

# Data Mining

18.06.2021

**Fakultät für Ingenieurwissenschaften**  
**Bereich Elektrotechnik und Informatik**

C. Werner, J. Prothmann

[www.hs-wismar.de](http://www.hs-wismar.de)





# Gliederung I

- 1 Entscheidungsbäume
  
- 2 Implementierung
  - 2.1 Entscheidungsbäume



## Decision Tree Learner

- 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.

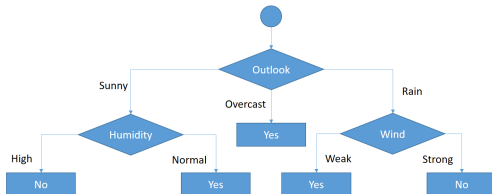


Bild 1: CART Tree Beispiel



## 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

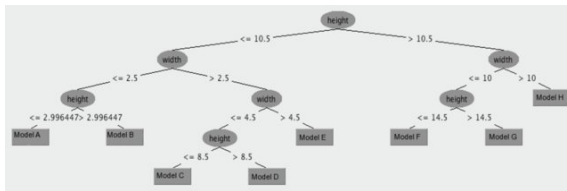


Bild 2: NB Tree Beispiel



## REPTree

- Weka-Knoten
- basiert auf C4.5 Algorithmus
- Generierung unter Berücksichtigung von:
  - Informationsgewinn
  - Varianz



## LMT

- Weka-Knoten
- Blätter: lineare Regressionsfunktionen
- stufenweiser Anpassungsprozess
- Automatische Auswahl relevanter Attribute





## DecisionStump

- Weka-Knoten
- einstufiger Entscheidungsbaum
- Vorhersage anhand des Wertes eines Eingabe-Features
- Knoten: Schwellenwert
- Blätter: Werte unterhalb und oberhalb des Schwellenwerts
- Einsatz als „schwache Lerner“ (z.B. Gesichtserkennung)



## J48Graft

- Weka-Knoten
- nutzt den C4.5++ Algorithmus
- Verbesserung durch „all-tests-but-one-partition“ (ATBOP)
- Reduzierte Rechenzeit
- Reduzierte Komplexität des Baums



## BFTree

- Weka-Knoten
- Best-First-Entscheidungsbaum
- „beste“ Knoten zuerst expandieren
- „beste“ Knoten: maximalen Reduktion der Unreinheit (z.B. Gini-Index)
- resultierende Baum nur in Reihenfolge unterschiedlich



## RandomTree

- Weka-Knoten
- zufällig ausgewählte Attribute an den Knoten
- kein Pruning



# RandomForest

- Weka-Knoten
- Kombination von Baumprädiktoren
- Abhängigkeit jedes Baumes von Werten eines Zufallsvektors
- Zufallsvektor: unabhängig und besitzt gleiche Verteilung für alle Bäume im ‘Wald’



# Gesamtworflow

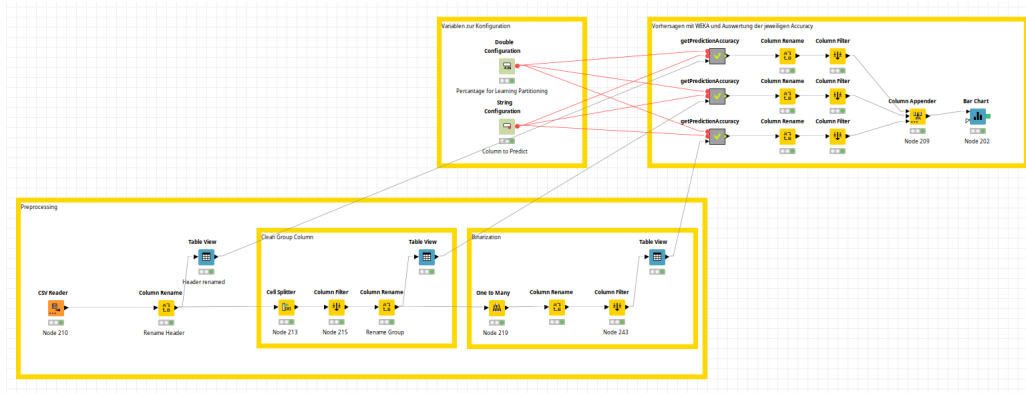


Bild 3: Gesamtworflow



## Ermittlung der Accuracys

- Knoten zur Eingabe von:
  - vorherzusagender Spalte
  - Trainingsdatenaufteilung
- Extrahierung der Accuracys und Anzeige in Balkendiagramm

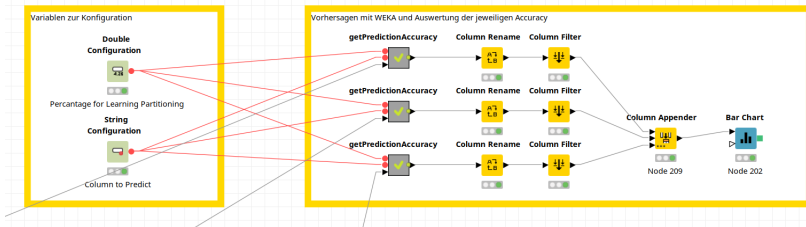


Bild 4: Ermittlung der Accuracys

## Metaknoten 'getPredictionAccuracy'

- Aufteilung in Trainings und Testdaten
- Extrahierung der Accuracys

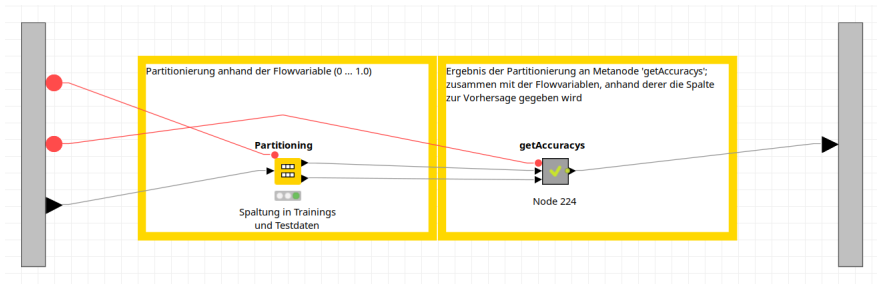


Bild 5: Inhalt des Metaknoten 'getPredictionAccuracy'





## Metaknoten 'getAccuracys'

- Metaknoten für verschiedene Entscheidungsbäume
- Gleiche Trainings- und Testmenge
- Zusammenführung der Accuracys in eine Tabelle

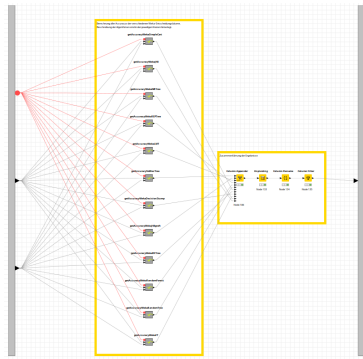


Bild 6: Inhalt des Metaknoten 'getAccuracys'

## Beispielhafter 'getAccuracyWeka\*' Knoten

- Lerner
- Vorhersage
- Scoring
- Extrahierung der Accuracy

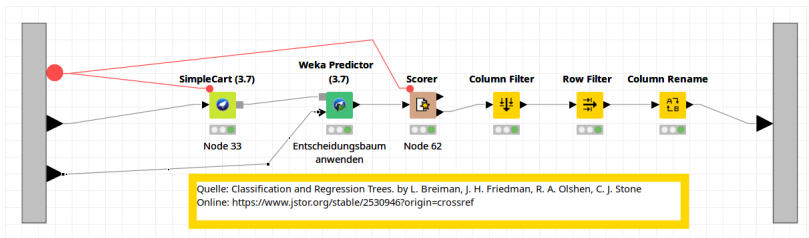


Bild 7: Inhalt des Metaknoten 'getAccuracyWekaSimpleCart'



## Accuracy Chart

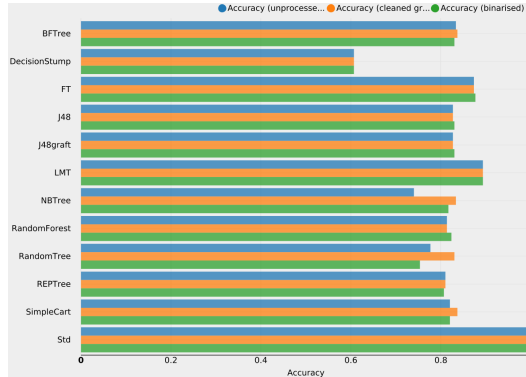


Bild 8: Balkendiagramm für 'gender' und Trainingssatz von 70%