

Applied Analytical Data Science - Übung 3 -

Dr. Jörg-Uwe Kietz,
Vorlesung an der Univ. Zürich,
Mittwoch, 14:00-15:45 Uhr Vorlesung,
16:00-17:30 Uhr Übung

<http://www.kietz.ch/AADS/>

Vorhanden von der letzten Übung

- 1) Alle relevanten Daten in der nützlichsten Form fürs Data Mining
- 2) Keine fehlenden Werte mehr in den Daten

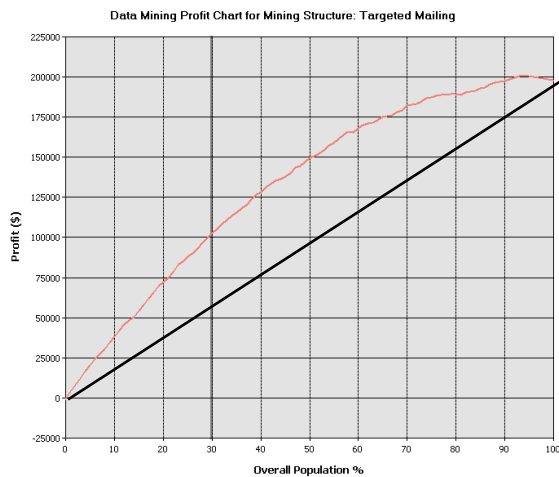
3. Aufgabe: Lernen von Spendet ja/nein

- Splitted die Lerndaten in 66% Training und 33% Test
- Klassifikationslernen funktioniert NICHT (gut) mit 95:5
 - Balanciert die Daten (positiv:negativ ~ 1:1) durch
 - duplizieren der positiven, oder
 - Auswahl der negativen (Ensemble über mehreren Samples)oder
 - Wandelt Target_D in ein “Weight”-Attribut um (Weight Attribut muss Werte zwischen 0 und 1 haben)
Weight = abs(Profit)
- Lernt ein Model auf den Trainingsdaten
- Benutzt „Apply Model“ auf den Test-Daten

Aufgabe 3

- RM produziert 3 neue Attribute
 - Predicted(Target_B), confidence(false), confidence(true)
- Benennt sie um und ändert die Role
Predicted_Target_B, confidence_false, confidence_true
um sie besser weiter benutzen zu können.
- Order = confidence_true (descending) oder
- Order = confidence_true (ascending)

Profit Chart



Order the population by estimated likelihood

Sum up the Profit gained mailing the best X%

Y-axis: Cummulated Profit
X-axis: Percentage of Pop.

Profit Chart Erstellung

- Benutzt das R-Script aus der 1. Übung

Vortrag jeder Gruppe in der nächsten Übung

- Welche Klassifikationsmethoden habt ihr probiert?
- Wie habt ihr die Daten Balanciert oder Gewichtet benutzt?
- Wie habt ihr sie Evaluert?
- Welche waren wie gut?
- Waren die Besten gut genug?
- Wo hat der Profit-Chart sein Maximum?
- Was kann man besser machen?

Abgabe der 3. Übung

- Vortrag am 18.04.18 (Mittwoch in der Übung)
- RM Process, Modelle und Resultate (Report, Graphs, Validation) via Email bis 16.04.18 (Montag abend)
Bitte einen gezippten Folder namens GrXUebY mailen.