

# **Vorstellung des Modells für die Kündigungsprognosen**

# Verwendete Software

- KNIME Analytics Platform
  - freie Software für interaktive Datenanalyse
  - graphische Benutzeroberfläche
  - ermöglicht das Aneinandersetzen von Modulen
  - erlaubt eine Einbindung verschiedener Verfahren des maschinellen Lernens

# Datenbasis

- Abonnement-Daten
- Kontakt-Daten
- Soziodemographische Daten
- Kundenbindungsdaten

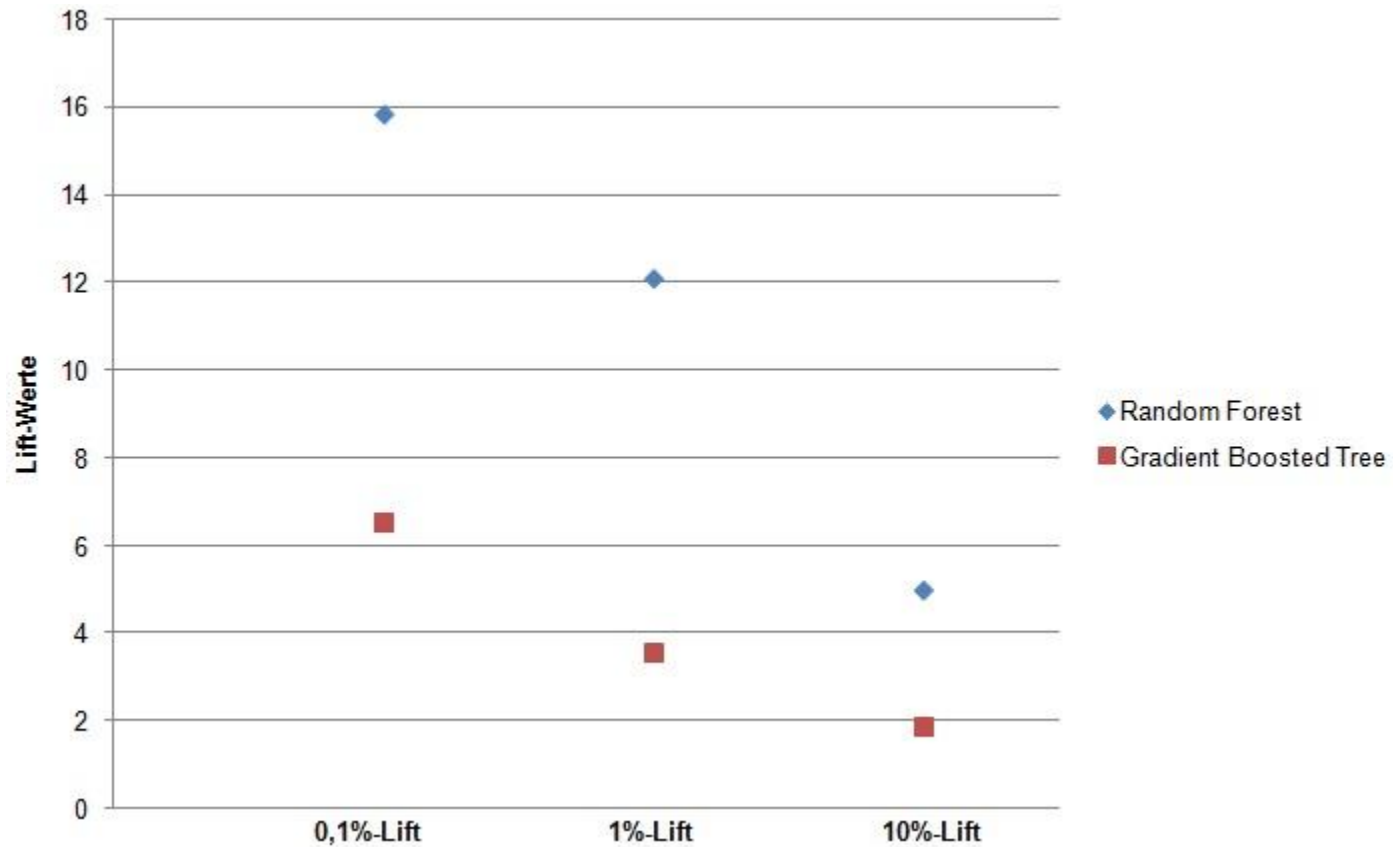
# Datensatz für das Modell

- Ca. 136000 Dateneinträge
- 116 Variablen
  - Zielvariable (Kündiger oder Nicht-Kündiger)
  - 115 Prädiktoren

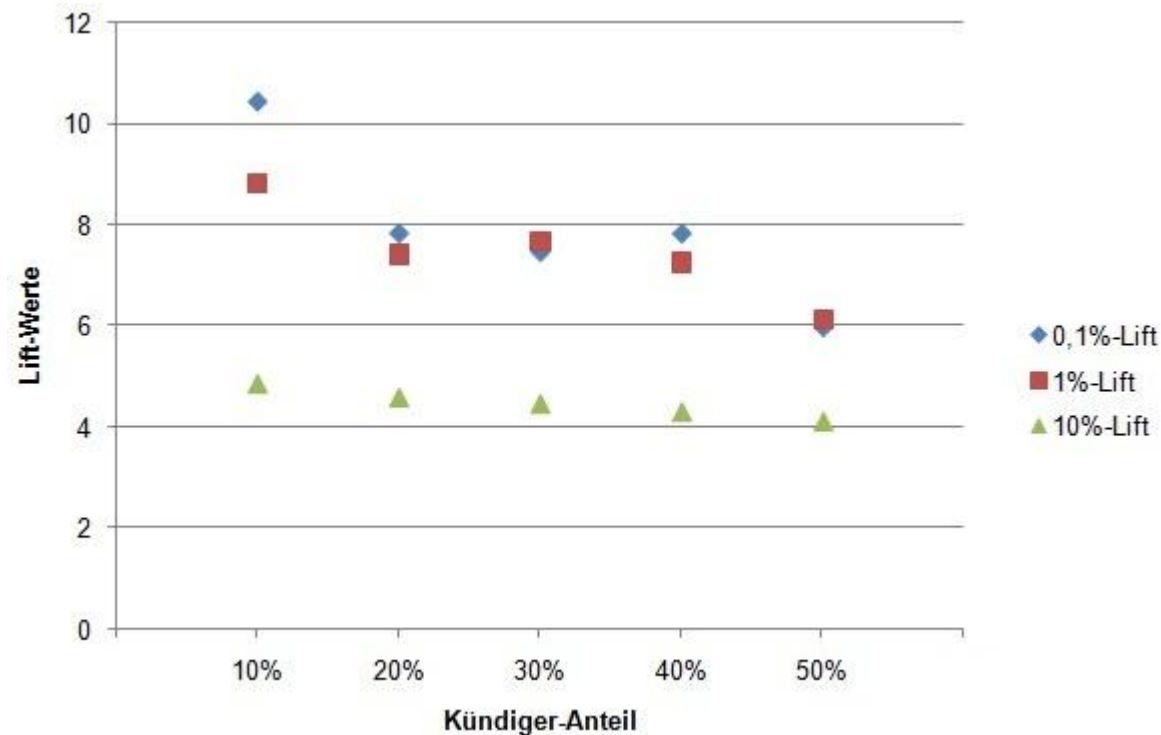
# Klassifikationsalgorithmen mit der Kreuzvalidierung



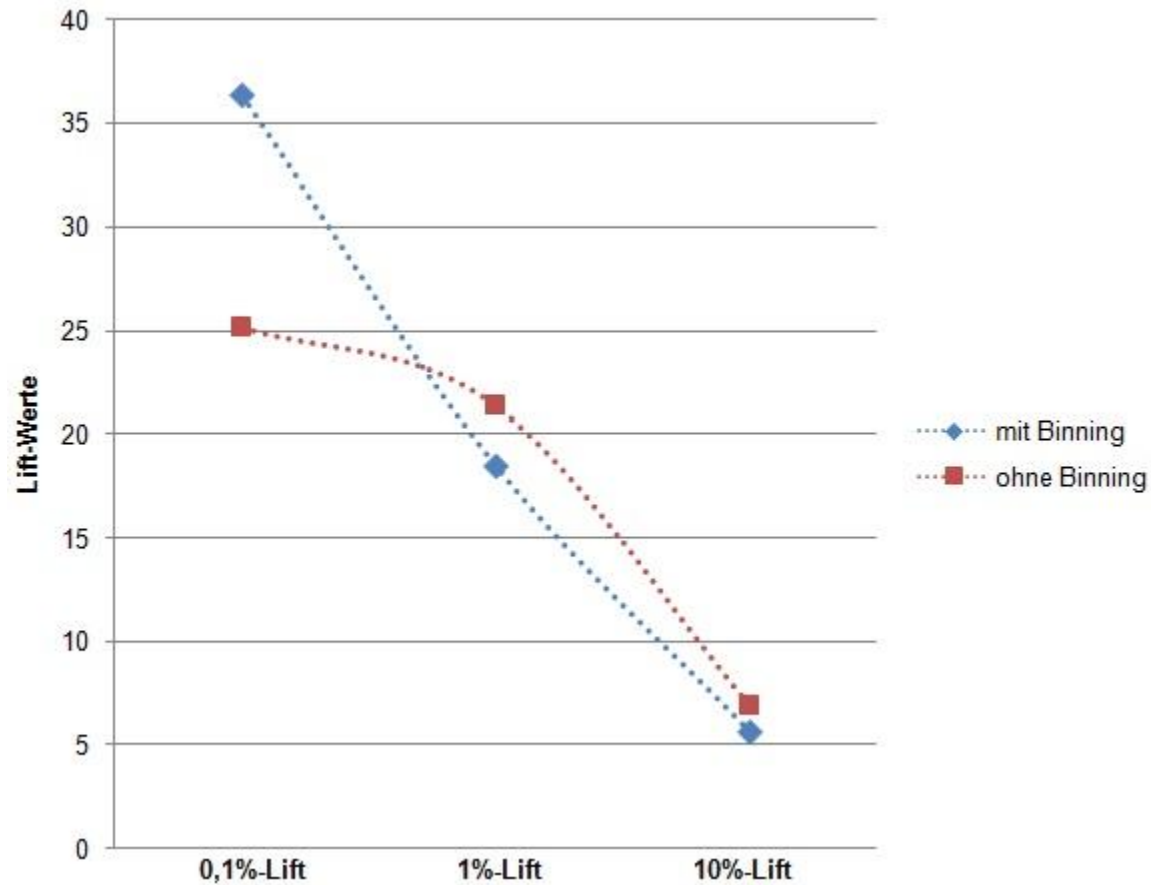
# Klassifikationsalgorithmen mit der out of time Validierung



# Random Forest mit unterschiedlichen Klassenverhältnissen



# Random Forest mit und ohne Binning





# Prognosemodell

- Aufbau:
  - Random Forest mit jeweils 600 Bäumen
  - 8 Zeiträume in den Trainingsdaten
- Kündigungsquote: 0,74%
  - ca. 1000 Kündiger unter 136000 Kunden
- Evaluierung:
  - 0,1%-Lift: 48,3 (49 Kündiger unter 136 Kunden)
  - 1%-Lift: 31,2 (314 Kündiger unter 1360 Kunden)
  - 10%-Lift: 8,1 (815 Kündiger unter 13600 Kunden)