BW2 Praktikum

Aufgabe 6 – Gruppe 1

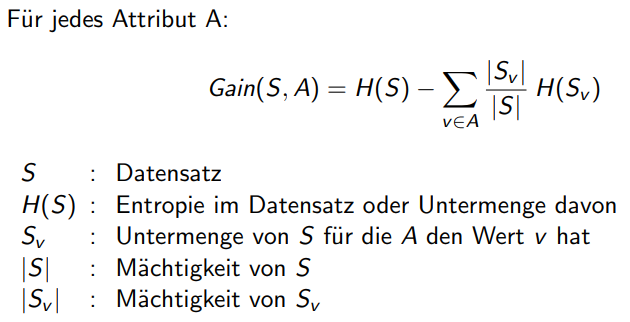
Adrian Helberg

Version 1.0

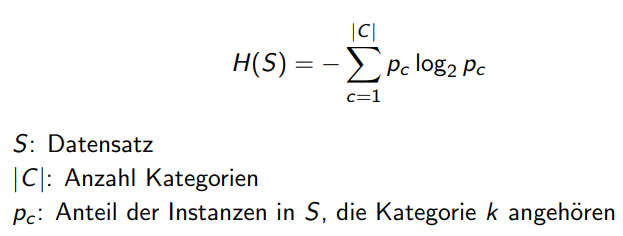
21.12.2018

# Aufgabe 9 b)

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:

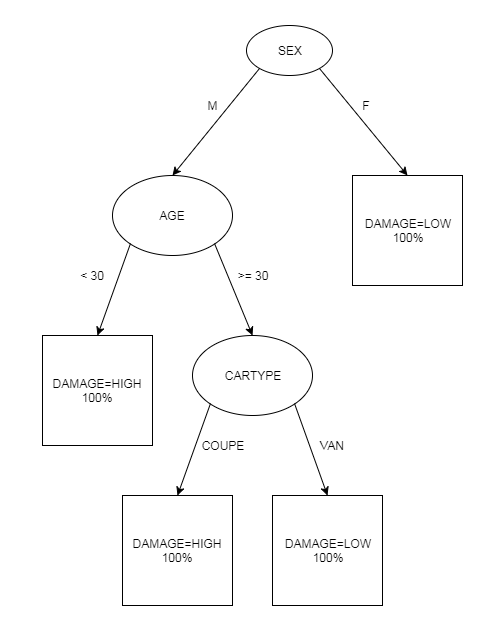


= 0.985228…

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.



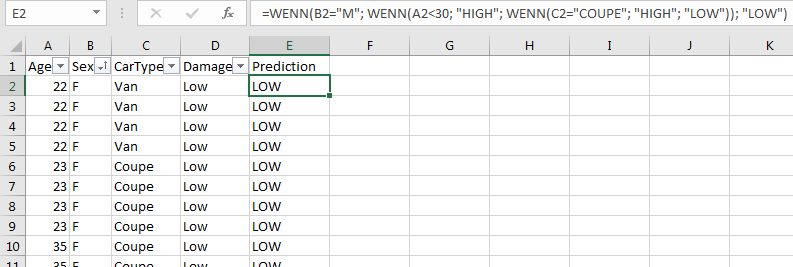
Der Entscheidungsbaum zeigt, dass das Geschlecht in den vorgegebenen Daten sehr aussagekräftig ist im Bezug auf das Attribute „DAMAGE“. Eine Versicherungsgesellschaft kann diese Datenanalyse nutzen, um teurere Versicherungsverträge für gewisse „Risikogruppen“ (Hier: Männer unter 30 Jahre und Männer, 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%

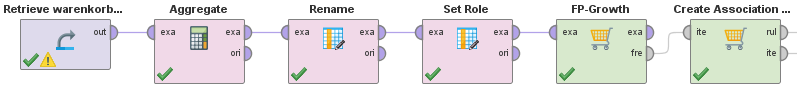
# Aufgabe 9 c)

Wichtige Kenngrößen:

|  |  |
| --- | --- |
| Support | Relative Häufigkeit der Beispiele, in denen die Regel anwendbar ist. |
| Konfidenz | Relative Häufigkeit der Beispiele, in denen die Regel richtig ist. |
| Lift | Angabe, wie hoch der Konfidenzwert für eine Regel den Erwartungswert übertriift, also die generelle Bedeutung einer Regel. |

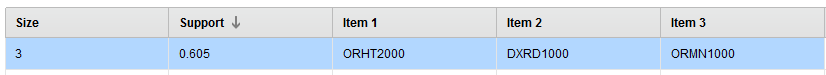
Beispiel 1:  
Minimum Support: 0.2  
Minimum Items pro Itemset: 3  
Maximum Items pro Itemset: 8  
Minimum Konfidenz: 0.8

Prozess:

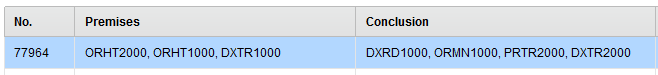


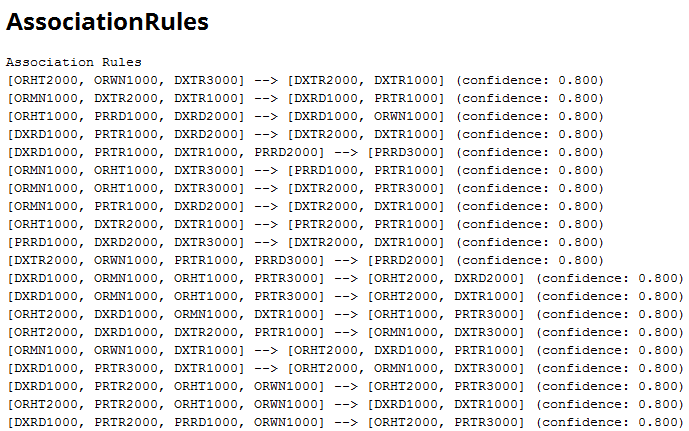
1. Gruppiere nach Order-Nummer und aggregiere Produkt ID als Konkatenation
2. Setze die Order-nummern als ID
3. FP-Groth und Association Rules Konfiguration siehe oben

FP-Growth (Maximaler Support):



Association Rules:

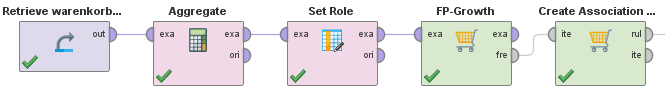



## Erläuterung

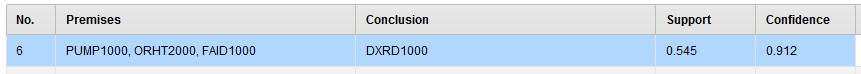
Die Regeln beschreiben das Kaufverhalten der Kunden.   
Z.B. wird, wenn die Produktkombination {ORHT2000, ORWN1000, DXTR3000} gekauft wird, mit einem Konfidenzwert von 80% auch die Kombination {DXTR200, DXTR1000} gekauft.

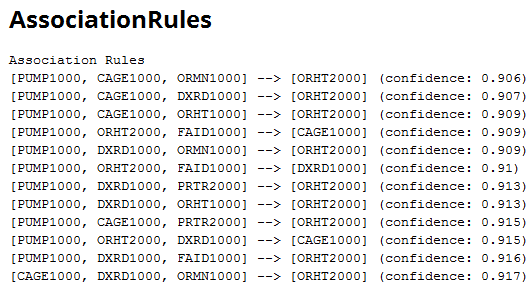
Aus diesen Assoziationen kann Dirt Bikes gezielt Werbung einsetzen, Produktkombinationen zusammenstellen und Über das Kaufverhalten der Kunden lernen, um so ein besserer Anbieter zu werden.

# Aufgabe 9 d)



1. Gruppiere nach „Time“ und CUSTOMER\_ID und aggregiere die PRODUCT\_ID als Konkatenation
2. Setze „CUSTOMER\_ID“ als Label und „TIME“ als ID
3. FP-Groth und Association Rules wie folgt:
   * Minimum Support: 0.8
   * Minimum Items pro Itemset: 3
   * Maximum Items pro Itemset: 8
   * Minimum Konfidenz: 0.9





## Erläuterung

Mit einem Konfidenzwert von 90,6% wird zu dem Produkt „Enduro 550 (silver)“ die Zubehörteile „LargeAir Pump“, „Water Bottle Cage“ und „Men’s Off Road Bike Fully“ gekauft.  
Um den Kaufwunsch des Kunden zu stärken, könnte Dirt Bikes die Kombination als Produkt-Paket verkaufen mit einem geringen Preisnachlass o.ä.