

Anhang II: Vorgehensweise zur monetären Bewertung alternativer Klassifikatoren

Eine monetäre Bewertung von Klassifikatoren erfordert eine Angabe der mit korrekten bzw. falschen Klassenprognosen verbundenen Kosten und Erträge. Für die DMC-Datensätze finden sich diese Informationen in den Aufgabenbeschreibungen des jeweiligen Wettbewerbs. Analog enthält die Beschreibung des Datensatzes GC eine Kostenmatrix.¹² Für die verbleibenden Datensätze AC, Adult und Coil sind keine entsprechenden Informationen bekannt. Daher wurde hier die in der Literatur übliche Annahme getroffen, dass die Fehlklassifikationskosten einer Klasse der A-priori-Wahrscheinlichkeit der anderen Klasse entsprechen. Hierdurch wird sichergestellt, dass Fehlprognosen der Minderheitsklasse härter bestraft werden. Die resultierenden Kosten und Erträge falscher bzw. korrekter Klassifikationen zeigt Tab. 7.

[Tab07]

Tab. 7 Kosten und Erträge falscher bzw. korrekter Klassifikationsprognosen

	Wahre Klasse	Prognose	
		-1	+1
AC	-1	0	0,4449
	+1	0,5551	0
GC	-1	0	1
	+1	5	0
Adult	-1	0	0,2393
	+1	0,7607	0
Coil	-1	0	0,06
	+1	0,94	0
DMC 2000	-1	0	-6
	+1	-95	95
DMC 2001	-1	1.110	662
	+1	-265	-25
DMC 2002	-1	72	66,3
	+1	0	43,8
DMC 2004	-1	1	-1
	+1	-1	1
DMC 2005	-1	15	13
	+1	-25	2

Ferner ist zu berücksichtigen, dass die Berechnung von Erträgen und Fehlklassifikationskosten eine diskrete Klassenzuordnung erfordert, wohingegen Klassifikationsmodelle oftmals lediglich eine Klassenzugehörigkeitswahrscheinlichkeit bzw. Konfidenz¹³ ausgeben (Fawcett 2006, S. 863). Folglich ist eine Konvertierung bzw. Nachverarbeitung erforderlich, um eine diskrete Einteilung der zu klassifizierenden Objekte zu erreichen. Im Mittelpunkt praktischer Anwendungen steht i. d. R. eine effiziente Allokation vorgegebener (Marketing-)Ressourcen. Das bedeutet, dass z. B. die Anzahl der zu kontaktierenden Kunden durch das verfügbare

¹² [http://archive.ics.uci.edu/ml/datasets/Statlog+\(German+Credit+Data\)](http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data))

¹³ Konfidenz meint in diesem Zusammenhang einen numerischen Wert, der die Sicherheit eines Modells ausdrückt, dass ein Objekt einer bestimmten Klasse angehört, aber nicht im Intervall [0,1] skaliert ist und somit keine echte Wahrscheinlichkeit darstellt.

Marketingbudget exogen vorgegeben ist und die vom Klassifikationsmodell generierten Konfidenzen verwendet werden, um die entsprechende Anzahl von Kunden auszuwählen. Da für die hier betrachteten Anwendungsfälle keine Budgetrestriktionen bekannt sind, wurde zur Simulation eines solchen Szenarios die Annahme getroffen, dass die Anzahl der jeweils zu kontaktierenden Kunden anhand der a priori Wahrscheinlichkeit der betriebswirtschaftlich relevanten Klasse bestimmt werden kann. Für jeden der neun Datensätze wurde diese Wahrscheinlichkeit aus den Trainingsdaten geschätzt und die korrespondierende Anzahl von Objekten mit größter Konfidenz aus der Testmenge ausgewählt. Diese Fälle wurden der Klasse 1 zugeordnet, alle anderen der Klasse 0. Anschließend konnte unter Verwendung der Ertrags- bzw. Kostenmatrizen je Datensatz und Klassifikator der aus der Anwendung des Verfahrens resultierende Gewinn ermittelt werden.