Weiterführende Informationen zu Phase 8.1

Hinweise zum Umgang mit dem Jupyter Notebook

Link <https://go.upb.de/auto-baum>

Zugangsdaten Benutzername: jupyterhub Passwort: jupyter@upb

Die SuS arbeiten in der Lernumgebung in Form eines Jupyter Notebook, das per „Click and play“ benutzt werden kann. Die Benutzung des Jupyter Notebooks ist angelegt wie eine Website und als Lernumgebung erstellt, so dass sich SuS weitgehend selbst durcharbeiten können. Die Lernumgebung umfasst das komplette Erstellen eines Entscheidungsbaums.

Im Idealfall wird die Lernumgebung von oben nach unten von SuS durchgearbeitet. Eingebettet sind verschiedene Aktivitäten, zu denen es automatisches Feedback für die SuS gibt.

Anmeldename für die Lernumgebung:

Passwort für die Lernumgebung:

Bevor die Aktivitäten/Aufgaben in der Lernumgebung sichtbar sind, müssen zunächst alle Zellen einmal ausgeführt werden. Das geschieht im Menü unter Cell  Run all



Die Aktivitäten der Lernumgebung erfasst folgende Elemente.

* + **Datengrundlage**
	Zuerst werden SuS eingeführt in die Struktur einer Datentabelle erfolgt und wie diese sich zu den Datenkarten verhält. In einer Aufgabe wird abgefragt, welche Wert von den Datenkarten falsch übertragen wurde, damit SuS sich mit beiden Repräsentationen und deren Zusammenhang beschäftigen.
	*Lösung: Lebensmittel* Apfel *mit dem Merkmal* Eiweiß
	+ **Daten importieren**
	Als nächstes müssen die Daten der Lebensmittel importiert werden. Dies geschieht über den Button 
	+ **Trainingsdatensatz herstellen**
	Anschließend vergeben SuS Label für den Trainingsdatensatz wie in Stunde 2. Hier ist es sinnvoll, die gleichen Label zu vergeben, die in der Klasse genutzt wurden. Dazu kann entweder der Kartensatz mit Label oder Arbeitsblatt 2 zu Hilfe genommen werden.
	Hinweis: Die Label können später wieder geändert werden!
	+ **Daten sortieren**Das Sortieren der Daten nach einem ausgewählten Merkmal geschieht in der nächsten Aktivität. Auf Knopfdruck werden die Daten aufsteigend nach den Werten des gewählten Merkmals sortiert. SuS erkennen, dass diese händische Aktivität viel schneller durch den Computer ausgeführt werden kann.
	+ **Entscheidungsregel aufstellen**
	Wie im Unterricht geht es als nächstes darum, einen Schwellenwert zu finden. Das geschieht erst händisch und kann anschließend automatisch durchgeführt werden. Die Ausgabe beinhaltet die Anzahl der Fehlklassifikationen.
	+ **Entscheidungsbaum automatisch erstellen**Danach wird durch den Computer automatisch ein Entscheidungsbaum erstellt. Die Tiefe des Baums (als Anzahl der Stufen) kann dabei eingestellt werden, sodass SuS Erfahrungen sammeln können mit verschiedenen Anzahlen von Stufen eines Baums. Die maximale Tiefe des Baums beträgt 6, weil Merkmale nicht auf verschiedenen Stufen wiederverwendet werden dürfen.
	Hinweis: Abhängig von den zuvor gesetzten Labeln für die Lebensmittel werden unterschiedliche Bäume erzeugt. Mitunter werden nicht alle Stufen benötigt, um die kleinste Fehlklassifikationsanzahl zu erzeugen.
	+ **Neue Lebensmittel klassifizieren**Am Ende der Lernumgebung können neue Lebensmittel durch den Baum klassifiziert werden.

Anmerkung zum Quellcode im Jupyter Notebook:

Im Jupyter Notebook arbeiten die SuS nur mit menübasierten Elementen. Der Quellcode ist versteckt. Sollte im Jupyter Notebook aber versehentlich der Code einer Zelle angezeigt werden, so kann das „Ausführen” dieser Zelle durch die Tastenkombination Shift+Enter den Code wieder verbergen und die „schöne” Ansicht kann wiederhergestellt werden.

Vertiefungsmöglichkeit:

* **Diskussion, welchen Einfluss der Mensch durch die Datenmanipulation auf den Entscheidungsbaum hat**

Zum Vertiefen können die Ausgangsdaten im Jupyter Notebook unterschiedlich mit Labeln versehen werden (z. B. zufällig, oder alle mit dem gleichen Label, etc.). Dadurch lässt sich erkennen, welchen Einfluss die Daten bzw. die Modellierung der Daten (hier der menschliche Einfluss durch Vergabe der Label) auf die Ergebnisse hat.

Zu jedem so erstellten Trainingsdatensatz wird ein Entscheidungsbaum erstellt, weil der Computer nicht „weiß“, ob seine Datengrundlage sinnvoll oder nicht ist. Bei einer „unsinnigen“ Trainingsgrundlage ergibt jedoch auch der automatisch erstellte Entscheidungsbaum keinen Sinn. Dies verdeutlicht, dass die Qualität der Daten(-modellierung) entscheidend für die Qualität eines Regelsystems ist, das durch maschinelles Lernen erstellt wird.