

Predictive Analytics mit PDI und R

Dr. David James, it-novum GmbH

Pentaho User Meeting, 06.03.2018, Frankfurt am Main

Predictive Analytics mit PDI und R

- Einführung PA (Beispiel, Usecases)
- Umsetzung mit PDI und R

Einführung

Beispiel : Klassifikation



Beispiel: Klassifikation



iris setosa



iris versicolor

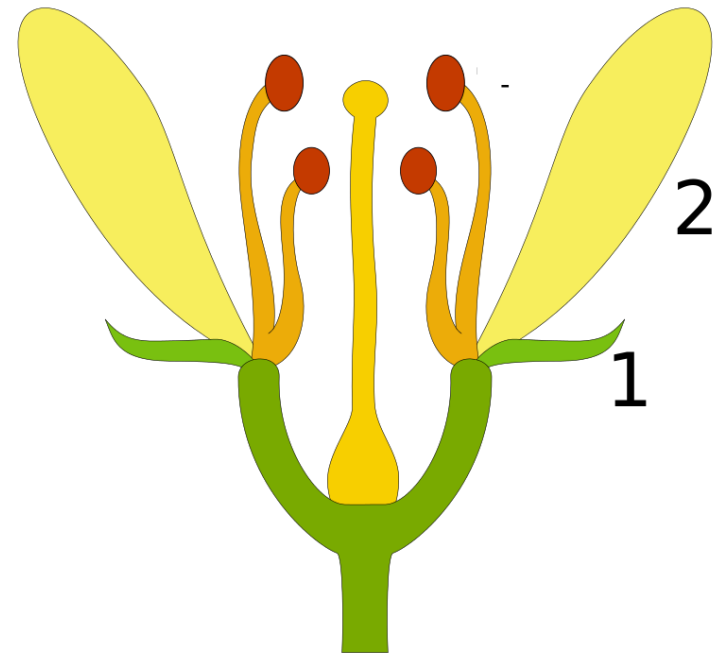


iris virginita

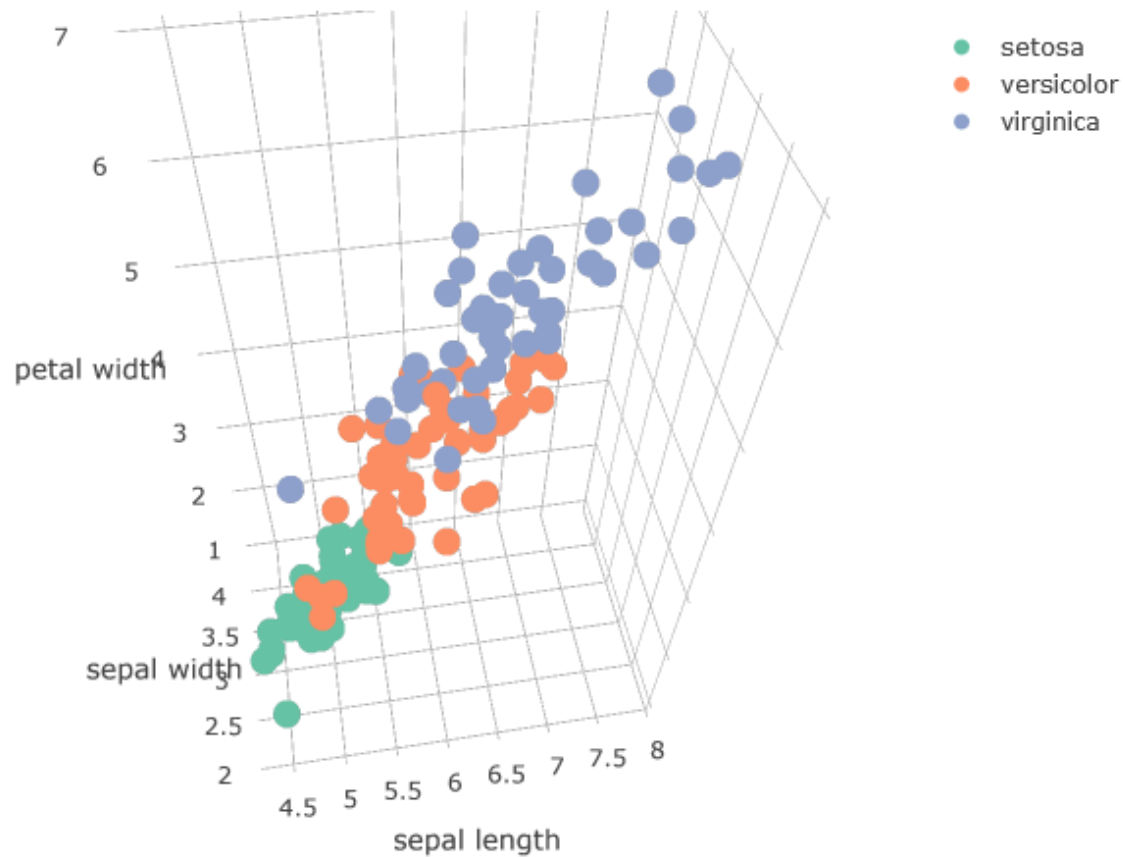
Iris Datensatz (Schwertlilien) [Ronald Fisher, 1936]

4 Kennzahlen / Features

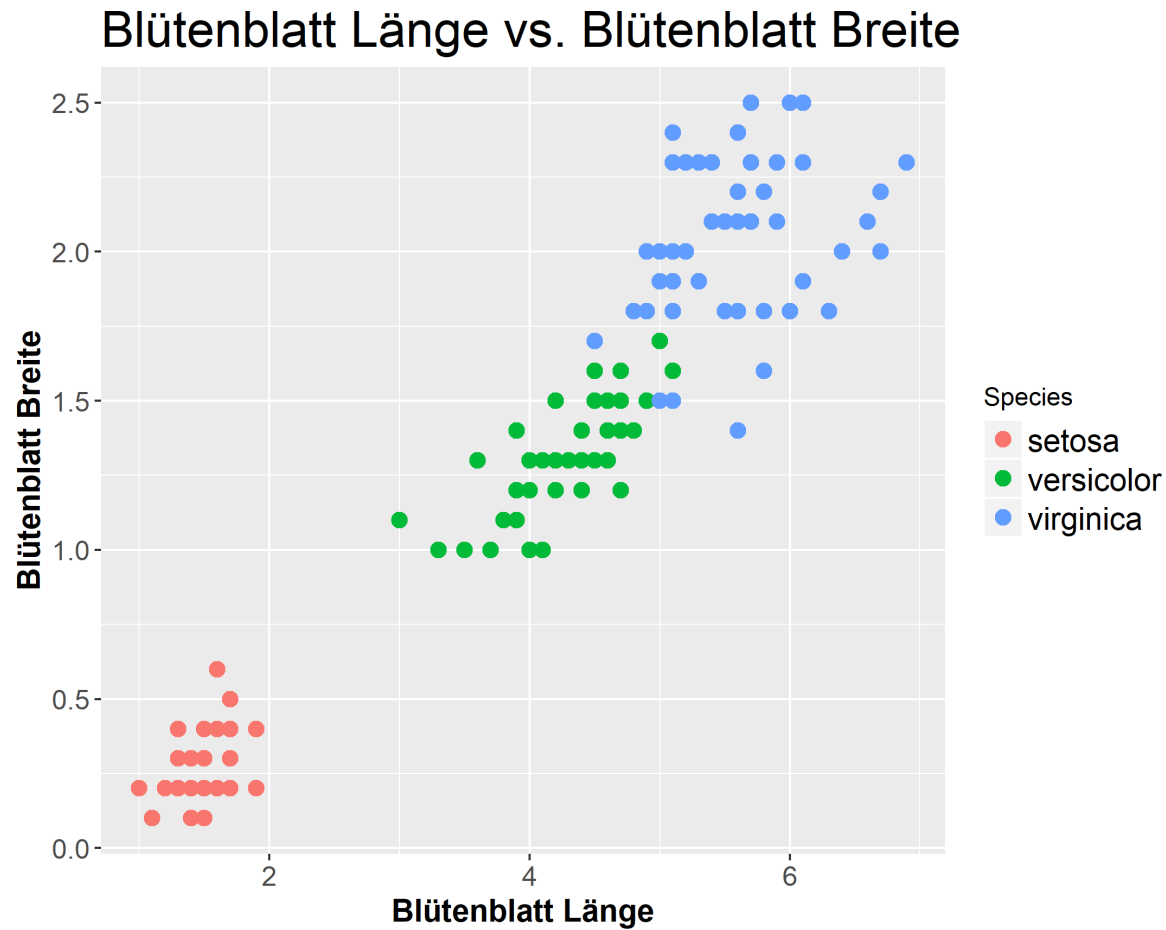
- 1) Kelchblatt (Länge und Breite)
- 2) Blütenblatt (Länge und Breite)



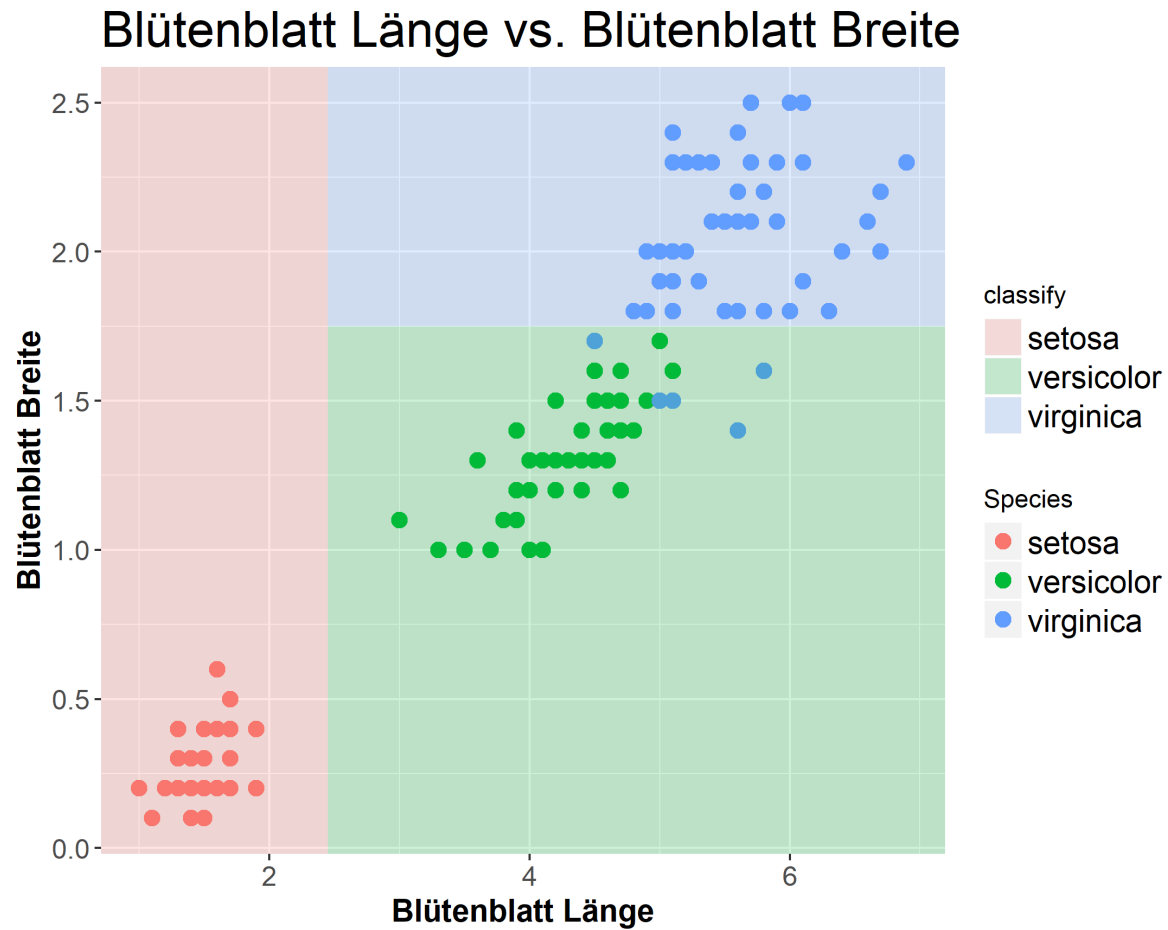
Beispiel: Klassifikation



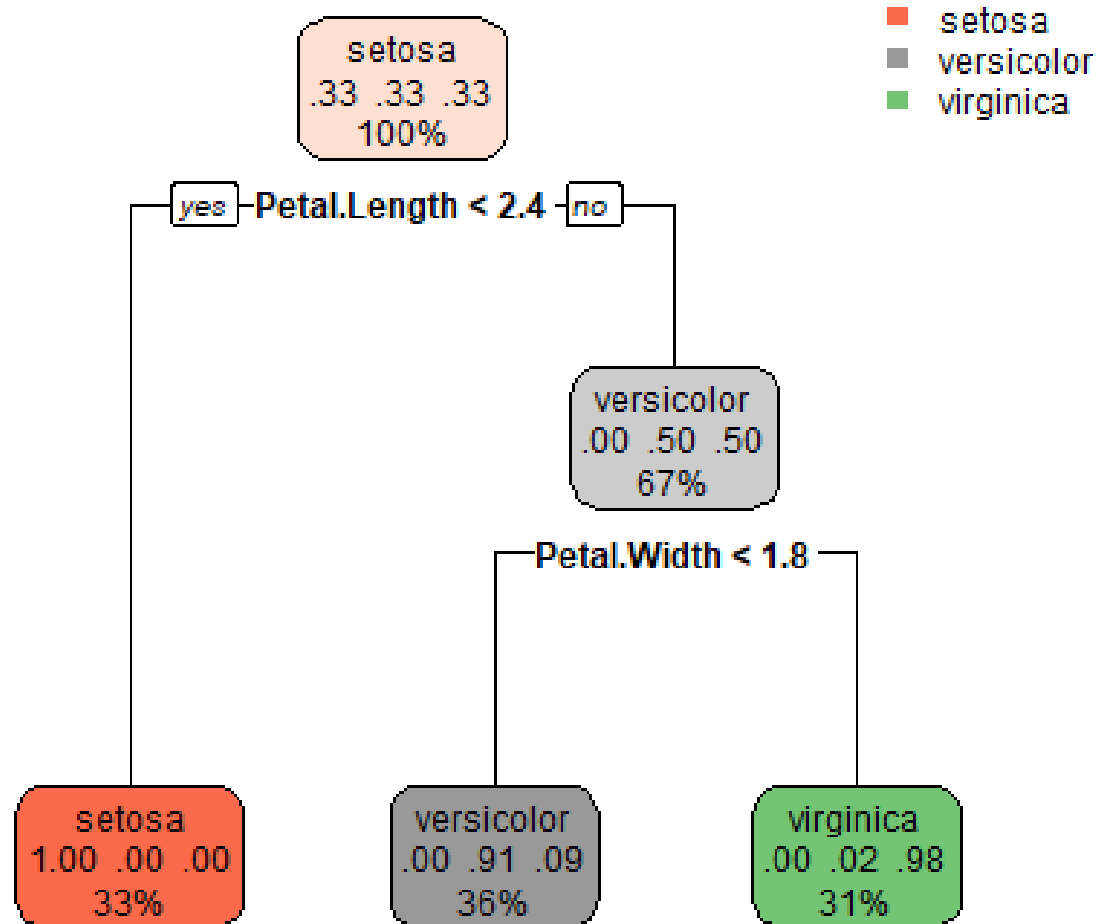
Beispiel: Klassifikation



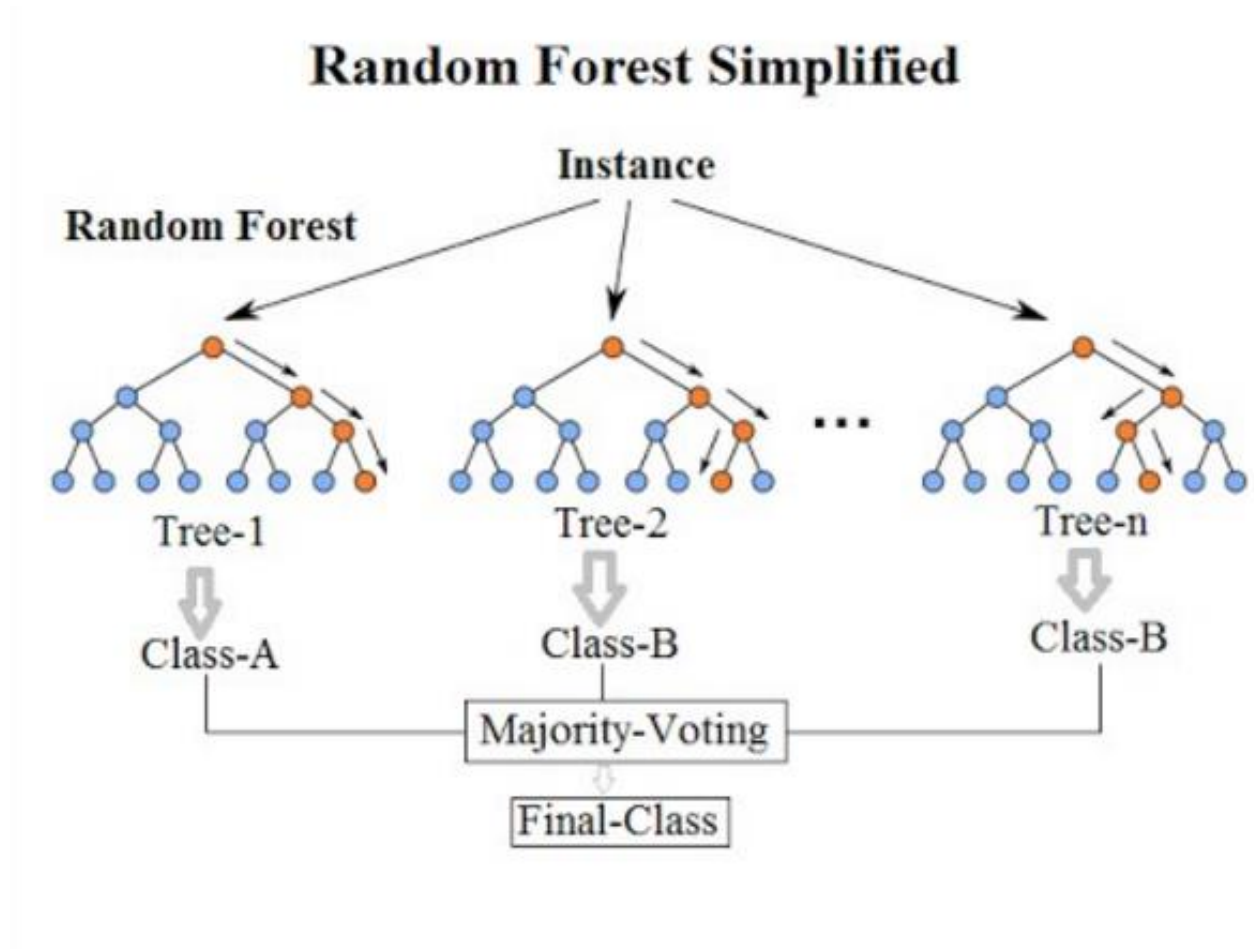
Beispiel: Klassifikation



Decision Tree (Entscheidungsbaum)



Random Forest



Klassifikation (Vorhersage von Klassen / Labels)

Methoden

- Decision Tree
- Random Forest
- Support Vector Machine
- Neural Networks
- ...

Einführung

Usecases



Predictive Analytics

Usecases

- Vorrorausschauende Wartung
- Betrugserkennung
- Ressourcen Planung
- Bildererkennung
- ...

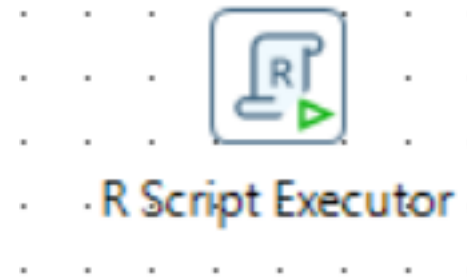
PDI und R



R

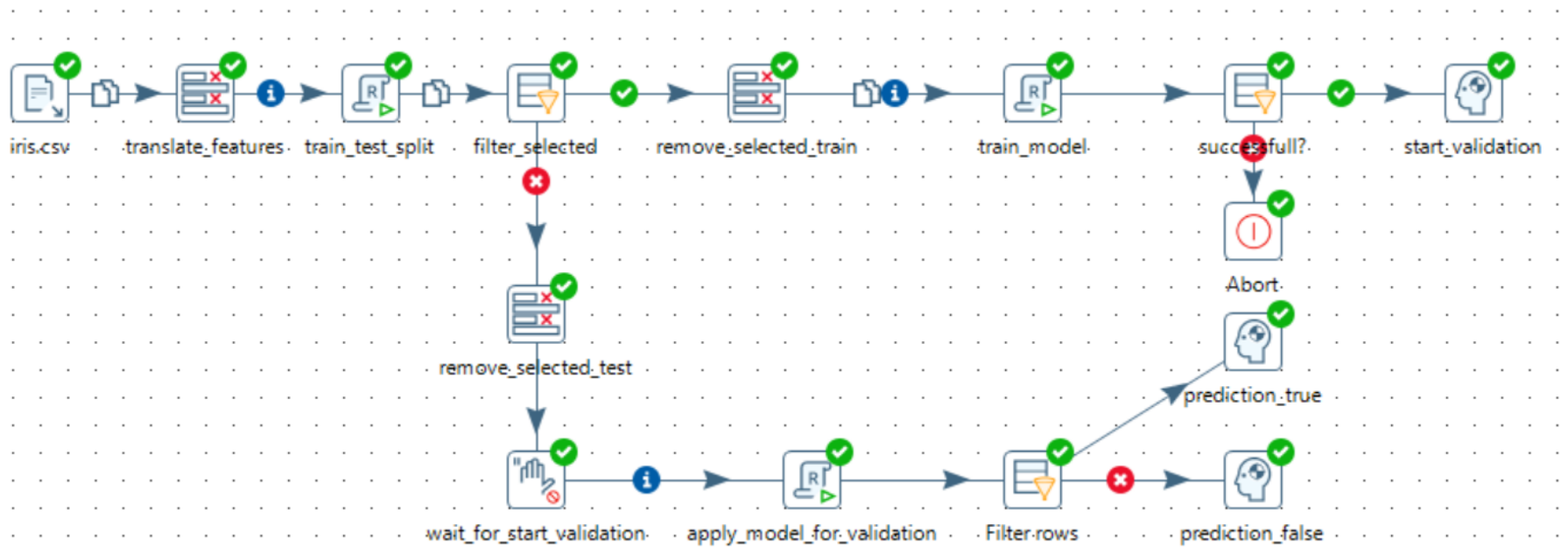
- Freie Statistik Programmiersprache
- Open Source
- Erscheinungsjahr 1995
- Verbreitet in Wissenschaft und Wirtschaft (TIOBE: 13)
- Stärken: Statistische Analysen, Visualisierung
- Schwächen: Datenintegration

R Script executor – the best of both Worlds



- Datenaufbereitung / Integration in PDI
- Statistik / Machine Learning in R
- Deployment der Modelle mit PDI

PDI und R



Vielen Dank für Ihre Aufmerksamkeit!