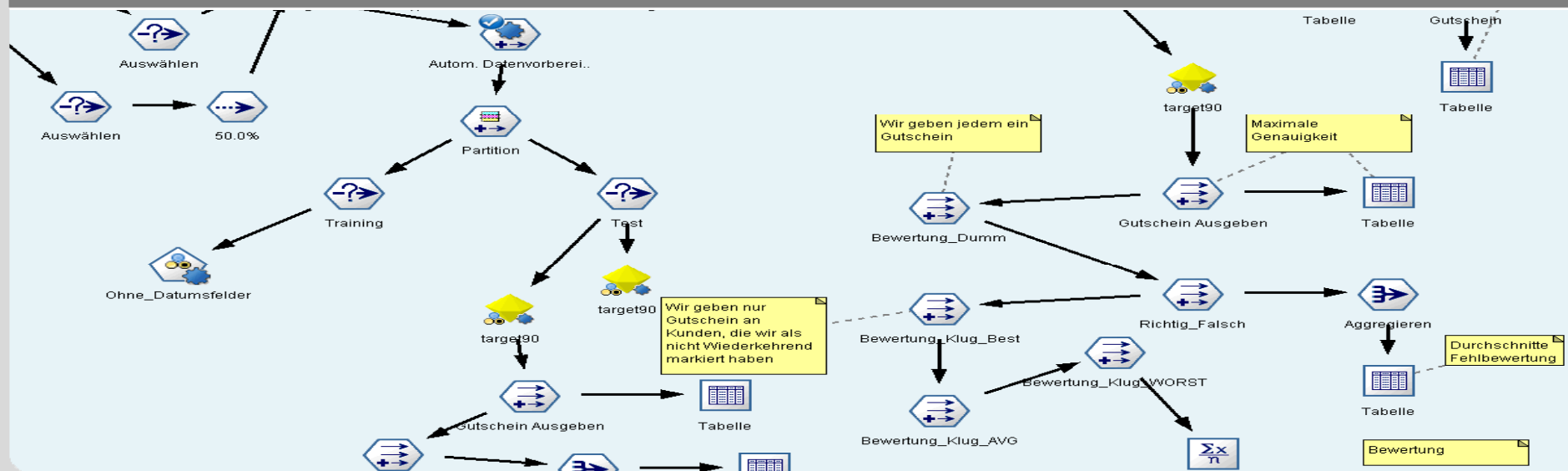


Data-Mining Cup - Zwischenpräsentation

Gruppe 4: Andreas, Andriy, Patrick, Raimund, Sven, Zhen

IPD, Lehrstuhl für Systeme der Informationsverwaltung, Prof. Böhm



Gliederung der Aufgabenstellung in zwei Teile

- Target90 möglichst genau vorhersagen
- Bewertungsfunktion

Target90 möglichst genau vorhersagen

- Eliminierung unwichtiger Spalten (Points)
- Umformen und Zusammenfassen von Datumsspalten in aussagekräftigere Darstellung:
 - $\text{DeliveryDateDiff} = \Delta(\text{DeliveryDateReal}, \text{DeliveryDatePromised})$

Δ	Bedeutung
<0	schneller geliefert als versprochen
0	Lieferdatum entspricht versprochenem Datum
>0	Lieferung verspätet

- Nullwerte bei DeliveryDateReal berücksichtigen
 - Downloads (nur $w3 \neq 0$ oder $w5 \neq 0$)
 - etc.

Target90 möglichst genau vorhersagen

■ Daten vervielfachen mit

1. target90 = 0

Verbesserung der Vorhersagegenauigkeit um ca. 1 – 2 Prozent

2. target90 = 1

Wesentliche Verschlechterung der Vorhersagegenauigkeit

■ Verwendung verschiedener Modelle

1. SVM

2. Neuronales Netz

3. Bayes

4. Kombination verschiedener Modelle

Bewertungsfunktion

- Definition der Bewertungsfunktion
 - Ansatz:
 - Vorhersage bei Target90 = 1 auf jeden Fall keinen Gutschein
 - Konzentration auf Vorhersage bei Target90 = 0
 - Wert von ca. 14200 (SVM)

- Maximierung der Bewertungsfunktion
 - Suche einer Funktion, die den möglichen Gewinn in Abhängigkeit von der Konfidenz der Vorhersage berechnet

Offene Fragen

- Tatsache, dass nur 10% der Gutscheine tatsächlich eingelöst werden, bereits in Gewinn von 1,50 Euro berücksichtigt?
 - Was bedeutet model
 - Was bedeutet es, wenn die Lieferung 9 Monate vor dem versprochenen Datum liegt (z.B. Cust.-No. 62744)
 - DeliveryDatePromised = 26.11.4746 (z.B. Cust.-No. 41577)
-
- Es existieren FAQ, evtl. weiterleiten von Fragen an die Veranstalter des DMC.