Aufgabenstellung „Briefaktion“

Eine Bank möchte ihren Gewinn erhöhen. Die Buchhaltung hat festgestellt, dass Kunden, die ein sog. Premium-Konto führen („Premium-Kunden“), den höchsten Beitrag zum Gewinn leisten. Daher soll die Anzahl der Premium-Kunden erhöht werden. Zu diesem Zweck ist eine personalisierte Massen-Mail geplant. Um allerdings

1. die Kosten für die Briefaktion zu minimieren,
2. gezielt nur die Kunden anzusprechen, die mit einer gewissen Wahrscheinlichkeit Interesse an einem Premium-Konto haben,
3. keine Kunden mit unnötiger Post zu belästigen, die höchstwahrscheinlich sowieso kein Premium-Konto gebrauchen können,

soll nur ein entsprechend gewählter Teil der Kunden angeschrieben werden.
Die Auswahl dieser Kunden ist Ihre Aufgabe!

Vorgehensweise:

Filtern Sie die Premium-Kunden und clustern Sie diese. Dadurch erhalten Sie Gruppen ähnlicher Premium-Kunden. Dann ordnen Sie alle übrigen Kunden diesen Clustern zu, d.h. Sie klassifizieren die Nicht-Premium-Kunden anhand der gefundenen Klassen. Interpretieren Sie Ihr Ergebnis.

Milestones:

<u>Milestone 1 (Abnahme: 8.1./12.1.03)</u>

- Implementierung geeigneter Datenstrukturen
- Implementierung eines Clustering Algorithmus

<u>Milestone 2 (Abnahme: 29.1./3.2.03)</u>

- Implementierung eines geeigneten Klassifikators

Datensatz:

Der Datensatz „banking“ enthält die Kundendaten unserer Bank. Das Portfolio dieser Bank enthält verschiedene Kontoarten, u. a. ein normales Giro-Konto (product=0), ein Premium Konto (product=1), ein Exklusiv-Konto (product=3), verschiedene Spar-Konten usw. Weiterhin führt die Bank verschiedene Options-Pakete im Angebot. Für jeden Kunden kennt die Bank zusätzlich sein Geschlecht (gender), sein Alter (age), die Anzahl der Geschwister (siblings), sein Einkommen (income), das gewählte Optionspaket (type), und die Art des Kontos.
Die Datei „banking.txt“ enthält diese sechs Felder jeweils durch Komma getrennt: gender, age, siblings, income, type, product.