

Projektpraktikum Data Science der Fernuni Hagen im SS 2025

Hier einmal meine Gedanken, ohne Anspruch auf Vollständigkeit oder Richtigkeit

Allgemeines

1. Mail an Frau Folz-Weinstein wegen Danksagung für das Zoom-Meeting, offener Fragen und Terminvereinbarung für das nächste Zoom-Meeting
2. Ggf. Hinweis Abgrenzung Diebstahl und Betrug? → hier dürfte es sich eher um Diebstahl handeln
3. Müssen wir ein Protokoll dieser Zoomsitzung schreiben und der Dokumentation beifügen?
4. Wir sollten die **genaue** Aufgabenstellung noch einmal mit dem Kunden abstimmen und uns bestätigen lassen.
5. Zielfunktion (Bewertungsfunktion) überprüfen (s.u.)
6. Welche Bedeutung haben die Daten des Kassensystems **ohne** Kontrolle? → z.B. Ermittlung der durchschnittlichen Warenkörbe etc.
7. Gibt es personenbezogenen Daten, die besonders geschützt oder anonymisiert werden müssen? → ggf. Hinweis in Dokumentation
8. Wie werden die Daten bei **uns** gesichert (Verschlüsselung Zugriff etc.) → ggf. Hinweis in Dokumentation
9. Meiner Meinung nach sollten wir dem Kunden auch allgemeine Maßnahmen gegen Ladendiebstahl empfehlen: [▷ Diebstahlschutz an der Selbstbedienungskasse](#)
10. Weiteres Problem: Diebstahl OHNE Kassenbon → wie ist so etwas zu beurteilen?
11. Benötigen wir eine GENAUE Darstellung des Scanvorgangs bzw. die **Art** der Selbstbedienungskasse → vgl. Abbildung in Dokumentation der Fernuni. Spielt das für uns eine Rolle?
12. Hat der Kunde eine Diebstahl **Versicherung**, die einen evtl. Schaden erstattet? → Vermutlich nein, aber ggf. fragen.
13. Gibt es weitere Auflagen, die wir berücksichtigen müssen (durch Gesetze, Satzung, Gesellschaftsvertrag) die zu berücksichtigen sind?
- 14.
15. Benötigen wir ein Quellenverzeichnis?

A. Daten

1. Gibt es ein Data-Dictionary oder Metadaten? (Spaltenbezeichnungen, Währung)
2. Wie wurden die Daten erhoben? Wie wurden die Stichproben der Kontrollen erhoben? (zufällig oder nach Anfangsverdacht?)
3. Datenquantität und -qualität ausreichend?
4. Spalten einfügen (z.B. Scandauer)
5. Können ggf. weitere Daten angefordert werden oder weitere Felder?

B. Bewertungsfunktion

Die Bewertungsfunktion ist aus meiner Sicht wesentlich für die Aufgabe, weil ein (Verständnis)-fehler möglicherweise zu völlig falschen Ergebnissen führt. Wir sollten die Bewertungsfunktion deshalb vorher mit dem „Kunden“ abstimmen bzw. uns diese bestätigen lassen. Optimal wäre aus meiner Sicht, eine **flexible** Bewertungsfunktion zu nehmen (Schaden und Nutzen kann angepasst werden, entweder feste Werte oder

Artikelpreise etc.)

Soweit ich das in der Veranstaltung richtig verstanden haben, soll die folgende Bewertungsfunktion zugrunde gelegt werden:

1. Korrekter Bon als korrekt eingestuft $\rightarrow 0$ €
2. Korrekter Bon als inkorrekt eingestuft $\rightarrow -10$ € (Kunde verärgert, MA vergeblich tätig)
3. Inkorrekt Bon als korrekt eingestuft \rightarrow **Schaden (Frage: wie berechnet sich der Schaden? Verkaufspreis, historischer Einkaufspreis oder Wiederbeschaffungskosten?)**
4. Inkorrekt Bon als inkorrekt eingestuft $\rightarrow +5$ € ?? Macht es Sinn, hier einen pauschalen Wert anzusetzen? M.E. sollte man vielleicht den verhinderten Schaden zugrunde legen).
5. Angenommen wird hierbei, dass eine durchgeführte Kontrolle immer richtige Ergebnisse liefert, also z. B. eine durchgeführte Kontrolle, die keinen Fehler entdeckt auch tatsächlich einen korrekten Einkauf beinhaltet.

C. Konsequenzen kann man aus der Info: 5 % der Scans sind inkorrekt

Achtung: Das sind erst einmal nur statistische Überlegungen, die mit Data Science nicht viel zu tun haben und sie dienen nur als Anregung

1. Das heißt: ein Einkauf mit 20 Scans enthält im Durchschnitt **einen** fehlerhaften Scan
2. Hier dürfte m.E. ein Fall der Poisson-Verteilung vorliegen, mit $\lambda = 1$). Mann könnte zum Beispiel wie folgt argumentieren:
3. Bei Einkäufen mit 20 Scans ($\lambda = 1$) ist somit die Wahrscheinlichkeit, dass der gesamte Einkauf fehlerfrei ist: $1/e$. also ca. 37 Prozent. Nach der Bewertungsfunktion wäre der Erwartungswert für solche Einkäufe damit: $0,37 * (-10) + 0,63 * (+5) = -0,55$, also negativ. Bei Einkäufen mit mehr als 20 Scans wäre dann der Erwartungswert für diese Einkäufe wohl positiv: z.B. bei 40 Scans pro Einkauf wäre $\lambda = 2$ etc.
6. Ggf. allgemeine Empfehlung, kleine Einkäufe generell nicht zu überprüfen und große Einkäufe immer zu kontrollieren und für alle übrigen Einkäufe eine gemischte Strategie
7. Aus betriebswirtschaftlicher Sicht müssen auch immer weitere Gesichtspunkte eine Rolle spielen: Wie groß ist der insgesamt zu erwartende Nutzen absolut? Auch die Kosten dieser Studie, die Kosten der technischen Umsetzung müssen berücksichtigt werden. Ggf. kann man zum Beispiel auch weitgehend auf Kontrollen verzichten, wenn die eingesparten Personalkosten (z.B. 60 TEUR pro Jahr höher sind, als die ersparten Kosten)