Weiterführende Informationen zu Stunde 3

Hinweise zur Durchführung der *Lebendigen Statistik*

Man kann aufgrund der begrenzten Anzahl an SuS i. d. R. nicht alle Karten für die lebendige Statistik verwenden. Man sollte dabei darauf achten, dass keine ungünstigen Stichproben der Karten genutzt werden. Manche Stichproben vereinfachen das Problem zu stark, sodass man zu einfach perfekte Regeln findet. Wir schlagen dafür einen Rumpfdatensatz vor, mit dem die Phase gut durchgeführt werden kann. Vorschläge für Rumpfdatensätze verschiedener Größe finden sich in den Dateien „Rumpfdatensatz\_22.csv“ und „Rumpfdatensatz\_28.csv“. Leichte Abweichungen von den Rumpfdatensätzen sind kein Problem.

Mit der gesamten Klasse wird die Lebendige Statistik durchgeführt, um Datensplits anhand von Schwellenwerten durchzuführen und datenbasierte Entscheidungsregeln daraus abzuleiten.

1. Jede SuS nimmt eine Datenkarte (inkl. farbiger Klammer) und repräsentiert dieses Lebensmittel.
2. Die Lehrkraft nennt einen Schwellenwert (z. B. Energie und 350 kcal). Jede SuS dessen Lebensmittel einen Energiewert ≤ 350 hat geht auf die linke Seite des Klassenzimmers, jede SuS mit einem Wert > 350 geht auf die rechte Seite.
3. In den so entstandenen Teildatensätzen wird die Anzahl der roten und grünen Klammern („eher empfehlenswert” oder „eher nicht empfehlenswert”) ermittelt, z. B. über Handzeichen. Die Lehrkraft hält die Mehrheitsverhältnisse an der Tafel fest.
4. Gemeinsam wird anhand der Mehrheitsverhältnisse für beide Teildatensätze eine Entscheidung getroffen („eher empfehlenswert” oder „eher nicht empfehlenswert”)
5. Die Lehrkraft verbalisiert die Entscheidungsregel („Wenn ein Lebensmittel 350 kcal oder weniger hat, klassifizieren wir es als ...”) und notiert sie als einstufigen Entscheidungsbaum an der Tafel.
6. Es wird ausgezählt, wie viele Lebensmittel in beiden Teildatensätzen zusammen nicht korrekt klassifiziert werden. Da diese Zahl noch recht hoch ist, formuliert die Lehrkraft das Ziel, eine neue Entscheidungsregel zu finden, mit der weniger Lebensmittel falsch klassifiziert werden. Dazu wird ein neuer Schwellenwert (durch die Lehrkraft oder SuS) für das Merkmal Energie vorgeschlagen, z. B. 300. Die Punkte 2-5 werden für den neuen Schwellenwert wiederholt.
7. Es wird ein Urteil gefällt, welcher Datensplit die bessere Entscheidungsregel liefert. Als Kriterium dient dazu die jeweilige Anzahl der falsch klassifizierten Lebensmittel. Der Fachbegriff hierfür ist Anzahl der Fehlklassifikationen. Es bleibt der Lehrkraft überlassen, obn dieser Fachbegriff benutzt oder nur umschrieben wird.
8. Die Punkte 6-7 werden erneut durchgeführt für einen dritten Schwellenwert.

Mögliches Tafelbild zum Schwellenwert 230 kcal:


Hintergrund: Entscheidungsregeln aufstellen

Ziel innerhalb der Unterrichtsreihe ist es, ein mehrstufiges Regelsystem zur Klassifikation von Lebensmitteln zu erstellen. Dafür lernen die SuS in dieser Stunde zunächst Entscheidungsregeln (einstufige Entscheidunsgbäume) aus den Daten abzuleiten. Dies wird mit dem Konzept des Datensplits umgesetzt, das in dieser Stunde eingeführt wird. Dabei werden die Datenkarten anhand eines Merkmals und eines sogenannten Schwellenwerts in zwei Teilgruppen aufgeteilt (z.B. Lebensmittel mit bis zu 10 g Fett und über 10 g Fett). In beiden Teilgruppen wird dann geschaut, ob die Mehrheit eher empfehlenswert oder eher nicht empfehlenswert ist. Wenn in den Teilgruppen unterschiedliche Label zu finden sind (was in den allermeisten Fällen so ist), gibt es in beiden Teilgruppen Lebensmittel, die von der Mehrheitsentscheidung abweichen. Diese werden als Fehler oder Fehlklassifikationen bezeichnet. Es gilt deshalb den Schwellenwert zu finden, der möglichst wenige Fehler hervorbringt (Anzahl der Fehlklassifikationen). Die Entscheidungsregeln werden in dieser Stunde exemplarisch mit Hilfe des Merkmals Energie erstellt.

Erarbeiten der Zielstellung für das Erstellen einer Entscheidungsregel:

Anhand von [Präsentation 2](https://unterrichtsmaterial-ddi.cs.upb.de/images/1/14/Pr%C3%A4sentation2_Datensplit_Einf%C3%BChrung.pptx) kann die Zielstellung beim Formulieren einer Entscheidungsregel erarbeitet werden. Der Optimalfall (Zielstellung) liegt vor, wenn man einen Schwellenwert findet, sodass auf einer Seite des Schwellenwerts möglichst nur eher nicht empfehlenswerte Lebensmittel zu finden sind und auf der anderen Seite nur eher empfehlenswerte. Eine solch "perfekte" Regel, mit der man von den vorliegenden Trainingsdaten kein Lebensmittel falsch klassifiziert, findet man meistens nicht, aber man versucht möglichst nah dranzukommen, sodass möglichst wenige Lebensmittel falsch klassifiziert werden.