**1. 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.

**2. 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 Kassiervorgänge 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.

**3. Bewertungsfunktion**

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

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

Was schließen wir aus dieser Bewertungsfunktion? Generell werden kleine Diebstähle, die nicht entdeckt werden, lediglich mit kleiner Schadenssumme bewertet. 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 könnte man auch unterschiedliche Modellvarianten berechnen.
* **Wertabhängige Bewertung des verhinderten Schadens:** Im Fall 3 könnte ein noch größerer Wert angenommen werden, wenn ein bewusst oder unbewusst fehlerhafter Scanvorgang entdeckt und diese Person nachhaltig zu mehr Achtsamkeit angeregt bzw. bewusster Diebstahl in Zukunft komplett verhindert wird. Hier gilt es abzuwägen, inwiefern ein Reputationsschaden durch irrtümliche Kontrollen größer und dementsprechend in der Bewertungsfunktion stärker gewichtet werden soll als ein verhinderter Schaden oder umgekehrt die Diebstahlprävention wichtiger als falsche Kontrollen ist. Hier gilt das grundsätzliche Prinzip, dass nicht gleichzeitig sowohl Diebstahl / falsche Scanvorgänge optimal erkannt und gleichzeitig irrtümliche Kontrollen minimiert werden können.

Unsere Vorschläge und weitere Details können im Laufe der konkreten Datenanalyse und der Modellentwicklung besprochen werden, um flexibel auf die aus den Daten gewonnen Erkenntnisse reagieren zu können. Zum jetzigen Zeitpunkt erscheint uns eine abschließende Entscheidung nicht sinnvoll.

**4. Abgrenzung des Projektumfangs**

Nicht Bestandteil des Projekts ist die Entwicklung oder Empfehlung hardwareseitiger Kontrollmechanismen, wie z. B. Gewichtssensorik oder optischer Auswertungssysteme. Ebenso erfolgt keine juristische Bewertung hinsichtlich Datenschutzes 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.

**5. Datenlage**Die Wertkauf GmbH hat

Die Wertkauf GmbH hat insgesamt sechs Dateien für das Projekt zur Verfügung gestellt. Ein begleitendes Data Dictionary fehlt, wodurch bei einzelnen Attributen potenziell Rückfragen notwendig sein werden, um Missverständnisse zu vermeiden.

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 aus dem Jahr 2024 vorliegen. 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 adressieren. 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:

* **customer\_feedback**: Hier fehlen Einträge vermutlich, weil nicht alle Kund:innen eine Bewertung abgeben.
* **valid\_to**: Rund 28 % der Produkte haben kein Gültigkeitsdatum.
* **weight**: Bei 2.505 Artikeln fehlt der Gewichtsangabe, obwohl nur 2.155 Produkte nach Gewicht verkauft werden. Somit fehlen bei etwa 6 % der nicht-gewichtsbasierten Artikel ebenfalls Gewichtsangaben.
* **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.

**6. Meilensteine**

**7. Risiken**

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

1. 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 jedoch in der Filiale gerade durch viele kleinere Verluste zu einem großen Schaden kommen, wäre dieses Modell nicht ideal.
2. 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 mittel komplexe Modelle sind die Daten vermutlich ausreichend. Jedoch könnten für sehr komplexe neuronale Netze zu wenig klassifizierte Trainingsdaten verfügbar sein.
3. 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.
4. 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.
5. Modellverständlichkeit und Akzeptanz:  
   Damit die Ergebnisse später wirklich genutzt werden, müssen sie auch nachvollziehbar sein – für alle Beteiligten. Wenn das Modell zu komplex ist, könnte es schwer werden, das Modell im Detail zu interpretieren.