Übung 3

Prof. Dr. Alexandros Nanopoulos, Rasoul Karimi
Wirtschaftsinformatik und Maschinelles Lernen (ISMLL)
Universität Hildesheim

Deadline : June 30

Aufgabe 1: Klassifikation-Welche Kreditkarte darf es denn sein? (30 Punkte)

Aufgrund Deiner tollen Arbeit hast Du Dein Ziel erreicht und bist nun optimal sozialversichterter Angestellter bei der Firma FoodMart Inc.. Nach der OLAP-Analyse sollst Du nun anhand der zur Verfügung stehenden Daten und Sql Server 2008 herausfinden, welchen Kunden welchen Kundenkartenstatus besitzen. Analysiere die Kundendaten dahingehend, welche kundenindividuellen Parameter wie z.B. Bildung, Geschlecht,. . . den Status ihrer Member Card (Golden, Silver,. . .) bestimmen. Danach erstelle mit den drei Modellklassen Entscheidungsbäume, und Naive Bayes Vorhersagemodelle, die diese Membercard-Stufe möglichst präzise vorhersagen. Für die Evaluation der Modelle verwende eine Aufteilung in Test- und Trainingsdaten im Verhältnis 3:7 und Cross- Validation. Beantworte dann folgende Fragen:

• Welche Parameter scheinen Einfluss zu haben und vor allem wie beeinflussen sie den Membercard Status?
• Warum verwendet man Cross-Validation?
• Wie unterscheiden sich die 2 verwendeten Modellklassen hinsichtlich
1. Vorhersagegenauigkeit?
2. Interpretierbarkeit der Resultate?
3. Schwierigkeitsgrad der Modellparametersuche?
4. Zugänglichkeit für nominale, ordinale und stetige Variablen (als Ziel- sowie als Inputvariable)?
5. Geschwindigkeit der Berechnung?

Aufgabe 2: Clustering-Kosten senken (30 Punkte)

Als nächstes gibt Dir Dein neuer Chef die Aufgabe die Produktpalette dahingehend zu untersuchen, obes in der Gesamtmenge der angebotenen Produkte größenmäßige Übereinstimmungen gibt, sodass in Zukunft ähnlich große Produkte in dann standardisierten Regalen dem Kunden angeboten werden könnten. Seine famose Idee ist nämlich ein neues Store-Konzept, wonach die Produkte nicht mehr nach Kategorie angeordnet werden, sondern nach ihrer Größe. Das hätte neben einem Kostenvorteil durch standardisierte Regale den Vorteil, dass die Marke FoodMart Inc. unverwechselbarer werden würde. Untersuche daher anhand der Attribute shelf-width, shelf-height und shelf-depth folgende Aufgabenstellung:

• Welche Produkte sind größenmäßig ähnlich und in Clustern zusammenfassbar?
• Wie viele Cluster hast Du gewählt und warum?
• Versuche die Einteilung der Produkte in die von Dir gefundenen Cluster zu verstehen. Wie kommt es zu diesem Ergebnis? Unterscheide zwischen inhaltlichen und Modellierungsgründen.