

Aufgabenblatt 2 (Klassifikation)

Aufgabe 1 – Entscheidungsbaum I

(25 Punkte)

Gegeben seien folgende Trainingsdaten:

Churn	Anruf im CallCenter	Beschwerde	Zahlungsverzug
Ja	Ja	Ja	Ja
Ja	Nein	Nein	Nein
Ja	Nein	Ja	Ja
Nein	Ja	Nein	Ja
Nein	Nein	Ja	Nein
Nein	Nein	Nein	Ja

- (a) Leiten Sie das Attribut Churn mit Hilfe eines Entscheidungsbaums ab. Nutzen Sie hierfür als Splitkriterium den Gini-Index und entwickeln Sie solange neue Knoten bis entweder kein Splitattribut mehr verfügbar ist oder in einem Knoten alle Daten in der gleichen Klasse sind. **(10 Punkte)**
- (b) Sagen Sie mit Hilfe des Entscheidungsbaums aus (a) die Empfehlung für folgende Beobachtungen voraus (Ja, Nein, Nein) und (Ja, Ja, Nein). **(5 Punkte)**
- (c) Leiten Sie das Attribut Churn wie in Aufgabe (a) ab, nutzen Sie jetzt das X^2 -Maß zur Auswahl der Splits. **(10 Punkte)**

Aufgabe 2 – Entscheidungsbaum II

(30 Punkte)

Gegeben seien folgende Trainingsdaten:

Empfehlung	Form	Farbe	Material
Kaufen	Formumspielend	rot	Leder
Kaufen	Körperbetont	gelb	Baumwolle
Kaufen	Körperbetont	rot	Leder
Nicht kaufen	Formumspielend	gelb	Leder
Nicht kaufen	Körperbetont	rot	Baumwolle
Nicht kaufen	Körperbetont	gelb	Leder

- (a) Leiten Sie das Attribut Empfehlung mit Hilfe eines Entscheidungsbaums ab. Nutzen Sie hierfür als Splitkriterium den Informationsgewinn und entwickeln Sie solange neue Knoten bis entweder kein Splitattribut mehr verfügbar ist oder in einem Knoten alle Daten in der gleichen Klasse sind. **(10 Punkte)**
- (b) Überführen Sie den Entscheidungsbaum aus (a) in eine Menge von Regeln. **(7 Punkte)**
- (c) Sagen Sie mit Hilfe des Entscheidungsbaums aus (a) die Empfehlung für folgende Beobachtung voraus (Formumspielend, gelb, Baumwolle). **(3 Punkte)**
- (d) Nutzen Sie einen Naive-Bayes-Klassifikator um für das Tupel aus (c) die Empfehlung vorherzusagen. **(10 Punkte)**

Aufgabe 3 - Verständnisfragen

(10 Punkte)

- (a) Beschreiben Sie, wie Support Vector Machines funktionieren. **(5 Punkte)**
- (b) Diskutieren Sie, wie die Qualität von Klassifikatoren durch den Benutzer deutlich gesteigert werden kann. **(5 Punkte)**