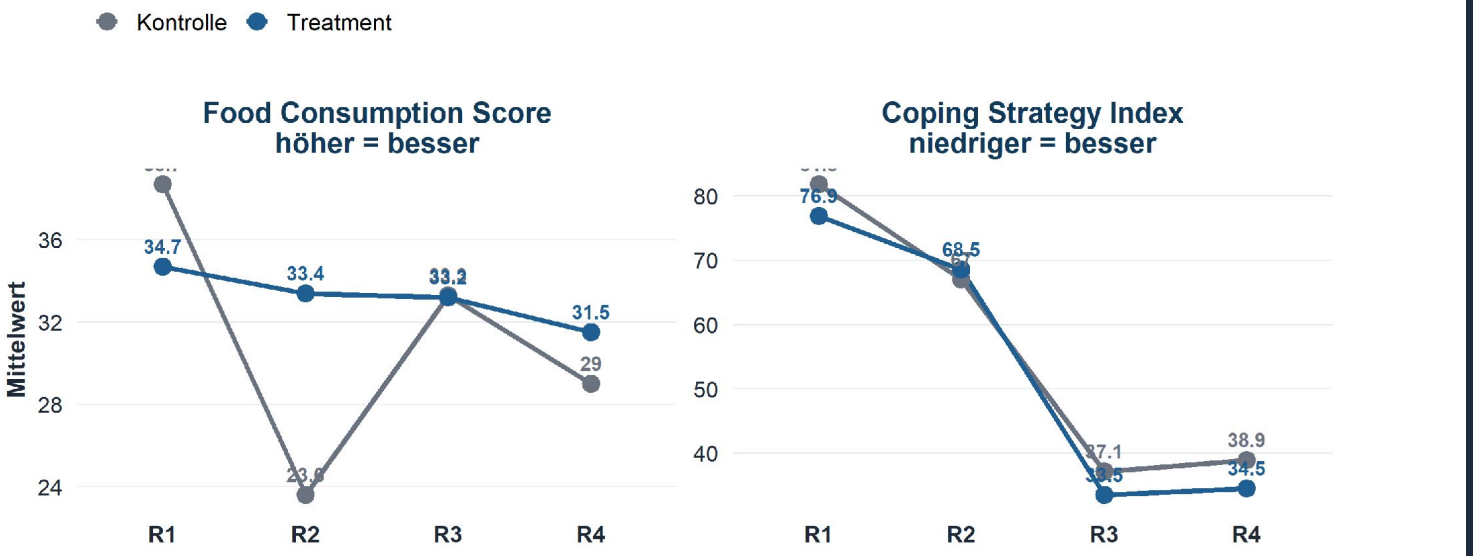
Für wen fanden wir einen vergleichbaren Match?



ERGEBNISSE & SIGNALE

Zwei Hauptsignale: FCS stabiler, CSI klarer (ungematchte Stichproben)

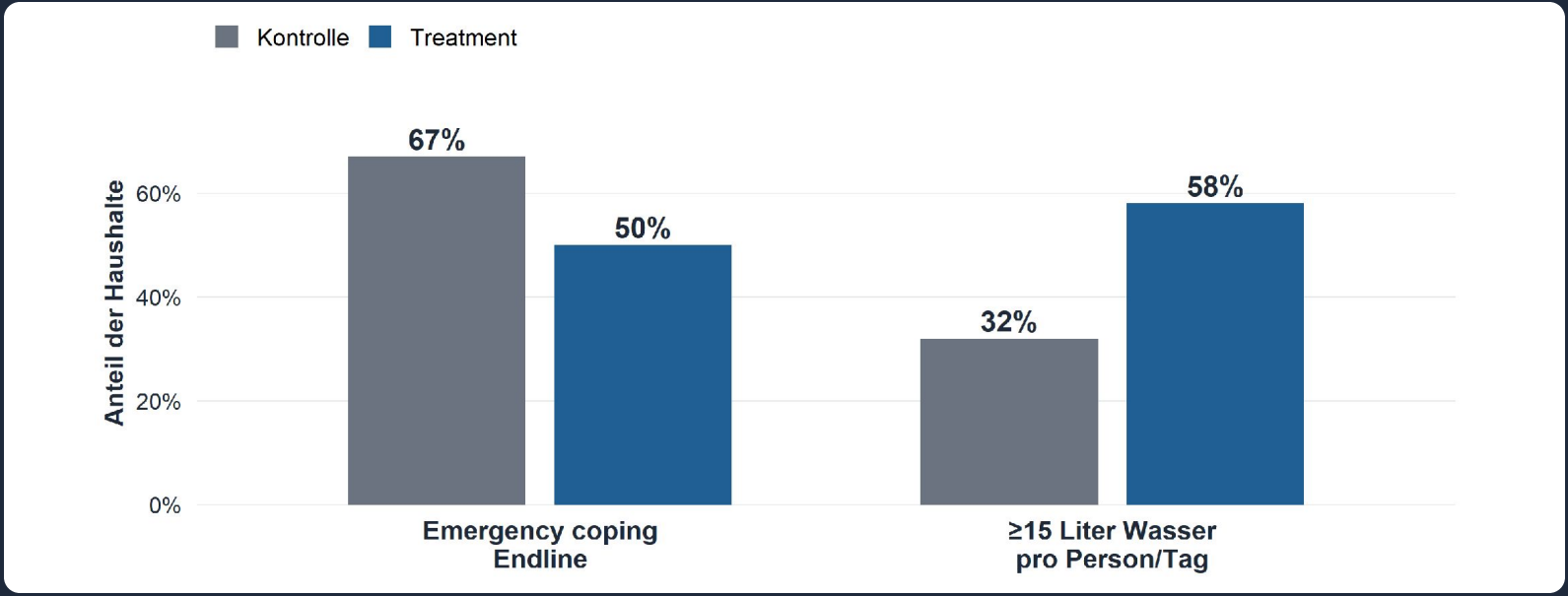
Deskriptive Mittelwerte der jeweiligen Erhebungsrunden · ungematchte Stichproben · wiederholte Querschnitte



ERGEBNISSE & SIGNALE

Deskriptiver Endline-Vergleich · ungematchte Treatment- und Vergleichsgruppe

Endline-Beispiele: hier deskriptiv, nicht als isolierter kausaler Effekt.



METHODIK & ANALYSE

Robustheit: Wie stabil ist die Interpretation über verschiedene Ansätze?

1 · Deskriptiv

- Verläufe über mehrere Erhebungen
- Endline: Treatment vs. Vergleich
- zeigt Richtung und Größenordnung

keine isolierte Kausalaussage

2 · Matching

- PSM auf der 4. Erhebung
- Nearest-Neighbour-Matching

449 Treatment + 449 Vergleich gematcht

238 Treatment-Beobachtungen ohne Match

- Vergleichbarkeit nach Matching geprüft

3 · DiD / Regression

- Veränderungen über die Zeit vergleichen
- Treatment vs. Vergleich
- unterschiedliche Exposition / Implementierungsphasen berücksichtigen
- Wirkungsschätzung unter Modellannahmen

Robustheit heißt nicht: dieselbe Zahl.
Sondern: Hält die Interpretation über plausible Spezifikationen hinweg?

METHODIK & ANALYSE

Wie funktioniert Random Forest?

Viele Entscheidungsbäume – eine gemeinsame, robustere Vorhersage

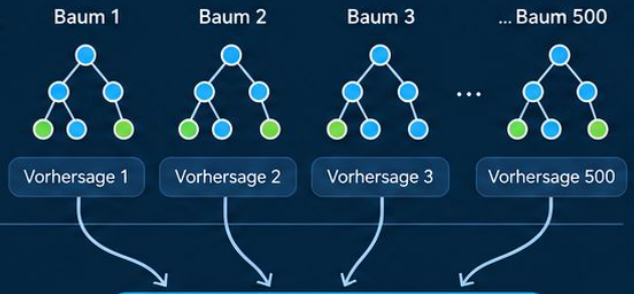
1 Zufällige Teilstichproben

Für jeden Baum wird aus den Daten zufällig eine Teilmenge von Haushalten und Variablen verwendet.

→ Verschiedene Teilstichproben der Daten

2 Viele Entscheidungsbäume

Jeder Baum findet Muster und teilt die Daten Schritt für Schritt in Gruppen.



3 Gemeinsame Vorhersage

Die Vorhersagen aller Bäume werden kombiniert. Dadurch werden zufällige Muster ausgeglichen, und die Vorhersage wird stabiler.

Endgültige Vorhersage
(z. B. Food Consumption Score)

Wichtigste Prädiktoren für den Food Consumption Score (FCS)

Variable Bedeutung im Random-Forest-Modell
(illustrative Rangfolge, keine numerischen Werte)

Hinweis: Darstellung zeigt eine illustrative Rangfolge der wichtigsten Variablen. Keine numerischen Importance-Werte.

Random Forest hilft uns, aus vielen verfügbaren Variablen die wichtigsten Prädiktoren zu identifizieren.
Es zeigt, welche Faktoren den FCS am besten vorhersagen – es beweist keine Kausalität.

ERGEBNISSE & SIGNALE

Was bleibt nach Matching und Modellierung?

Ernährungssicherheit

- **FCS:** günstiges Signal bleibt auch in den adjustierten Modellen bestehen
- Treatment-Implementierung positiv mit FCS verbunden

Coping

- **CSI/LCSI:** insgesamt günstigere Coping-Muster, aber weniger eindeutig als FCS
- Weniger schwere Coping-Strategien in Treatment

WASH

- **Wasserzugang:** besonders konsistentes positives Signal
- **58 % vs. 32 %** mit ≥15 L/Person/Tag an der Endline

Protection

- **Sicherheit:** positive Signale, aber heterogener
- Verbesserungen nicht über alle Protection-Indikatoren gleich stark

Gesamtbild: konsistente Hinweise auf einen positiven Projektbeitrag – am stärksten bei **Ernährungssicherheit** und **WASH**, differenzierter bei **Coping** und **Protection**.

Grenzen: Beobachtungsdaten · keine Randomisierung · verbleibende Imbalance nach PSM · wechselnde Stichproben · unbeobachtete Einflussfaktoren möglich

WORKFLOW & EFFIZIENZ

Durch Reproduzierbarkeit lässt sich Zeit sparen

Was wird reproduzierbar?

- Datenerhebung und Harmonisierung
- Cleaning- und Recoding-Regeln
- Indikatorberechnung
- Diagnostik und Modelle
- Grafiken und Bericht

Was spart das?

- **Tempo:** neu rechnen statt neu bauen
- **Qualität:** dieselben Regeln über mehrere Runden
- **Transparenz:** Annahmen und Transformationen werden sichtbar
- **Institutionelles Gedächtnis:** Wissen bleibt im Workflow

Kernerkenntnis:
Automatisierung verwandelt einmaligen Aufwand in dauerhaften Zeitgewinn und Verlässlichkeit.

OPEN SOURCE & TRANSPARENZ

Open Source: Transparenz wird überprüfbar

Ein öffentlicher Code ermöglicht es:

- Transformationen nachzuvollziehen
- Annahmen zu prüfen
- Modelle neu laufen zu lassen
- Fehler oder Inkonsistenzen zu finden
- Alternativen zu testen
- Komponenten weiterzuverwenden

Code & Dokumentation

Vollständiger Quellcode sowie Daten-Transformationen stehen zur unabhängigen Überprüfung zur Verfügung:

Repository Link:
https://github.com/movimentar/impact-analysis-sahara

Aber: Öffentlicher Code macht eine Analyse nicht automatisch richtig — und eine Evaluation nicht automatisch unabhängig.

EVALUATIONSMODELL

Kontinuierliche Evaluierbarkeit ≠ kontinuierliche Evaluation

Traditionelles Muster

- Implementation
- Monitoring-Daten sammeln
- Evaluator kommt
- rekonstruiert & analysiert
- berichtet

Evaluation-ready

- Evaluationslogik
- Wiederholte Daten
- QA + Code
- Evidenz wächst
- Evaluation + Triangulation + Urteil

Der Unterschied: Die Evaluation beginnt nicht bei null.

FAZIT & SCHLUSSFOLGERUNG

Schluss: schneller durch frühere Evaluierbarkeit

Daten & Methoden

- Wiederholte operative Daten können evaluativ wertvoll werden.
- Quasi-experimentelle Methoden stärken die Wirkungsargumentation — wenn Grenzen sichtbar bleiben.
- Random Forest hilft bei Screening und Diagnostik, nicht bei kausaler Entscheidung.

Code & Transparenz

- Reproduzierbarer Code verkürzt Analysezeit und macht Entscheidungen überprüfbar.
- Open Source erweitert Transparenz um öffentliche Prüfung und Wiederverwendung.

Kernerkenntnis: Die Antwort auf Zeitdruck ist keine schnellere Evaluation, sondern frühere Evaluierbarkeit.

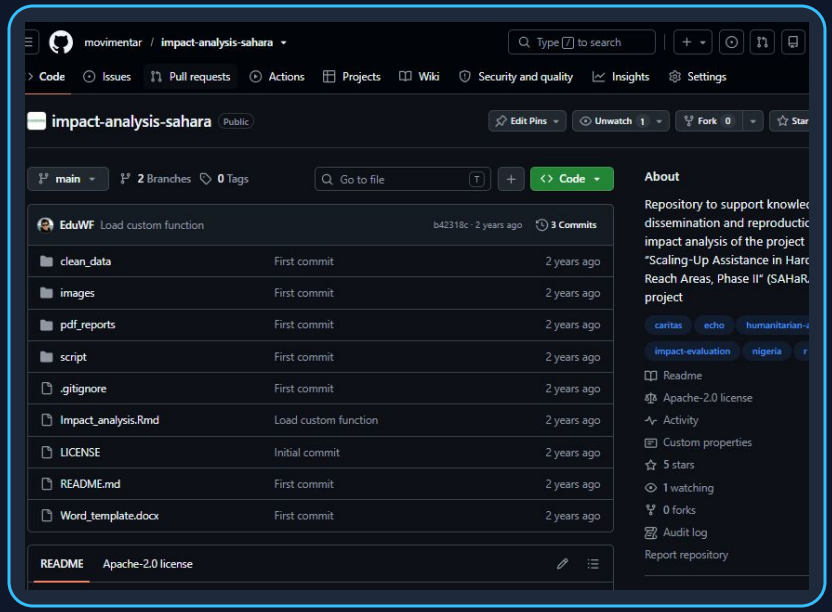
VIELEN DANK

Fragen & Diskussion

Dr. rer. pol. Eduardo Wirthmann Ferreira
movimentar GmbH

contact@movimentar.eu · www.movimentar.eu

GitHub Repository

https://github.com/movimentar/impact-analysis-sahara

Code · anonymisierte Daten · reproduzierbarer Workflow