

Revenue maximisation by intelligent couponing

Thomas Friedrich, Lukas Schmauch

Friedrich-Schiller-Universität Jena

20. Januar 2020

Übersicht

1 Aufgabenstellung und Datensatz

2 Klassifikation und Evaluation

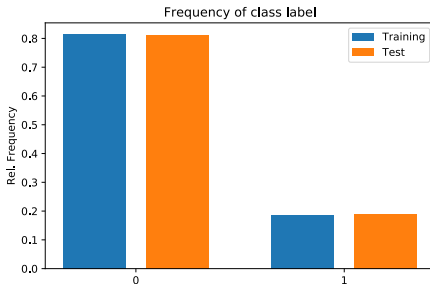
Ausgangssituation

- Data Mining Cup 2010
- **Ziel:** Gewinnmaximierung anhand der Ermittlung der Wiederkäufer

		Vorhergesagt	
		kein Wiederkäufer(0)	Wiederkäufer(1)
Tatsächlich	kein Wiederkäufer(0)	1.5	0
	Wiederkäufer(1)	-5	0

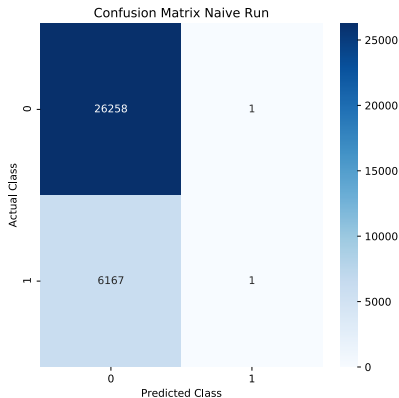
Der Datensatz

- 64854 Einträge
- 50:50 Train/ Test
- 38 Merkmale
- 20 numerisch
- 18 kategorisch



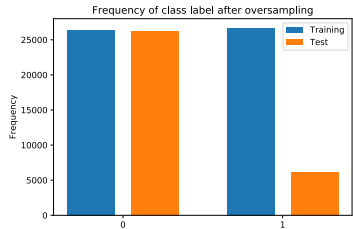
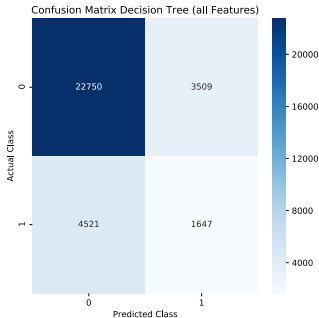
Konfusionsmatrix Naiver Run (Decision Tree)

- **erzielter Umsatz:** 8552.00 €
- **Umsatzsteigerung:** 0.04 %
- **Accuracy:** 81.00 %

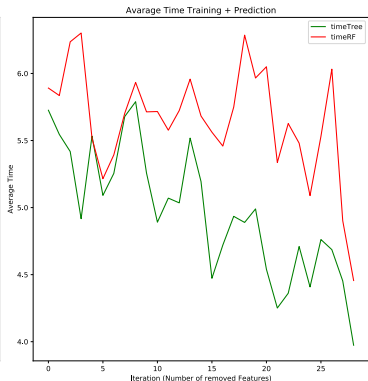
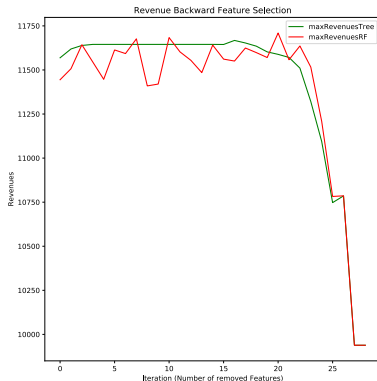


Konfusionsmatrix nach Oversampling (Decision Tree)

- **erzielter Umsatz:** 11520.00 €
- **Umsatzsteigerung:** 34.76 %
- **Accuracy:** 75.23 %



Backward Feature Elimination und Zeitmessung



Zusammenfassung der Ergebnisse

Klassifikator	Umsatz	Steigerung
Ohne Optimierung	8548.50 €	-
Decision Tree Naiv	8552.00 €	0.04 %
Decision Tree Oversampling	11667.50 €	36.49 %
Random Forest Oversampling	11709.50 €	36.98 %
Optimaler Klassifikator	39388.50 €	-

Quelle: Aufgabenstellung und Datensatz



Data Mining Cup 2010

<https://www.data-mining-cup.com/reviews/dmc-2010/>