Ausgangslage und Problemstellung:
Ein Kreditanbieter nutzt ein **maschinelles Lernmodell** mit Informationen über Antragsteller von **Konsumkrediten**, um vorherzusagen, ob sie über einen **Zeitraum von zwei Jahren pünktliche Zahlungen leisten werden**. Die Vorhersage des Modells wird genutzt um zu entscheiden, ob eine antragstellende Person für einen Kredit in Frage kommt oder nicht. Zu den erklärenden Variablen gehören Alter, Einkommen, Vermögen etc.
Es stellt sich heraus, dass ein grösserer Anteil von Anträgen von **Personen mit Migrationshintergrund aufgrund dieser Merkmale abgelehnt werden**.

Aufgabe:
Welche **Ethischen Grundorientierungen** aus dem Ethik Kodex sehen Sie in diesem Fall betroffen, bzw. **verletzt**, und warum? Machen Sie zwei konkrete Vorschläge für die Stärkung der ethischen Grundorientierungen und Umsetzung der prozeduralen Werte in diesem Fall.
Dabei muss nicht zwischen den vier Phasen des Daten-Lebenszyklus' unterschieden werden.

Lösungsvorschlag:
Besonders die Grundorientierung der **Gerechtigkeit** wird in diesem Fall tangiert: Es besteht das Problem von indirekter Diskriminierung, in diesem Fall von Personen mit Migrationshintergrund. Dies ist in diesem Fall besonders relevant, da das erstellte Modell die Entscheidungsfindung beim Kreditanbieter unterstützt.

Empfehlungen

- **Gerechtigkeit**:
  - Es wird sichergestellt, dass der **Trainingsdatensatz vielfältig genug** ist, sodass die indirekte Diskriminierung nicht aufgrund schlechter Trainingsdaten herrührt.
  - Das Modell wird auf indirekte **Diskriminierung** bezüglich sensibler Merkmale der antragstellenden Personen **untersucht** und die Ergebnisse der Audit-Stelle zur Verfügung gestellt.
    Es wird erklärt, welche Formen indirekter **Diskriminierung** aufgrund der Minimierung **ökonomischer Risiken in Kauf genommen werden**.
- zur **Transparenz**: Den antragstellenden Personen
  - wird zu Beginn des Antragsprozesses der **Einsatz des statistischen Modells erklärt.**
  - werden die **Einflussfaktoren** auf die Entscheidung des Modells **offengelegt**
  - wird mit dem **Entscheid automatisch** von einer Software berechnete, individuellen Änderungen an ihrem Profil aufgezeigt, die die Entscheidung des KI-Modells verändert hätten.