BW2 Praktikum

Aufgabe 6 – Gruppe 1

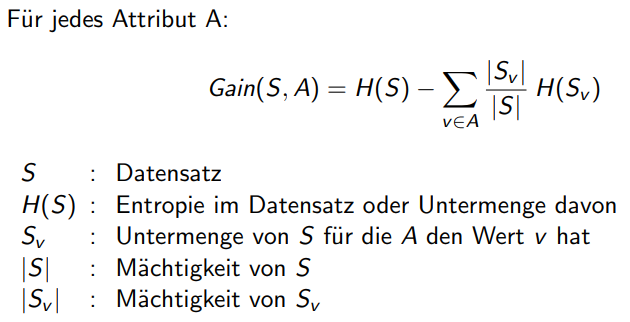
Adrian Helberg

Version 1.0

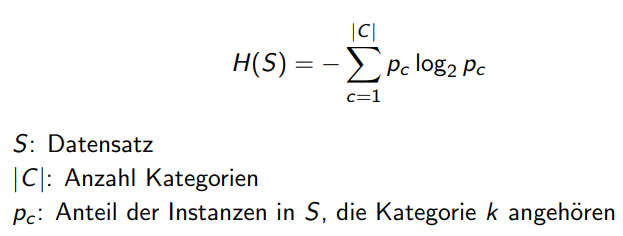
14.12.2018

# Aufgabe 9

Informationsgewinn:



Entropie im Datensatz:



Gesamtentropie berechnen (DAMAGE):



= 0.9612347…

Entropie für die Teildatensätze berechnen:



= 0.650022…

Gesamtentropie zusammensetzen:

Alle fehlenden Attribute berechnen:

\* AGE wird in <30 und >=30 aufgeteilt

*„SEX“ erzielt den größten Informationsgewinn, daher wird es als Baumwurzel gewählt.*

Gesamtentropie berechnen (SEX):



= 0.995727…

Entropie für die Teildatensätze berechnen:

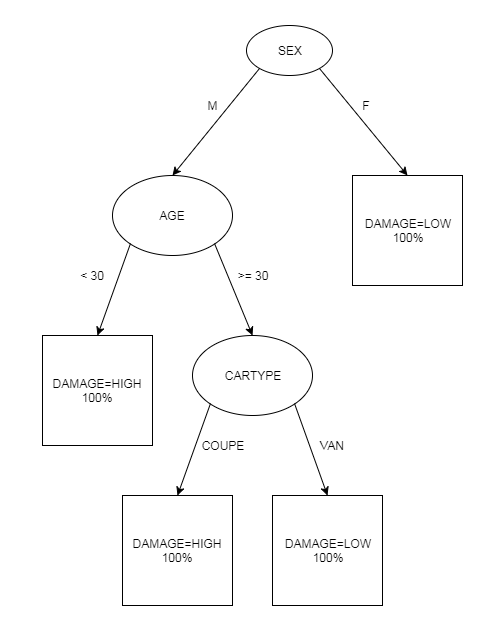


= 0.985228…

Gesamtentropie zusammensetzen:

Alle fehlenden Attribute berechnen:

„AGE“ erzielt den größten Informationsgewinn, daher wird es als weiterer Knoten gewählt.  
Alle übrigen Informationen können ohne Berechnungen in den Baum eingefügt werden.



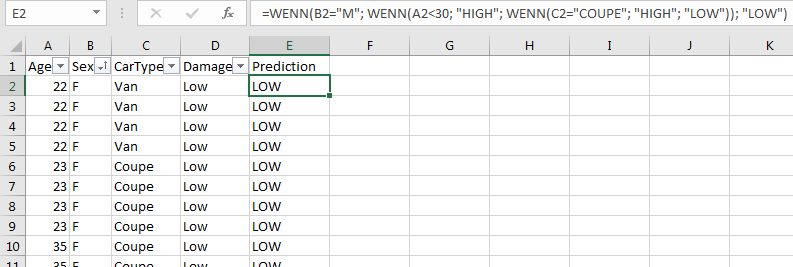
Der Entscheidungsbaum zeigt, dass das Geschlecht in den vorgegebenen Daten sehr aussagekräftig ist im Bezug auf das Attribute „DAMAGE“. Eine Versicherungsgesellschaft kann diese Datenanalyse nutzen, um teurere Versicherungsverträge für gewisse „Risikogruppen“ (Hier: Männer unter 30 Jahre und Männer, die ein Coupe fahren) zu erstellen, um eine Kostendeckung zu erreichen, sollte ein Schaden am Auto auftreten. Billigere Verträge gehen z.B. an die Gruppe „Frauen“.

Abgeleitete Regeln:

***IF*** *SEX = M* ***THEN   
 IF*** *AGE < 30* ***THEN*** *DAMAGE* ***IS*** *HIGH* ***ELSE  
 IF*** *CARTYPE = COUPE* ***THEN*** *DAMAGE* ***IS*** *HIGH* ***ELSE*** *DAMAGE* ***IS*** *LOW****ELSE*** *DAMAGE* ***IS*** *LOW*

Error Matrix (52 Datensätze):

|  |  |  |
| --- | --- | --- |
| DAMAGE | HIGH | LOW |
| HIGH | 20 | 0 |
| LOW | 0 | 32 |



Der Algorithmus hält einen Vorhersagewert von 100%