

Projektauftrag

Verlustprävention an Selbstbedienungskassen im Einzelhandel

Team: Raphael Schaffarczyk, David Zurschmitten, Matthias Bald

Auftraggeber: Wertkauf GmbH

Abgabedatum: 21.04.2025

Projektauftrag

Verlustprävention an Selbstbedienungskassen im Einzelhandel

1. Projektbezeichnung

Betrug an Self Checkout-Kassen

Hinweis zur Terminologie:

Der Projekttitel folgt der offiziellen Bezeichnung durch die Wertkauf GmbH. Inhaltlich umfasst das Projekt jedoch sämtliche Ursachen von Verlusten an Selbstbedienungskassen – nicht nur vorsätzliche Betrugsversuche, sondern auch unbeabsichtigte Fehlbedienungen, technische Störungen oder systemische Schwächen. Zur besseren Lesbarkeit wird im weiteren Verlauf teilweise allgemein von *Verlusten* gesprochen, sofern keine explizite Differenzierung notwendig ist.

2. Problemstellung und Ausgangslage

Die Wertkauf GmbH setzt seit mehreren Jahren stationäre Selbstbedienungskassen (SBK) ein, um die Filialprozesse effizienter zu gestalten und das Einkaufserlebnis der Kundinnen und Kunden zu verbessern. Die SBK-Systeme befinden sich im Ausgangsbereich der Märkte; mobile Self-Scanning-Wagen oder Handscanner kommen derzeit nicht zum Einsatz. Artikel mit Gewichtserfassung werden über integrierte Waagen erfasst, ein automatischer Abgleich des Gesamtgewichts mit dem gescannten Warenkorb erfolgt jedoch nicht.

SBK-Systeme gelten als besonders anfällig für Warenverluste, da die Verantwortung für den vollständigen Kassiervorgang vollständig bei der Kundschaft liegt, während Kontrollmaßnahmen nur eingeschränkt stattfinden. Verluste entstehen sowohl durch absichtliches Auslassen von Artikeln (z. B. Diebstahl) als auch durch unbeabsichtigte Fehlbedienungen oder technische Störungen. Eine differenzierte und belastbare Einschätzung der Ursachen und des Umfangs dieser Verluste liegt derzeit nicht vor.

Die bislang durchgeführten Kontrollmaßnahmen in den Filialen der Wertkauf GmbH beschränken sich auf zufällig ausgewählte Stichproben. Diese liefern lediglich punktuelle Erkenntnisse und lassen keine systematische Aussage darüber zu, ob die Maßnahmen tatsächlich geeignet sind, Verluste effektiv zu reduzieren oder wirtschaftlich sinnvoll eingesetzt werden.

3. Projektauftrag und Ziel

Im Rahmen dieses Projekts wurden wir von der Wertkauf GmbH beauftragt, eine datengetriebene Lösung zur Verlustprävention im Kontext von Selbstbedienungskassen zu entwickeln. Ziel ist es, den durch unvollständige oder fehlerhafte Kassivorgänge verursachten betriebswirtschaftlichen Schaden zu verringern.

Dazu sollen auf Basis der von der Wertkauf GmbH bereitgestellten Transaktionsdaten Muster und Zusammenhänge identifiziert werden, die auf potenziell fehlerhafte oder manipulative Abläufe hinweisen. Diese können sich beispielsweise in bestimmten Warengruppen, Zeitfenstern oder typischen Nutzungsverhalten zeigen. Die daraus gewonnenen Erkenntnisse sollen in konkrete Handlungsempfehlungen sowie technische Vorschläge zur Überwachung und Risikobewertung überführt werden.

Ein Kernelement der Lösung ist die Entwicklung eines Algorithmus, der verdächtige Transaktionen kennzeichnet und eine gezielte Nachkontrolle anstoßen kann. Die Bewertung des Algorithmus und seiner Vorschläge erfolgt auf Basis einer definierten Bewertungsfunktion, die im folgenden Abschnitt detailliert erläutert wird.

Die geplanten methodischen Schritte, Analysen und Meilensteine werden im weiteren Verlauf des Projektauftrags beschrieben. Die Zielsetzung umfasst dabei nicht nur die technische Machbarkeit, sondern auch die betriebswirtschaftliche Sinnhaftigkeit der vorgeschlagenen Lösung.

4. Bewertungsfunktion

Die Bewertungsfunktion dient als wirtschaftliche Grundlage für Modellbewertung und Handlungsempfehlungen.

Um die Effektivität möglicher Kontrollstrategien und die Leistungsfähigkeit des entwickelten Algorithmus bewerten zu können, legt die Wertkauf GmbH folgende wirtschaftliche Annahmen zugrunde:

- Eine **nicht durchgeführte Kontrolle bei einem korrekten Einkauf** gilt als betriebswirtschaftlich neutral und wird mit **0,00 €** bewertet.
- Eine **nicht durchgeführte Kontrolle bei einem inkorrekten Einkauf** führt zu einem **wirtschaftlichen Schaden**, der mit dem Wert der entgangenen Ware beziffert werden soll.
- Eine **durchgeführte Kontrolle, die einen inkorrekten Einkauf identifiziert**, wird mit einem **positiven Nutzenwert von +5,00 €** angesetzt.
- Eine **durchgeführte Kontrolle bei einem korrekten Einkauf** wird mit **-10,00 €** bewertet, da hierbei sowohl potenzieller Kundenärger als auch zusätzlicher Personalaufwand berücksichtigt werden.

Die vorliegende Bewertungsfunktion lässt vermuten, dass kleine Warenverluste, die nicht entdeckt werden, lediglich mit kleiner Schadenssumme bewertet werden sollen. Hingegen werden fälschliche Kontrollen bei kleinen Warenwerten überproportional bestraft. Dies könnte dazu führen, dass bei potenziell **kleinen** Schadenssummen **keine Kontrollen** durchgeführt werden. Bei **hochpreisigen** Artikeln wiederum wird eine ausbleibende Entdeckung mit einem hohen negativen Wert (Warenwert) bewertet, sodass eine **Kontrolle** sich eher lohnt.

Wir begrüßen diese betriebswirtschaftliche Perspektive. Aus unserer Sicht ist es jedoch sinnvoll, die Bewertungsfunktion **flexibel** und **kontextsensitiv** zu gestalten. Daher schlagen wir folgende Anpassungen vor:

- **Variable statt fixer Beträge:**

Die von der Wertkauf GmbH vorgeschlagenen Werte können als Standard beibehalten werden, sollten jedoch im Modell parametrierbar sein, um je nach Geschäftsfall angepasst werden zu können. So könnten auch **unterschiedliche Modellvarianten** berechnet werden.

- **Wertabhängige Bewertung des verhinderten Schadens:**

Im Fall 3 könnte ein höherer Wert angesetzt werden, wenn durch das System ein fehlerhafter – sei es bewusst oder unbewusst ausgeführt – Scanvorgang aufgedeckt wird. Ziel ist es, die betreffende Person zu mehr Achtsamkeit zu bewegen bzw. potenzielle Diebstähle künftig zu verhindern. Dabei ist sorgfältig abzuwägen, in welchem Maße ein möglicher Reputationsschaden durch unbegründete Kontrollen schwerer wiegt als der Nutzen der verhinderten Schäden – oder ob im Gegenteil der präventive Aspekt der Diebstahlerkennung überwiegt. Grundsätzlich gilt: Eine gleichzeitige Optimierung sowohl bei der Erkennung von Diebstahl oder fehlerhaften Scanvorgängen als auch bei der Vermeidung irrtümlicher Kontrollen ist nicht vollständig möglich – ein Zielkonflikt bleibt bestehen.

Die Ausgestaltung des weiteren Bewertungsmodells möchten wir gerne in enger Abstimmung mit Ihnen als Kunde vornehmen.

Unsere bisherigen Vorschläge sowie weiterführende Details sollen in der anschließenden Datenanalyse und Modellentwicklung gemeinsam mit Ihnen konkretisiert werden. So kann flexibel auf neue Erkenntnisse aus den vorliegenden Daten reagiert werden. Eine abschließende Festlegung erscheint uns zum jetzigen Zeitpunkt noch verfrüht.

5. Abgrenzung des Projektumfangs

Nicht Bestandteil des Projekts ist die Entwicklung oder Empfehlung hardwareseitiger Kontrollmechanismen, wie z. B. Gewichtssensoren oder optischer Auswertungssysteme. Ebenso erfolgt keine juristische Bewertung in Hinsicht auf Datenschutz oder Zulässigkeit von Kontrollvorgängen.

Die durchgeführten Analysen basieren ausschließlich auf **anonymisierten Transaktionsdaten**, die durch die Wertkauf GmbH zur Verfügung gestellt wurden. Weitere Abgrenzungen und Risiken werden in einem separaten Abschnitt behandelt.

6. Datenlage

Die Wertkauf GmbH hat insgesamt sechs Dateien für das Projekt zur Verfügung gestellt.

Bei der Sichtung der übermittelten Kassendaten wurde festgestellt, dass aktuell keine umfassende Dokumentation (z. B. in Form eines Data Dictionary) zur Verfügung steht. Dies erschwert die eindeutige Interpretation einzelner Felder, insbesondere im Hinblick auf Feldbedeutungen und Währungseinheiten. Wir empfehlen, sofern möglich, eine Ergänzung der Metadaten, um die Datenqualität und Auswertbarkeit nachhaltig zu verbessern.

Die Daten bestehen aus zwei CSV-Dateien mit Stammdaten zu Filialen und Artikeln sowie vier Dateien im Parquet-Format, welche Transaktions- und Positionsdaten enthalten – jeweils aufgeteilt in Trainings- und Testdaten.

Die Testdaten stammen aus den Jahren 2022 und 2023, während die Trainingsdaten das Jahr 2024 betreffen. Die Trainingsdaten enthalten zwei zusätzliche Spalten, die in den Testdaten nicht vorhanden sind:

- **label** – gibt an, ob eine Kontrolle durchgeführt wurde und welches Ergebnis diese hatte,
- **damage** – enthält vermutlich die Schadenshöhe in Euro, falls bei einer Kontrolle ein fehlerhafter Scan oder Betrugsversuch festgestellt wurde.

Diese beiden Spalten stellen offensichtliche Zielvariablen dar, die durch ein Modell vorhergesagt werden könnten, um die gegebene Problemstellung zu lösen. Es könnten ein Klassifikationsmodell zur Erkennung potenzieller Betrugsfälle sowie ein Regressionsmodell zur Schätzung der zu erwartenden Schadenshöhe im Fall eines Betrugs kombiniert werden, um eine fundierte und differenzierte Entscheidungsunterstützung zu ermöglichen.

Von insgesamt 1.481.783 Transaktionen im Trainingsdatensatz wurde bei 148.025 eine Kontrolle durchgeführt; davon wurden 4.656 als Betrug klassifiziert (ca. 3,14 %). Die Spalte **damage** ist konsistent mit der **label**-Spalte: Transaktionen ohne Kontrolle weisen keinen Wert auf, Kontrollen mit unauffälligem Ergebnis haben den Wert 0, und nur bei als Betrug markierten Fällen wird ein positiver Schadensbetrag angegeben.

Die Anzahl klassifizierter Betrugsfälle ist relativ gering, was die Entwicklung eines leistungsfähigen Modells erschweren könnte, insbesondere im Hinblick auf die Erkennung seltener Muster.

Die Daten weisen insgesamt einen hohen Vollständigkeitsgrad auf. Nennenswerte Anteile fehlender Werte betreffen lediglich folgende Spalten:

1. **customer_feedback**: Hier fehlen Einträge vermutlich, weil nicht in sämtlichen Fällen eine Bewertung abgegeben wurde.
2. **valid_to**: Rund 28 % der Produkte haben kein Gültigkeitsdatum.
3. **weight**: Bei 2.505 Artikeln fehlt die Gewichtsangabe, obwohl nur 2.155 Produkte nach Gewicht verkauft werden. Somit fehlen bei etwa 6 % der nicht-gewichtsbasierten Artikel ebenfalls Gewichtsangaben.

4. **camera_certainty, camera_product_similar:** Fehlende Werte in diesen Spalten treten über den gesamten Zeitraum hinweg auf, was auf temporäre Ausfälle des Kamerasystems hindeutet. Die Verteilung dieser fehlenden Werte ist in gelabelten und nicht-gelabelten Daten ähnlich, was positiv zu werten ist: Die gelabelten Daten bilden offenbar eine repräsentative Stichprobe der Gesamtdaten.

Eine weitere Auffälligkeit betrifft die Konsistenz zwischen Transaktionsdaten und den zugeordneten Positionen: Die in der Spalte **n_lines** (aus der Transaktionsdatei) angegebene Anzahl Positionen stimmt in mehreren Fällen nicht mit der tatsächlich verknüpften Anzahl Positionen überein.

Preise und Geldbeträge liegen im Fließkommaformat vor, was zwar unüblich ist, jedoch keine grundlegenden Probleme erwarten lässt. Abgesehen von einigen Formatabweichungen bei Zeitstempeln ist die Datenqualität insgesamt als gut einzuschätzen.

In der bisherigen Analyse ergaben sich keine Hinweise darauf, dass die Daten grundsätzlich ungeeignet wären, um ein Modell zur Lösung der Aufgabenstellung zu entwickeln.

7. Meilensteine

Meilenstein	Phase	Start	Ende
1. Projektdefinition und Zielklärung, Projektauftrag inkl. Präsentation	Define	07.04.2025	29.04.2025
2. Datenzugang, Qualitätsprüfung und erste Datenanalysen und Präsentation der Ergebnisse; Definition der REST-Schnittstelle	Acquire	07.04.2025	18.05.2025
3. Datenaufbereitung, Feature Engineering, Modellentwicklung	Structure	19.05.2025	15.06.2025
4. Dokumentation der Ergebnisse und Übergabe des über eine REST-Schnittstelle lauffähigen Programmcodes	Communicate	16.06.2025	29.06.2025

Phasenergebnisse im Detail

Zu beachten ist, dass sich die Meilensteine an der vorgesehenen Gesamtlaufzeit des Projekts orientieren und keinen Anspruch auf Ausschöpfung aller erdenklichen oder sinnvollen Ansätze stellen.

1. Meilenstein

Ziel dieser Phase ist die präzise Formulierung des Projektziels inklusive betriebswirtschaftlicher und technischer Rahmenbedingungen. Der Projektauftrag wird formuliert, mit dem Kunden abgestimmt und anschließend zusammenfassend präsentiert. Die Bewertungsfunktion zur späteren Modellbewertung wird definiert und diskutiert. Potenzielle Zielkonflikte wie das Spannungsfeld zwischen Diebstahlprävention und Vermeidung unnötiger Kontrollen sowie potenzielle Probleme mit den verfügbaren Daten werden in dieser Phase identifiziert und analysiert.

Artefakte:

1. abgestimmter Projektauftrag (PDF)
2. Präsentationsfolien über den Projektauftrag und die Projektstruktur (PowerPoint oder PDF)

2. Meilenstein

Diese Phase umfasst den Zugang zu den bereitgestellten Transaktions-, Artikel- und Filialdaten sowie deren detailliertere Sichtung hinsichtlich Vollständigkeit, Konsistenz und Interpretierbarkeit. Kritische Aspekte wie die fehlende Dokumentation (Data Dictionary), der Anteil klassifizierter Daten und die initiale Einschätzung der Modellierbarkeit werden hier auf ihren Einfluss auf mögliche Ergebnisse untersucht. In dieser ersten Analysephase werden zudem grundlegende Kenntnisse über den Datensatz erlangt, die für die spätere Modellentwicklung wichtig sind. Hier werden sowohl freie explorative Analysen durchgeführt als auch statistische Tests auf Zusammenhänge.

Die gewonnen Erkenntnisse werden bereits in einer kurzen Präsentation vorgestellt. Sollte es hier zu überraschenden Einsichten kommen, gibt es die Möglichkeit, den weiteren Verlauf des Projekts anpassend zu gestalten.

Die für die spätere Übergabe notwendige REST-Schnittstelle wird festgelegt, um am Ende der Projektlaufzeit einen lauffähigen Prototyp an die Wertkauf GmbH übergeben zu können.

Artefakte:

3. Data Audit Report (fehlende Werte, Formatprobleme etc.)
4. Explorative Datenanalyse (erste Hypothesen über Datenmuster)
5. Präsentation der Erkenntnisse (Folien mit Visualisierungen)
6. Dokumentation zur geplanten REST-Schnittstelle

3. Meilenstein

Zunächst erfolgt die systematische Bereinigung und Strukturierung der Daten. Auf Basis dieser strukturierten Daten werden geeignete Features ausgewählt, die sich zum Teil bereits aus der Arbeit in Meilenstein 2 ergeben. Z. B. Zeitmerkmale, Artikelkategorien oder verdächtige Verhaltensmuster. Anschließend werden verschiedene Modellklassen auf diesen Daten getestet, um schrittweise komplexere und mächtigere Prognosemodelle zu entwerfen. In der Endphase könnte es sich hierbei um Ensemble-Modelle oder neuronale Netze handeln. Es ist jedoch möglich, dass aufgrund der Beschaffenheit der verfügbaren Trainingsdaten (z.B. zu wenige, zu korrelierte oder unbrauchbare Daten) bestimmte Modelle nicht trainierbar sind. Hier gilt es, ein sinnvolles Maß an Komplexität zu wählen. Ziel ist die Entwicklung eines belastbaren Prototyps zur Markierung auffälliger Transaktionen. Dabei wird die betriebswirtschaftliche

Bewertungsfunktion in die Optimierungsstrategie des Modells integriert und schließlich auch zur systematischen Evaluierung des Prototyps herangezogen.

Aus dem Prototyp abgeleitete Handlungsempfehlungen orientieren sich sowohl an wirtschaftlicher Effizienz als auch anhand der logischen Nachvollziehbarkeit, um schlecht generalisierende Modelle und ein mögliches Reputationsrisiko durch deren Nutzung zu minimieren.

Artefakte:

7. Feature-Katalog mit Beschreibung, Typ und Berechnungsmethode
8. Modellübersicht (getestete Modelle inkl. Parameter)
9. Bewertungsbericht (Precision, Recall, economic loss/gain etc.)
10. Visualisierung der Modelllogik

4. Meilenstein

Zum Abschluss des Projekts werden alle Ergebnisse systematisch dokumentiert und in einem übersichtlichen Bericht aufbereitet. Dieser enthält sowohl Erklärungen zu den herausgearbeiteten besten Modellen als auch wirtschaftliche Handlungsempfehlungen. Der finale Prototyp, Codebasis und Handlungsleitfäden werden zur weiteren Nutzung an die Wertkauf GmbH übergeben. Zentraler Punkt ist die Übergabe des über die zuvor definierte REST-Schnittstelle aufrufbaren Prototyps.

Artefakte:

11. REST-Schnittstelle
 12. Alle Programmskripte (Python)
 13. Dokumentation
-

8. Risiken

Im Verlauf des Projekts können verschiedene Herausforderungen auftreten, die Einfluss auf die Ergebnisse und deren praktische Umsetzbarkeit haben könnten:

- **Grenzen der Bewertungsfunktion**
Wie oben bereits geschildert, ist ein Modell, das auf die vom Kunden gewünschte Bewertungsfunktion optimiert wird, vermutlich eher bestrebt, kleinere Fehler bzw. Diebstähle unkontrolliert zu lassen. Sollte es jedoch in der Filiale insbesondere durch viele kleinere Verluste zu einem großen Schaden kommen, wäre dieses Modell nicht ideal.
- **Begrenzte Aussagekraft der klassifizierten Daten**
Die kontrollierten und mit Klassifikation versehenen Transaktionen machen nur 3% der gesamten Datenmenge aus. Es ist möglich, dass sie nicht alle typischen Muster und Fälle repräsentieren. Das kann dazu führen, dass das Modell nicht gut auf weiteren Datensätzen (z.B. den Testdaten) generalisiert. Für einfache bis mittelkomplexe Modelle sind die Daten vermutlich ausreichend. Jedoch könnten für

sehr komplexe neuronale Netze die vorliegenden Trainingsdaten möglicherweise nicht ausreichend sein.

- **Eingeschränkte Übertragbarkeit auf andere Filialen**

Die Analyse basiert auf Daten aus einem bestimmten Filialumfeld. Da sich Kundenverhalten, Prozesse oder Technik in anderen Filialen unterscheiden können, ist nicht sicher, ob die Ergebnisse dort genauso gut funktionieren. Auch findet durch das Modelltraining anhand der gemischten Daten eine Mittelwertbildung statt, sodass keine differenzierte Einzelmodelle, sondern ein allgemeines Modell entwickelt wird.

- **Unausgewogene Datenverteilung**

Da nur ein kleiner Teil der Transaktionen fehlerhaft ist, ist die Verteilung der Klassen sehr unausgeglichen. Das kann sich negativ auf die Trainings- und Testergebnisse des Modells auswirken.

- **Modellverständlichkeit und Akzeptanz**

Damit die Ergebnisse später in der Praxis genutzt werden können, müssen sie auch nachvollziehbar sein – für alle Beteiligten. Wenn das Modell zu komplex ist, könnte es schwer werden, das Modell im Detail zu interpretieren.

- **Fehlendes Data Dictionary**

Da kein Data Dictionary zur Verfügung steht, besteht die Möglichkeit, dass bestimmte Variablen oder Werte nicht die erwartete Bedeutung haben oder falsch interpretiert werden. Insbesondere besteht das Risiko, dass zentrale Kennzahlen wie beispielsweise „damage“ betriebswirtschaftlich nicht sinnvoll definiert sind. Eine Klärung ist daher frühzeitig vorzunehmen.

- **Externe Vorgaben und rechtliche Rahmenbedingungen**

Regelungen wie Datenschutzvorschriften, jugendschutzrechtliche Bestimmungen, vertragliche oder gesellschaftsrechtliche Einschränkungen sowie Auflagen oder mögliche Kompensationen durch Versicherungen, die von der Wertkauf GmbH abgeschlossen wurden, wurden im Rahmen des Projekts nicht berücksichtigt.

9. Ressourcen

Für die Bearbeitung des Projekts ist keine zusätzliche technische Infrastruktur erforderlich. Alle drei Teammitglieder verfügen über geeignete Arbeitsgeräte sowie die notwendigen Entwicklungsumgebungen. Die Analyse und Modellentwicklung erfolgten primär in Python unter Verwendung frei verfügbarer Tools und Bibliotheken. Auch der Prototyp soll in Python umgesetzt werden.

10. Abgabe des entwickelten Prototyps

Der Prototyp wird in einer Form ausgeliefert, die eine reproduzierbare Ausführung in der Zielumgebung des Kunden gewährleistet. Um sicherzustellen, dass das Modell unabhängig von der konkreten Systemkonfiguration zuverlässig und konsistent funktioniert, wird eine der folgenden Auslieferungsformen verwendet:

- Containerisierte Bereitstellung (z. B. via Docker): Das Modell wird inklusive aller notwendigen Abhängigkeiten, Konfigurationsdateien sowie der spezifischen Python-Version in einem Docker-Container verpackt. Dies ermöglicht eine einheitliche Ausführung auf allen gängigen Plattformen, die Container-Technologie unterstützen.
- Alternativ kann das Modell als Python-Paket mit begleitender virtueller Umgebung (z. B. via venv oder conda) ausgeliefert werden. Eine vollständige Liste aller benötigten Abhängigkeiten inklusive exakter Versionsnummern wird durch eine requirements.txt oder ein environment.yml dokumentiert. Zusätzlich wird ein Startskript bereitgestellt, das die Nutzung des Modells erleichtert.

Die endgültige Auslieferungsform wird frühzeitig im Projektverlauf in Abstimmung mit dem Kunden festgelegt. Dabei werden die technischen Voraussetzungen sowie die bevorzugte Integrationsform auf Kundenseite berücksichtigt.