

Data Mining

18.06.2021

Fakultät für Ingenieurwissenschaften
Bereich Elektrotechnik und Informatik

C. Werner, J. Prothmann

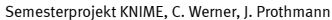
www.hs-wismar.de





Gliederung I

1 Entscheidungsbäume



- Standardknoten von Knime
- Zielattribut: nominal
- Entscheidungsfindungsattribute: nominal, numerisch
- Qualitätsmaße für Splitberechnung:
 - Gini-Index
 - Gain-Ratio
- Pruning möglich



SimpleCart

- Weka-Knoten
- Erzeugung von Binärbäumen
- Pruning möglich
- Je höher der Informationsgehalt eines Attributs in Bezug auf die Zielgröße, desto weiter oben im Baum findet sich dieses Attribut.

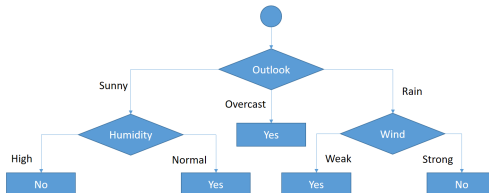


Bild 1: CART Tree Beispiel



J48

- Weka-Knoten
- C4.5 Algorithmus von J. Ross Quinlan
- Ähnlich zu CART, jedoch kein Binärbaum
- Deutlich breiter und weniger tief als CART
- Pruning möglich



NBTree

- Weka-Knoten
- Hybridalgorithmus aus Entscheidungsbaum- und Naive-Bayes-Klassifikatoren
- „klassische“ Knoten
- Blätter enthalten Naive-Bayes'sche Klassifikatoren

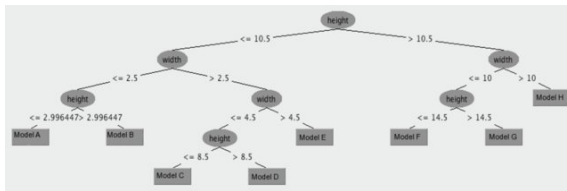


Bild 2: NB Tree Beispiel



REPTree



LMT



DecisionStump



J48Graft



BFTree



RandomTree



RandomForest



FT



Test

- mehr Energie und Sorgfalt bei Suchanfragenformulierung
- Wahl der passenden Suchbegriffe
- Nutzung spezieller Befehle und Operatoren
- „Effective internet searching is part science, part art, part skill and part luck.“ (Bradley 2013, S.18)

