

# Projektauftrag: Verlustprävention an Selbstbedienungskassen im Einzelhandel

(nach DASC-PM v1.1)

## 1. Projektbezeichnung

Titel: Betrug an Self Checkout-Kassen

## 2. Projektbeteiligte

Projektgruppe:

- Raphael Schaffarczyk (Datenanalyse)
- David Zurschmitt (Modellierung & Evaluation)
- Matthias Bald (Dokumentation & Projektkoordination)

Projektgeber: xxxxxxxx GmbH

Betreuer: Prof. Dr. Christian Beecks, Lehrgebiet Data Science an der FernUniversität in Hagen  
Frau Sabine Folz-Weinstein, Max Pernklau

## 2a. Stakeholder & Beteiligte

- Projektteam: Drei Studierende der Hochschule
- Projektgeber: Die xxxxxxxx GmbH, Fachabteilung für Filialsicherheit
- Stakeholder: Marktleiter, IT-Abteilung, Geschäftsführung
- Nutzer: Sicherheitsverantwortliche, Filialleitung

Optional: Stakeholder-Matrix wird im Anhang ergänzt.

## 2b. Projektorganisation

Das Team arbeitet nach dem DASC-PM-Modell iterativ. Wöchentliche Abstimmungen im Team, zweiwöchentliche Meetings mit der Betreuerin Frau Folz-Weinstein und Präsentation und Abgabe der Artefakte je nach Phase.

Rollenverteilung: Raphael – EDA, David – Modellierung, Matthias Bald – Kommunikation und Doku.

Dokumentation via OneDrive, Versionierung über GitHub privat.

## 3. Ausgangslage & Projektidee

Immer mehr Einzelhändler setzen auf Selbstbedienungskassen (SBK), um Personalressourcen zu sparen. Gleichzeitig steigen jedoch die Diebstahlzahlen, insbesondere durch bewusstes Nicht-Scannen von Artikeln.

Die xxxxxxxx GmbH stellt historische Kassendaten zur Verfügung, die im Rahmen des Projekts auf Muster untersucht werden sollen.

Neben vorsätzlichen Diebstahlsversuchen durch Kundinnen und Kunden steht im Rahmen des Projekts auch die allgemeine Verlustprävention an Self-Checkout-Kassen im Fokus. Diese umfasst unter anderem unbeabsichtigte Fehlvorgänge wie das versehentliche Nicht-Scannen von Artikeln, technische Probleme bei der Artikelerfassung oder Bedienfehler. Ziel ist es daher nicht nur, potenziellen Diebstahl zu erkennen, sondern auch systematisch Ursachen für Kassier-Abweichungen zu identifizieren und darauf basierend Verbesserungspotenziale für Prozesse und Technik aufzuzeigen.

#### 4. Ziel des Projekts

Ziel ist es, ein Machine-Learning-Modell zu entwickeln, das mit hoher Genauigkeit potenziell fehlerhafte Transaktionen erkennt.

Dabei sollen typische Merkmale auffälliger Einkäufe identifiziert und bewertet werden.

#### 5. Fachlicher Hintergrund (Domain-Analyse)

Im Einzelhandel entstehen jährlich Verluste in Milliardenhöhe durch Inventurdifferenzen.

Studien zeigen, dass SBK-Stationen besonders anfällig sind.

Typische Betrugsmethoden sind z. B. das Scannen günstiger Produkte statt teurerer („Bananen-Trick“) oder das vollständige Auslassen von Artikeln.

Die xxxxxxxx GmbH setzt derzeit auf manuelle Stichprobenkontrollen, was ineffizient ist.

#### 6. Datenbasis

Es stehen 15.000.000 Transaktionen als Trainingsdaten aus den Jahren 2022/2023 bzw.

Testdaten aus dem Jahr 2024 zur Verfügung, davon wurden rund cccccc Transaktionen stichprobenartig kontrolliert.

Die Daten enthalten Zeitstempel, Artikelinformationen, Anzahl gescannter Artikel, Preis, Benutzerkennung sowie ein Label (korrekt/inkorrekt).

Die Daten enthalten keine personenbezogenen Informationen und wurden DSGVO-konform anonymisiert.

#### 7. Methodischer Ansatz

1. Vorverarbeitung & Datenbereinigung
2. Explorative Datenanalyse
3. Feature Engineering (u. a. Warenkorbgröße, Preis pro Artikel, Häufigkeit)
4. Modellierung mittels Random Forest & Isolation Forest
5. Evaluierung über Precision, Recall, F1-Score, ROC AUC
6. Ableitung von Regeln & Modellinterpretation

## 8. Zeitplan / Projektphasen

Meilenstein	Verantwortlich	Frist
Projektauftrag	Matthias Bald	17.04.2025
Datenaufbereitung	Raphael Schaffarczyk	10.04.2025
Identifikation und Evaluation geeigneter Analyseverfahren		
Sicherstellung technischer Umsetzbarkeit		
Abschlusspräsentation	Alle Teilnehmer	08./09.07.2025

## 9. Ressourcen

Alle Teammitglieder verfügen über grundlegende Kenntnisse in Python und Statistik und Maschinellem Lernen mit: Python (pandas, scikit-learn, matplotlib), JupyterLab  
Infrastruktur: Google Colab (GPU), GitHub für Codeverwaltung  
Datenzugriff via gesichertem SFTP

## 10. Risiken & Herausforderungen

- Ungleichverteilung der Klassen (nur 6 % fehlerhafte Transaktionen)
- Modellinterpretierbarkeit muss gewährleistet bleiben
- Zeitliche Begrenzung des Projekts

## 11. Erfolgskriterien

- F1-Score > 0.70 auf Testdaten
- Gute Visualisierbarkeit und Nachvollziehbarkeit der Ergebnisse
- Feedback des Projektgebers zur Praxistauglichkeit