

Protokoll Lernfall 1

Problembestimmung

- Kategorisierung der Variablen (nominal, ordinal, etc.)
- Wie kann man die relevanten Variablen visualisieren?
- Welche ist die jährliche Kündigungswahrscheinlichkeit der Kunden?
- Wie nutzen wir die Daten, um an die relative Kündigungshäufigkeit zu kommen?
- Unterschied: relative Kündigungshäufigkeit vs. Kündigungswahrscheinlichkeit?
- Ist_gekündigt als eine eigene Variable aufschlüsseln
- Wie können wir die Beziehung der relativen Kündigungshäufigkeit zu den anderen Variablen analysieren?

Problemanalyse und Diskussion

- Sterbefälle nicht rausfiltern
- Kündigungswahrscheinlichkeit nicht berechnen
- Vorgehensweise:
 - a. Klassifizierung der Variablen
 - b. Vorbereiten der Daten (z.B ist_gekündigt Variabel)
 - c. Visualisierung der Variablen und Verifizierung der Behauptungen
 - d. Bestimmung der relativen Kündigungshäufigkeit in Abhängigkeit von anderen Variablen
 - e. Unterschied zwischen relativer Kündigungshäufigkeit und Kündigungswahrscheinlichkeit
 - Relative Kündigungshäufigkeit: Bezug auf vorhandenen Daten
 - Kündigungswahrscheinlichkeit: Vorhersage auf zukünftige Kündigungen basierend auf bestehenden Daten
 - f. (Berechnung der jährlichen Kündigungswahrscheinlichkeit)
- Visualisierungen: Scatter Plots, Histogramme, Balkendiagramme, Korrelationskoeffizient
- Beziehung zw. relativen Kündigungshäufigkeit und anderen Variablen: Korrelationskoeffizient

To-Do's

- Kategorisierung der Variablen
- Ist_gekündigt als Variable festlegen
- Histogramme für stetige Variablen
- Balkendiagramme für kategorische
- Korrelationskoeffizienten zwischen den verschiedenen Features

- Bestimmung der relativen Kündigungshäufigkeit
- Setzen der relativen Kündigungshäufigkeit in Abhängigkeit von anderen Variablen
- Unterschied zwischen relativer Kündigungshäufigkeit und Kündigungswahrscheinlichkeit aufschreiben